CrossMark

No-reference image quality assessment
with center-surround based natural scene statistics

Jun Wu[1] · Zhaoqiang Xia[1] (ID) **· Huifang Li[1] ·**
Kezheng Sun[2] · Ke Gu[3] · Hong Lu[3]

Abstract In this paper, we propose an efficient no-reference image quality assessment (NR-IQA) method dubbed Center-Surround based Blind Image Quality Assessment (CS-BIQA). Our proposed method employs the Difference of Gaussian (DoG) model to decompose images into several frequency bands, considering the center-surround effect and multi-channel attribute of human visual system (HVS). The integrated natural scene statistics (NSS) features can be further derived from all DoG bands. After that, regression models between the integrated features and associated subjective assessment scores are learned on the training dataset. Subsequently, the learned models are used to predict the quality scores of test images. The main contribution of this paper is twofold. Firstly, the empirical distributions of DoG bands of images are proven to be a Gaussian-like distribution. And thus, the

---

✉ Zhaoqiang Xia
xiazhaoqiang@gmail.com

Jun Wu
wujun318@mail.nwpu.edu.cn

Huifang Li
lhuifang@nwpu.edu.cn

Kezheng Sun
528106tyhj@163.com

Ke Gu
guke.doctor@gmail.com

Hong Lu
zdhxlh@njit.edu.cn

[1] School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China

[2] School of Information Technology and Electrical Engineering, Jiangsu Vocational College of Business, Nantong, China

[3] School of Computer Science and Engineering, Nanyang Technological University, Singapore, Singapore

NSS features can be employed to represent the perceptual quality of images. Secondly, different types of distortions are observed to affect different frequency components of images. So, the integrated features extracted from multi-frequency bands are employed in CS-BIQA to achieve stronger distinguishable capability of image quality. Excessive experiments are conducted to indicate that our proposed CS-BIQA metric can represent the perceptual characteristics of HVS. The results on popular IQA databases demonstrate that the CS-BIQA metric is competitive with the state-of-the-art relevant IQA metrics. Furthermore, our proposed method has very low computational complexity, making it more suitable for real-time applications.

1 Introduction

A picture paints a thousand words. So pictures or images are the principal information carriers in our daily life. High-quality images deliver clear and complete information to observers while images with low quality usually induce confusion. Usually, acquiring high-quality images faces many challenges as pristine images will be distorted in image processing tasks (e.g., image acquisition, compression, restoration, transmission). Consequently, a reliable image quality assessment (IQA) method is significant to ensure image quality in these tasks. Depending on the availability of reference information, IQA can be classified into three categories: 1) full-reference (FR) IQA which needs whole reference [12], 2) reduce-reference (RR) IQA which needs partial reference [11], and 3) no-reference (NR) IQA which evaluates image quality without any reference. In this paper, we focus on the NR-IQA problem.

NR-IQA metrics can be classified into two categories: 1) distortion-specific (DS) metric and 2) general-purposed metric. The DS-IQA metrics commonly tackle one or a few particular kinds of distortions, such as blocking artifacts [16], ringing effects [18], blurring [32] or contrast-change [10]. Thus, the applications of DS-IQA metrics are rather limited. On the other side, the general-purposed IQA metrics aim at evaluating the quality of image without any prior information of distortions [9, 21, 22]. Therefore, the general-purposed metrics have stricter conditions and wider applications.

In this paper, we propose a NR-IQA metric by considering the center-surround effect and multi-channel attribute of human visual system (HVS). There is considerably evidence that the ganglion cells in retina have the center-surround receptive fields [1, 2]. The cells exhibit lateral inhibition which can be well simulated by the Difference of Gaussian (DoG) model. Moreover, image distortions are observed to have specific statistical characteristics on different frequency components of images. This phenomenon will be further discussed in Section 3.2. Therefore, the statistical features are extracted from every DoG bands for obtaining a faithful representation of image quality in our proposed method. Then the features of all DoG bands are combined into the comprehensive feature. Finally, the comprehensive feature and associated human opinion scores of training dataset are fed into the support vector regression (SVR) to learn a regression model. The quality score of distorted image can be predicted using the trained regression model.

The rest of this paper is organized as follows. Section 2 presents the related work of general-purposed NR-IQA and existing DoG-based IQA metrics. Section 3 presents the

details of our proposed method. Section 4 presents the experimental results, and Section 5 concludes the paper.

2 Related work

2.1 General-purpose NR-IQA metrics

In last decade, enormous of elaborated general-purposed NR-IQA metrics have been proposed. These metrics can be divided into two categories, which are called as *opinion-aware IQA* (OA-IQA) and *opinion-unaware IQA* (OU-IQA), depending on the availability of subjective opinion scores (human opinion). OU-IQA metrics only use unlabeled distorted images while they always need auxiliary methods to acquire quality benchmarks. Comparatively speaking, OA-IQA methods use subjective opinion scores of images, so the latent relationship between the images and subjective opinion scores can be learned from the training data intuitively and precisely. In order to evaluate the image quality more effectively and efficiently, many powerful quality-aware features are designed to represent the quality of images. These studies almost share a unified architecture which trains a model in training procedure and predicts the image quality in testing procedure.

Opinion-unaware method In [21], Mittal et al. construct a collection of quality-aware features from a corpus of high-quality images. Then these features are fitted with a global multivariate Gaussian (MVG) model. By using the pre-trained MVG model as the benchmark of image quality, the quality of test image can be expressed as the distance between MVG model of test image and benchmark. In [36], Zhang et al. improve NIQE in two aspects, including by using enhanced features and adopting local MVG model. The integrated Local NIQE (IL-NIQE) shows state-of-the-art performance on several IQA databases. In [19], Lu et al. extend NIQE to shearlet domain. In [13], Gu et al. employ a reliable modified PCQI metric [34] to calculate the labels of training images. Five factors, including image contrast, sharpness, brightness, colorfulness, and naturalness are considered to extract 17 features. After that, the SVR is employed to learn the regression model and predict the image quality. Although OU-IQA methods have achieved preliminary successes, most of these literatures only demonstrate their availability on limited types of distortions.

Opinion-aware method In [22] and [23], the statistical features are extracted from wavelet subbands. The probabilities of distortion categories are estimated by a multi-class support vector machine (SVM). Then, the SVR tool or other distortion-specific IQA methods are suggested to calculate a quality score for each distortion category. Finally, the total image quality score is represented as a probability-weighted summation of individual quality scores. Both BIQI [22] and DIIVINE [23] assume that the categories of distortions in the test dataset are represented in the training dataset, which is not the case in many practical applications. Meanwhile, Saad et al. [28] propose a BLIIND metric based on the hypothesis that the statistical features of images in DCT domain can vary in a predictable way as the image quality changes. So, the statistical features are extracted to represent the contrast and structure of images, and the MVG distribution is used as a probability prediction model to predict the image quality. Later on, more sophisticated features are developed to extend BLIINDS to BLIINDS-II [29]. In [20], the divisive normalization strategy is applied to reduce the correlation between surrounding pixels of images. The empirical distribution of locally normalized luminance coefficients and the pairwise products of these coefficients

are utilized to design the features for image quality prediction. In [9], quality-aware features are extracted from three aspects, including the structural degradation, the HVS-inspired features and the possible losses of naturalness in distorted images. In [35], Xu et al. employ a codebook-based framework for adaptively extracting quality-aware features instead of the handcraft ways. As SVR is a commonly used and efficient machine learning tool in learned based methods, it is inevitably employed by these methods [9, 13, 20, 35] to learn the mapping function between features of training images and associated quality scores for predicting the quality scores of test images.

2.2 DoG-based IQA metrics

Over the years, the new IQA metrics put more emphasis on combining human visual psycho-physiological properties. In [5], images are divided into the salient and no-salient regions by an improved *Itti* algorithm. Then the natural scene statistic (NSS) features are extracted from two kinds of regions, separately. Finally, the features are combined to predict the image quality by SVR. However, the visual salient mechanism actually affects the importance of image regions in IQA. The NSS features extracted from the salient and no-salient regions have no apparently complementary effect. Except these salient-based methods, the DoG model simulated the exhibit lateral inhibition of the ganglion cells in retina has been preliminarily used in IQA methods. In [24], Pei et al. propose an FR-IQA method to calculate image quality scores on each DoG band by the classical FR-IQA metrics, in which the quality scores of all DoG bands constitute a new feature. Then, a random forest regression approach is employed to predict the quality of images. Pei and Chen [24] obtains a remarkable performance, however, it is limited by the reference information. In [6], an FRIQUEE metric is proposed. The single-layer DoG filter is performed on the sigma map of images and six statistical features are extracted. Then, these DoG-based features are integrated with amounts of other features to predict the image quality by a deep belief network. Ghadiyaram and Bovik [6] achieves high performance on various IQA databases. However, only single-layer DoG filter is performed on the sigma maps of images, so the multi-frequency characteristic of DoG model is not used in FRIQUEE. Meanwhile, the sigma map mainly describes the high-frequency attribute of an image, while the low-frequency attribute of images is not considered. Moreover, the FRIQUEE method costs a large amount of time for extracting the abundant features, making this method unpractical. In [30], multi-scale dictionaries are learned from DoG bands of left training images by sparse representation algorithms. The sparse representation coefficients of images are utilized to generate the quality-aware feature to predict the image quality by SVR. Due to the features used in [30] are computed from a set of learned dictionaries, the images employed to learn the dictionaries become very important, but there are still no criterion to choose the training images.

3 Proposed method

In the low-level vision of retina, the ganglion cells have the center-surround receptive fields. The cell exhibits lateral inhibition: light in the center is excitatory while light in the surround is inhibitory [1], and this attribute can be well simulated by DoG model. Moreover, the visual responses of HVS can be approximated by a set of DoG models where the number of DoG ranges from 2 to 4 [17]. Through DoG decomposition, images can be divided into several frequency bands. Image distortions are observed to affect image quality

differently on individual DoG bands. By analyzing the degradation induced by distortions on each individual DoG band, the quality of test images can be predicted more precisely. Our proposed method includes three procedures: DoG decomposition, feature extraction and model regression.

3.1 DoG decomposition

DoG is a feature enhancement algorithm that utilizes the subtraction of one blurred version of an original image from another less blurred version of the original image. As a feature enhancement algorithm, the DoG model can be used to improve the visibility of edges and other details presented in an image. Firstly, we decompose an image into several different-frequency bands by the DoG filter. The filter is defined as

$$DOG_{\sigma 1,\sigma 2} = G_{\sigma 1} - G_{\sigma 2} \tag{1}$$

where G_σ is a Gaussian kernel with standard deviation σ. The DoG band of image I can be calculated by

$$I_{DOG_{\sigma 1,\sigma 2}} = I * DOG_{\sigma 1,\sigma 2} = I * G_{\sigma 1} - I * G_{\sigma 2} \tag{2}$$

where $*$ is a convolution operator.

By (2) and several Gaussian kernels with different σ, we can decompose image I into different DoG bands, which can be represented as

$$I_{DOG_i} = \begin{cases} I - I * G_{\sigma 1} & i = 0 \\ I * G_{\sigma i} - I * G_{\sigma i+1} & i = 1 \sim N - 2 \\ I * G_{\sigma i} & i = N - 1 \end{cases} \tag{3}$$

From (3), we can observe that the summation of all DoG bands is identical to the original image. This attribute shows that the DoG model does not lose any information of original image and thus guarantees the full utilization of image information.

The standard deviation σ and the kernel size S of Gaussian kernel can be defined as

$$\sigma_i = k^{i-1}, \quad i = 1 \sim N - 1. \tag{4}$$

$$S_i = 6k^i, \quad i = 1 \sim N - 1. \tag{5}$$

In this paper, the level of DoG is set to be $N = 2$ as the increase of N does not improve the performance significantly. The parameter k in (4) and (5) is set to 1.6. This configuration ensures the computational conveniences and effectiveness of the proposed method.

Finally, we divide an image into high-frequency band and low-frequency band, denoted as DoG_h and DoG_l. As illustrated in Fig. 1, (a) is a reference image, (d) is a distorted version of (a). (b), (e) are the high-frequency bands corresponding to (a), (d), and (c), (f) are the low-frequency bands of (a), (d). We can observe that, (b), (e) represent the detail portion of images, and (c), (f) consist of the basic structures of images. Meanwhile, Fig. 1 shows that the DoG bands of reference image and distorted image present different visual content.

3.2 Feature extraction

It has been proven that the distribution of locally normalized luminance of a natural image closely follows the Gaussian-like distribution, and this distribution can excellently indicate the degree of quality degradation of distorted image [36]. The natural scene statistics (NSS) model is usually used to represent this attribute of a natural image. The locally normalized luminance of image can be computed by local mean subtraction and divisive normalization [27].
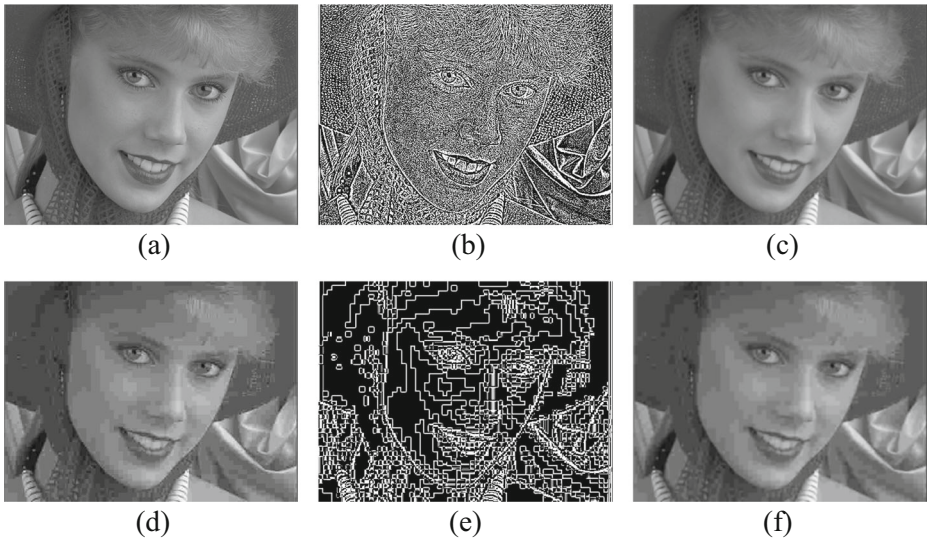
Fig. 1 DoG bands of an exemplar image. **a** is a reference image; **d** is a distorted version of (**a**); **b** and **e** are the high-frequency bands corresponding to (**a**) and (**d**); **c** and **f** are the low-frequency bands of (**a**) and (**d**)

Given a gray image $I(x, y)$, $x \in 1, 2...M$, $y \in 1, 2...N$, M and N are the height and width of an image, respectively. Then, the locally normalized luminance of $I(x, y)$ can be calculated by

$$\hat{I}(x, y) = \frac{I(x, y) - \mu(x, y)}{\sigma(x, y) + C} \tag{6}$$

where $C = 1$ is a constant to prevent instabilities. The $\hat{I}(x, y)$ is also called the mean subtracted contrast normalized (MSCN) coefficients of image $I(x, y)$. $\mu(x, y)$ and $\sigma(x, y)$ are the local mean value and local standard deviation of $I(x, y)$ respectively. They are calculated by

$$\mu(x, y) = \sum_{i=-I}^{I} \sum_{j=-J}^{J} \omega_{i,j} I(x + i, y + j) \tag{7}$$

$$\sigma(x, y) = \sqrt{\sum_{i=-I}^{I} \sum_{j=-J}^{J} \omega_{i,j} (I(x + i, y + j) - \mu(x, y))^2} \tag{8}$$

where $\omega = \{\omega_{i,j} | i = -I, ...I, j = -J, ..., J\}$ defines a unit-volume Gaussian window.

Figure 2a shows the MSCN coefficients' distribution of a reference image and its four commonly distorted versions in spatial domain. Through DoG decomposition, these images are split into DoG_h and DoG_l parts as shown in Fig. 2b and c. We can observe that the empirical distribution of each DoG band also closely follow a Gaussian-like distribution. Based on this observation, the NSS model is employed to extract NSS features from each individual DoG band. The NSS features used in this context contains two parts, including statistics of MSCN coefficients and statistics of MSCN pairwise products.
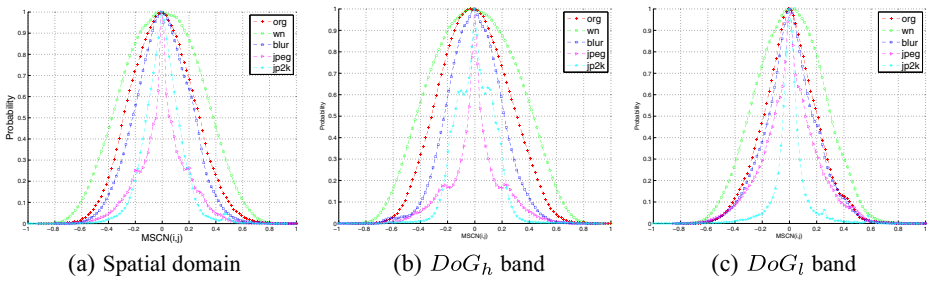
(a) Spatial domain   (b) DoG_h band   (c) DoG_l band

Fig. 2 MSCN coefficients' distribution of a reference image and four commonly distorted images

Statistics of MSCN coefficients The MSCN coefficients' distribution of an image can be captured by a generalized Gaussian distribution (GGD) with zero mean. The density function of GGD is given by

$$f(x; \alpha, \sigma^2) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} exp\left(-\left(\frac{|x|}{\beta}\right)^\alpha\right) \tag{9}$$

where

$$\beta = \sigma\sqrt{\frac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}} \tag{10}$$

and $\Gamma(\cdot)$ is the gamma function,

$$\Gamma(a) = \int_0^\infty t^{\alpha-1}e^{-t}dt \qquad a > 0. \tag{11}$$

The parameter α is a shape parameter, and σ^2 is the variance of GGD. The presence of distortion will change the values of α and σ^2. Consequently, (α, σ^2) constitutes the first part of the quality-aware feature used in our method:

$$f_C = [\alpha, \sigma^2] \tag{12}$$

Statistics of MSCN pairwise products The pairwise products of adjacent MSCN coefficients [20] are suggested to compute along four orientations, separately, as follows

$$H(x, y) = \hat{I}(x, y)\hat{I}(x, y + 1) \tag{13}$$

$$V(x, y) = \hat{I}(x, y)\hat{I}(x + 1, y) \tag{14}$$

$$D1(x, y) = \hat{I}(x, y)\hat{I}(x + 1, y + 1) \tag{15}$$

$$D2(x, y) = \hat{I}(x, y)\hat{I}(x + 1, y - 1) \tag{16}$$

where $x \in \{1, 2, ..., M\}$ and $y \in \{1, 2, ..., N\}$.

The empirical distribution of pairwise products has been shown to obey an asymmetric generalized Gaussian distribution (AGGD). The AGGD with zero mean is given by

$$f(x; v, \sigma_l^2, \sigma_r^2) = \begin{cases} \frac{v}{(\beta_l+\beta_r)\Gamma(\frac{1}{v})} exp\left(-\left(\frac{-x}{\beta_l}\right)^v\right) & x < 0 \\ \frac{v}{(\beta_l+\beta_r)\Gamma(\frac{1}{v})} exp\left(-\left(\frac{x}{\beta_r}\right)^v\right) & x \geq 0 \end{cases} \tag{17}$$

$$\beta_l = \sigma_l\sqrt{\frac{\Gamma(\frac{1}{v})}{\Gamma(\frac{3}{v})}}, \qquad \beta_r = \sigma_r\sqrt{\frac{\Gamma(\frac{1}{v})}{\Gamma(\frac{3}{v})}}. \tag{18}$$

where ν is shape parameter. σ_l^2 and σ_r^2 are scale parameters.

Additionally, the mean value η of AGGD is computed to provide a best AGGD fitting by

$$\eta = (\beta_r - \beta_l) \frac{\Gamma(\frac{2}{\nu})}{\Gamma(\frac{1}{\nu})}. \tag{19}$$

Thus, the second part of our proposed quality-aware feature is constituted of the parameters $(\eta, \nu, \sigma_l^2, \sigma_r^2)$. These parameters are extracted respectively along the orientations of H, V, $D1$ and $D2$.

$$f_H = [\eta_H, \nu_H, \sigma_{lH}^2, \sigma_{rH}^2] \tag{20}$$

$$f_V = [\eta_V, \nu_V, \sigma_{lV}^2, \sigma_{rV}^2] \tag{21}$$

$$f_{D1} = [\eta_{D1}, \nu_{D1}, \sigma_{lD1}^2, \sigma_{rD1}^2] \tag{22}$$

$$f_{D2} = [\eta_{D2}, \nu_{D2}, \sigma_{lD2}^2, \sigma_{rD2}^2] \tag{23}$$

The quality-aware feature $f = [f_C, f_H, f_V, f_{D1}, f_{D2}]$ is extracted on each DoG band, which can be summarized in Table 1. The 18 parameters in f can be estimated using the moment-matching based approach [31]. Moreover, an additional feature is extracted at a down-sampling image with factor 2, considering that image quality is affected by viewing distance and image resolution [8]. Eventually, a 36-dimensional quality-aware feature can be computed from each DoG band.

The advantage of extracting features on DoG bands In Section 3.1, we have introduced the advantages of using DoG model from the aspect of human visual psycho-physiological properties. In this paragraph, we will discuss the superiorities of using DoG model in terms of the discriminatory ability of features. Clearly, an eligible feature in IQA should have two characteristics: **a)** different categories of distortions should have distinctive features; **b)** the features should distinguish the intensities of distortions.

From this point of view, we find that the above-mentioned NSS feature in spatial domain has two drawbacks: **a)** the feature extracted on spatial domain shows low discrimination for some specific distortions. An example is illustrated in Fig. 3. The image "women" and four distorted versions of it in TID database, including additive Gaussian noise (#1), different additive noise in color components (#2), spatially correlated noise (#3) and masked noise (#4) are shown in (a)-(e). The statistical curves of them in spatial domain are shown in (f). Apparently, distortion #2 and #3 heavily overlap each other, which are nearly indistinguishable. Besides, the gap between distortion #1 and #4 is too small to be distinguished. **b)** Some other distortions with different levels of intensities are indistinguishable. Taking JPEG 2000 compression noise for example illustrated in Fig. 4a. The statistical curves of distortions with four intensities are plotted. It clearly demonstrates that the statistical curves of the second to fourth level are heavily overlapped.

Table 1 Summary of extracted features

Feature ID	Feature description	Computation procedure
f_C	Shape, variance	Fitting GGD to MSCN coefficients
f_H	Mean, shape, left variance, right variance	Fitting AGGD to H pairwise products
f_V	Mean, shape, left variance, right variance	Fitting AGGD to V pairwise products
f_{D1}	Mean, shape, left variance, right variance	Fitting AGGD to D1 pairwise products
f_{D2}	Mean, shape, left variance, right variance	Fitting AGGD to D2 pairwise products
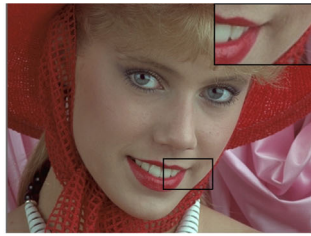
(a) Reference image
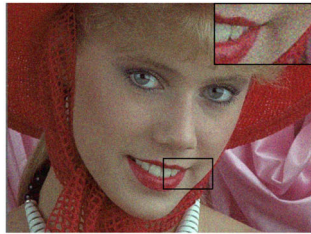
(b) # 1 distorted image
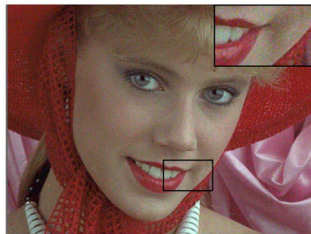
(c) # 2 distorted image

(d) # 3 distorted image

(e) # 4 distorted image

(f) The empirical distributions of (a)-(e) on spatial domain

(g) The empirical distributions of (a)-(e) on DoG_h

(h) The empirical distributions of (a)-(e) on DoG_l

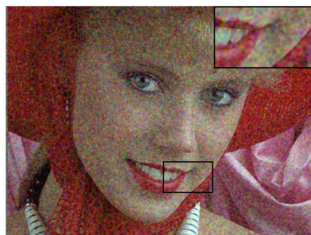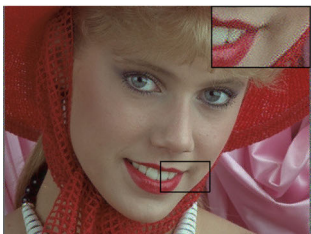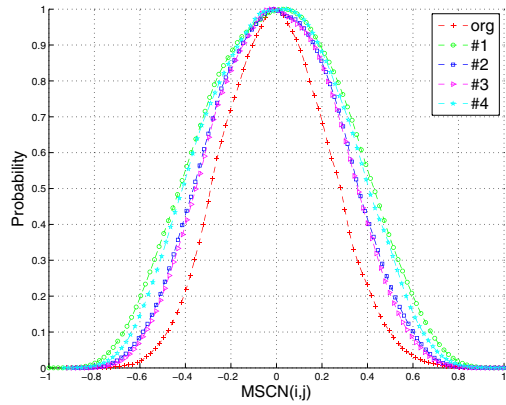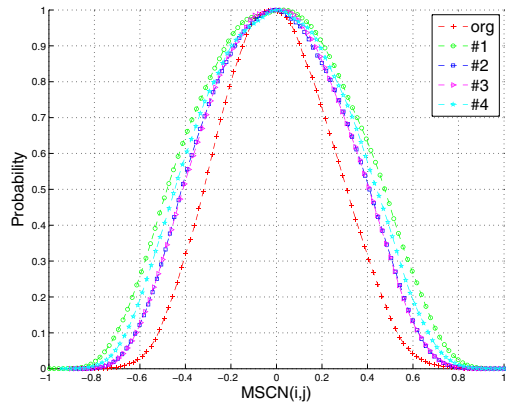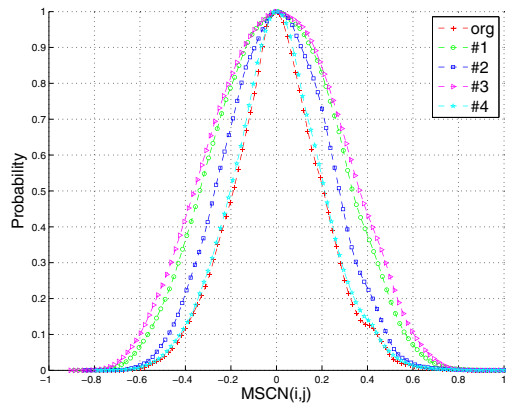**Fig. 3** The empirical distributions of different types of distortions on spatial domain and DoG bands

(a) The empirical distributions on spatial domain

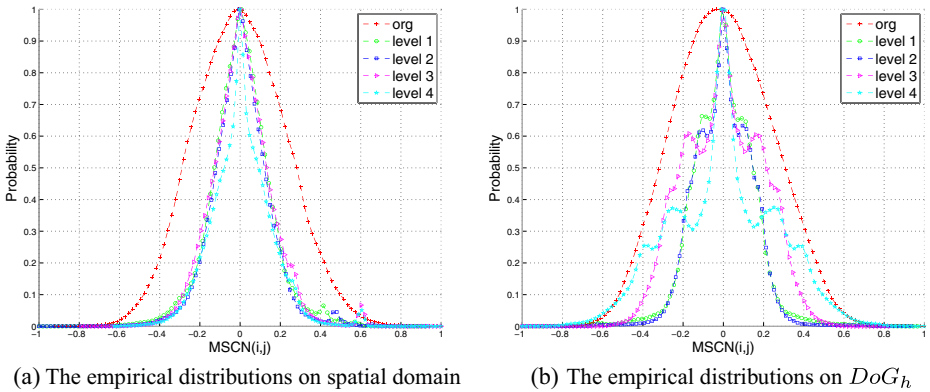(b) The empirical distributions on DoG_h

Fig. 4 The empirical distributions of JPEG 2K compressed images with four compress rates

In our proposed method, we utilize the DoG model to overcome these drawbacks. Images are decomposed from spatial domain to several DoG bands. Then the quality-aware features are extracted from all DoG bands to simulate the multiple strategies of HVS. The imperfection in single spatial domain can be solved by using complementary information in multiple DoG bands. More concretely, for the drawback **a)**, the indistinguishable curves (#2 and #3, #1 and #4) can be easily discriminated in DoG_l as shown in Fig. 3h. Although there still exist some overlaps in Fig. 3h, the overlap parts in DoG_l are separated in DoG_h as shown in Fig. 3g. On the other hand, with regard to drawback **b)**, some image distortions influence the high-frequency component of images. The high-frequency degradations are always masked by the low-frequency components of images in spatial domain. However, the masked high-frequency degradations are obviously distinguishable in DoG_h. Figure 4b shows that the statistical curves which heavily overlapped in spatial domain (Fig. 4a) are easily discriminable in DoG_h. In summary, by utilizing the complementary information in DoG_h and DoG_l, the above mentioned drawbacks can be overcome and a more effective quality-aware feature for IQA will be obtained.

3.3 Model regression and prediction

The quality-aware features f are extracted from DoG_h and DoG_l separately, denoted as f_{DoG_h} and f_{DoG_l}. Then, they are combined as $F = [f_{DoG_h}, f_{DoG_l}]$. The training features $F_{train} = [F_1; ...; F_{tn}]$ are extracted from the training dataset, where tn is the number of images in the training dataset. Finally, F_{train} and associated subjective opinion scores are fed into SVR to train the regression model.

The LIBSVM package [3] is used to implement the SVR with a radial basis function (RBF) kernel. To predict the quality of a test image, the feature F_{test} is first extracted. Then, the learned regression model is applied to map F_{test} to the predicted quality score.

4 Experimental evaluation

To evaluate the performance of our proposed method, we conduct thorough experiments on six popular IQA databases: TID2013 [26], TID [25], LIVE Multiply Distorted [14], LIVE [33], CSIQ [15] and BID [4]. The first five databases consist of reference images,

distorted images and associated subjective opinion scores. The BID database consists of realistic camera images containing many categories of blur distortions, and thus it only has distorted images and associated subjective opinion scores. The subjective opinion scores are generally given by two forms: mean opinion score (MOS) and difference mean opinion score (DMOS). The information regarding the distorted image content and subjective opinion scores of these datasets is summarized in Table 2. It is worth noting that LIVE Multiply Distorted (MD) IQA database consists of two parts which we denote by LIVE MD1 and LIVE MD2 [36]. Images in LIVE MD1 are distorted by blur followed by JPEG, and images in LIVE MD2 are distorted by blur followed by noise.

Our proposed CS-BIQA metric is compared with several state-of-the-art NR-IQA metrics, including BLIIND-II [28], BRISQUE [20], NIQE [21], IL-NIQE [36], and the existing DoG-based IQA metrics FRIQUEE [6] and 3D-IQA [30] are also compared. Since the source code is unreleased, the 3D-IQA metric method is implemented by ourself. Moreover, only DoG-based features in FRIQUEE, denoted as FRI-DoG for short, are used to evaluate the quality for fair comparison. The source codes of other metrics are downloaded from the websites offered in the literatures.

Three commonly used metrics are employed to evaluate the performance of all IQA metrics, including the Pearson linear correlation coefficient (PLCC), the Spearman's rank-order correlation coefficient (SRCC) and the Kendall rank-order correlation coefficient (KRCC). Among these performance metrics, PLCC is adopted to evaluate the prediction accuracy. SRCC and KRCC are utilized to assess the prediction monotonicity. Before calculating PLCC between objective scores and subjective opinion scores, we adopt the regression analysis approach which recommended by video quality experts group. More concretely, we choose a logistic mapping function with five-parameters:

$$f(x) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + exp(\beta_2(x - \beta_3))} \right) + \beta_4 x + \beta_5 \tag{24}$$

where x denotes the raw objective score which computed by the IQA metric, $f(x)$ denotes the mapped score for the performance evaluation, and β_i, $i = 1, 2, ..., 5$ are the parameters to be fitted. A better IQA metric is expected to obtain higher values of PLCC, SRCC and KRCC.

4.1 Experiments on individual databases

We first evaluate the overall performance on each individual database. Each database is separated into two groups: the distorted images associated with 80% of the reference images

Table 2 Benchmark IQA databases used to evaluate IQA metrics

Database	Reference Image No.	Distorted Image No.	Distortion Types No.	Contains multiply-distortions
TID2013	25	3000	24	YES
TID	25	1700	17	NO
CSIQ	30	866	6	NO
LIVE	29	779	5	NO
LIVE MD1	15	255	1	YES
LIVE MD2	15	255	1	YES
BID	0	585	Various blurring	YES

are used for training and the remaining 20% images are used for testing. Since NIQE and IL-NIQE have specific training datasets, which are different from these commonly used databases. So we only report their performance on the partitioned testing subset to promise the fairness. Considering the influence of training subsets, each database is randomly partitioned and repeat 1000 times. These random subsets are used to remove the impact of training subsets. The median values of the PLCC, SRCC and KRCC are shown in Table 3. Moreover, the average results of all databases are listed too, and the weighted average results are weighted by the size of each database. For each criterion, the best two are highlighted in boldface.

From the observation of Table 3, we can draw the following conclusions. Firstly, on the most of databases and the average levels, the results of CS-BIQA always lie in the best two positions. This demonstrates that the proposed IQA metric can effectively predict the quality of images in the databases. Secondly, on the databases which contained more types of distortions (e.g. TID2013), the advantage of CS-BIQA is more obvious. This demonstrates

Table 3 The performance evaluation on each individual database, in which the top two of results are highlighted in boldface

		BLIIND-II	BRISQUE	NIQE	IL-NIQE	FRI-DoG	3D-IQA	CS-BIQA
TID2013	PLCC	0.6040	**0.6688**	0.3752	0.5746	0.4941	0.5831	**0.7428**
	SRCC	0.4733	**0.5814**	0.3128	0.4944	0.3292	0.5005	**0.6705**
	KRCC	0.3348	**0.4215**	0.2129	0.3494	0.2331	0.3481	**0.4909**
TID	PLCC	0.6703	**0.6881**	0.3114	0.5614	0.4839	0.5750	**0.7527**
	SRCC	0.5569	**0.6240**	0.2429	0.4593	0.2873	0.5068	**0.6955**
	KRCC	0.4023	**0.4559**	0.1672	0.3256	0.2040	0.3559	**0.5150**
CSIQ	PLCC	0.7989	0.7697	0.7213	**0.8570**	0.5492	0.6522	**0.8491**
	SRCC	0.7303	0.6947	0.6278	**0.8120**	0.4303	0.5909	**0.7894**
	KRCC	0.5483	0.5194	0.4627	**0.6224**	0.3052	0.4250	**0.6067**
LIVE	PLCC	0.9048	0.8872	**0.9081**	0.9050	0.5893	0.8161	**0.9236**
	SRCC	0.8992	0.8775	**0.9039**	0.8964	0.4213	0.8082	**0.9228**
	KRCC	**0.7289**	0.7015	0.7274	0.7150	0.3195	0.6100	**0.7714**
LIVE MD1	PLCC	**0.9374**	0.9144	0.9154	0.9157	0.8599	0.8960	**0.9247**
	SRCC	**0.8956**	0.8809	0.8666	0.8860	0.8150	0.8709	**0.9010**
	KRCC	**0.7280**	0.7111	0.6879	0.7051	0.6051	0.6828	**0.7333**
LIVE MD2	PLCC	0.8992	0.9027	0.8555	**0.9045**	0.8649	0.8315	**0.9097**
	SRCC	**0.8798**	0.8770	0.7923	0.8760	0.8520	0.8082	**0.8901**
	KRCC	**0.7051**	0.7038	0.6094	0.6980	0.6673	0.6040	**0.7152**
BID	PLCC	0.5229	**0.6071**	0.4511	0.4836	0.5343	0.5266	**0.6738**
	SRCC	0.5028	**0.5807**	0.4552	0.4935	0.5155	0.5220	**0.6494**
	KRCC	0.3474	**0.4058**	0.3083	0.3422	0.3491	0.3646	**0.4620**
Direct average	PLCC	0.7625	**0.7769**	0.6481	0.7431	0.6251	0.6972	**0.8252**
	SRCC	0.7054	**0.7309**	0.6004	0.7026	0.5215	0.6582	**0.7884**
	KRCC	0.5421	**0.5599**	0.4539	0.5368	0.3833	0.4843	**0.6135**
Weighted average	PLCC	0.6885	**0.7194**	0.4975	0.6549	0.5366	0.6285	**0.7829**
	SRCC	0.5976	**0.6557**	0.4421	0.5919	0.3903	0.5696	**0.7302**
	KRCC	0.4435	**0.4884**	0.3229	0.4376	0.2807	0.4078	**0.5530**

that the CS-BIQA metric can handle more complicated situations than the other competing metrics. Finally, the performance of CS-BIQA is higher than the FRI-DoG and 3D-IQA, this demonstrates that, our proposed features extracted on the all DoG bands of pristine images are more effectively than the DoG-based features proposed in FRI-DoG and 3D-IQA.

Moreover, the box plot of SRCC values of competing metrics on TID2013 is shown in Fig. 5. Box plot shows the median value and the distribution range of SRCC values. We can observe that CS-BIQA has the highest median value and moderate stability. In addition, the lower bound of CS-BIQA is higher than the upper bounds of other competing methods.

4.2 Experiments on each type of distortion

In this experiment, the SRCC values of all IQA metrics on each individual type of distortion in TID2013 are reported in Table 4. There are 24 types of distortions in TID2013, and they are Additive Gaussian Noise (#1), Additive Noise in Color Components (#2), Spatially Correlated Noise (#3), Masked Noise (#4), High Frequency Noise (#5), Impulse Noise (#6), Quantization Noise (#7), Gaussian Blur (#8), Image Denoising (#9), JPEG Compression (#10), JPEG2000 Compression (#11), JPEG Transmission Errors (#12), JPEG2000 Transmission Errors (#13), Non Eccentricity Pattern Noise (#14), Local Block-wise Distortions (#15), Mean Shift (#16), Contrast Change (#17), Change of Color Saturation (#18), Multiplicative Gaussian Noise (#19), Comfort Noise (#20), Lossy Compression of Noise Images (#21), Color Quantization with Dither (#22), Chromatic Aberrations (#23), and Sparse Sampling and Reconstruction (#24). The entire TID database is employed as the training dataset.
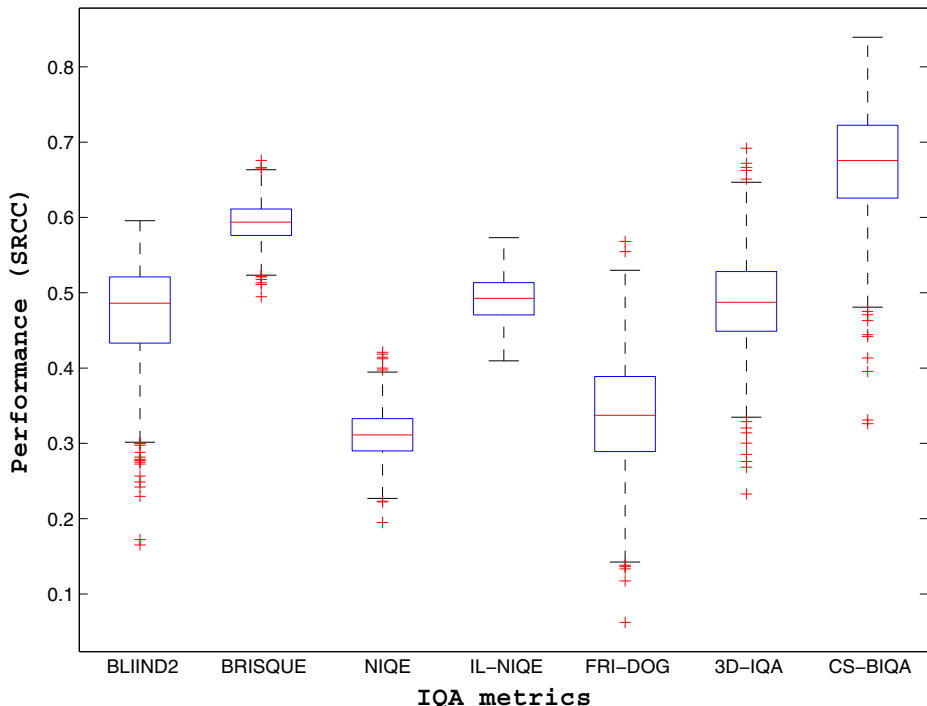

Fig. 5 Boxplot of SRCC values on TID2013

Table 4 Performances on individual distortion types in TID2013, in which the top two of results are highlighted in boldface

Distortion type	BLIIND-II	BRISQUE	NIQE	IL-NIQE	FRI-DoG	3D-IQA	CS-BIQA
#1	0.7226	0.8523	0.8194	**0.8760**	0.5609	0.4037	**0.8727**
#2	0.6497	**0.7090**	0.6699	**0.8159**	0.2580	0.0978	0.6584
#3	0.7674	0.4908	0.6660	**0.9233**	0.5457	0.6304	**0.8690**
#4	0.5127	0.5748	**0.7464**	0.5120	0.1366	0.2534	**0.5814**
#5	0.8245	0.7528	0.8449	**0.8685**	0.8275	0.7564	**0.9300**
#6	0.6501	0.6299	0.7434	**0.7551**	**0.8604**	0.5769	0.6377
#7	0.7816	0.7984	**0.8500**	**0.8730**	0.2675	0.4971	0.7656
#8	**0.8557**	0.8134	0.7954	0.8142	0.6281	0.8775	**0.9315**
#9	0.7116	0.5864	0.5903	**0.7500**	0.4800	0.7361	**0.8409**
#10	**0.8643**	**0.8521**	0.8402	0.8349	0.1292	**0.8521**	0.7881
#11	**0.8984**	0.8925	0.8891	0.8578	0.6488	0.8721	**0.9347**
#12	0.1170	**0.3150**	0.0028	0.2827	0.0381	0.6419	**0.8559**
#13	0.6209	0.3594	0.5102	0.5248	0.6116	**0.7199**	**0.8506**
#14	0.0968	**0.1453**	0.0698	0.0805	0.1074	0.2011	**0.3864**
#15	0.2098	**0.2235**	0.1269	0.1357	0.0137	0.0643	**0.6841**
#16	0.1284	0.1241	0.1626	**0.1845**	0.0444	0.0639	**0.3822**
#17	0.1505	0.0403	0.0180	0.0141	0.4245	**0.6645**	**0.7076**
#18	0.0178	0.1093	**0.2460**	0.1628	0.1259	**0.1981**	0.0595
#19	0.7165	**0.7242**	0.6940	0.6932	0.4555	0.3011	**0.8006**
#20	0.0178	0.0081	0.1548	0.3599	**0.3879**	**0.5996**	0.2213
#21	0.7193	0.6852	**0.8011**	**0.8287**	0.5342	0.7079	0.4920
#22	0.7358	**0.7640**	**0.7832**	0.7487	0.0937	0.7411	0.6030
#23	0.5397	0.6160	0.5612	0.6793	**0.7545**	0.6949	**0.8005**
#24	0.8164	0.7841	**0.8341**	**0.8650**	0.6266	0.7537	0.8009
Hit count	3	7	6	10	3	5	15

For each individual type of distortion, the best two results are highlighted in boldface. In addition, we report the hit count (the number of times ranked in the top 2 for each distortion type [35]) of each metrics.

From the results of Table 4, we can draw the following conclusions. Firstly, CS-BIQA has the highest hit count, which verifies again that our proposed method can handle most types of distortions than the other competing metrics. Secondly, our proposed method can evaluate some specific distortions that the other competing metrics are nearly disabled, for example, JPEG transmission errors (#12) and contrast change (#17).

4.3 Database independence

To evaluate the database independence of CS-BIQA metric, we train the metric on LIVE dataset and then test on TID2013, TID, CSIQ, LIVE MD1 and LIVE MD2 databases. The SRCC values are tabulated in Table 5. For each criterion, the best two results are highlighted in boldface.

Table 5 The evaluation of database independence, in which the top two of results are highlighted in boldface

		BLIIND-II	BRISQUE	NIQE	IL-NIQE	FRI-DoG	3D-IQA	CS-BIQA
	PLCC	0.4973	0.5010	0.3991	**0.5884**	0.4327	0.1430	**0.5879**
TID2013	SRCC	0.3767	0.4297	0.3120	**0.4938**	0.2861	0.1205	**0.5002**
	KRCC	0.2620	0.2982	0.2122	**0.3491**	0.1968	0.0790	**0.3523**
	PLCC	0.5341	0.4886	0.3247	**0.5642**	0.3979	0.0637	**0.5779**
TID	SRCC	0.3383	0.3966	0.2437	**0.4631**	0.2778	0.0531	**0.4900**
	KRCC	0.2399	0.2787	0.1672	**0.3282**	0.1925	0.0358	**0.3455**
	PLCC	0.7425	0.7275	0.7158	**0.8537**	0.5212	0.0069	**0.7490**
CSIQ	SRCC	0.6086	0.5311	0.6268	**0.8144**	0.3814	0.0363	**0.6545**
	KRCC	0.4510	0.3995	0.4608	**0.6222**	0.2684	0.0247	**0.4782**
	PLCC	0.8757	0.8748	**0.9094**	**0.9045**	0.8252	0.5327	0.9006
LIVE	SRCC	0.8205	0.8180	**0.8707**	**0.8912**	0.7722	0.3339	0.8565
MD1	KRCC	0.6357	0.6242	**0.6829**	**0.7019**	0.5667	0.2347	0.6705
	PLCC	0.3651	0.3430	0.8483	**0.8969**	0.6293	0.0042	**0.8552**
LIVE	SRCC	0.1913	0.0647	0.7946	**0.8824**	0.6518	0.0347	**0.8052**
MD2	KRCC	0.1315	0.0413	0.6058	**0.6958**	0.4708	0.0234	**0.6085**

From Table 5, we can draw the following conclusions. Firstly, OU-IQA methods are trained on the specific training datasets that are different from the commonly used IQA databases. Thus the database independence should be the largest advantage of OU-IQA methods. Even so, our proposed OA-IQA demonstrates competitive database independence with IL-NIQE metric which is one of the most advanced OU-IQA methods. Secondly, LIVE database only have five types of distortions, and the reference images in LIVE database are different from the test database in this experiment. Nevertheless, our proposed method still keeps competitive performance. So, it is shown that the CS-BIQA metric has well generalization capability.

4.4 Statistical significance and hypothesis testing

We evaluate the statistical significance of the performance of competing metrics. The hypothesis based on paired t-test on the 100 SRCC values which are obtained from 100 train-test trials [7]. The results are shown in Table 6. A value of '0' means that the row and

Table 6 Hypothesis Testing on TID2013

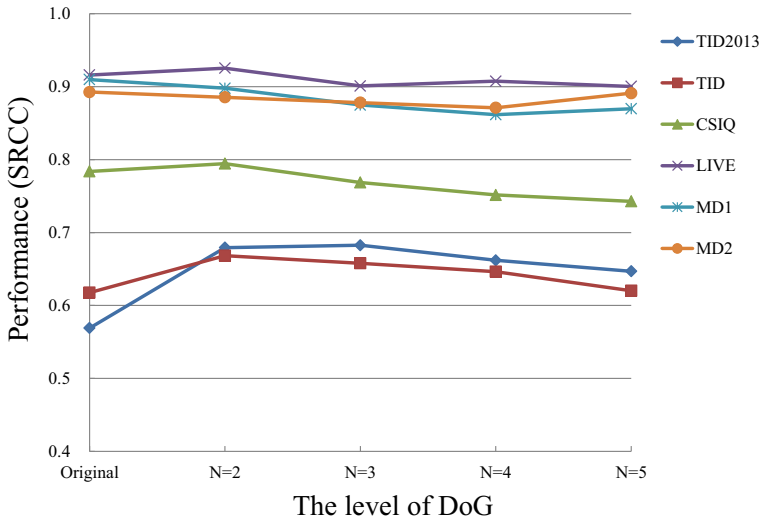|	BLIIND-II	BRISQUE	NIQE	IL-NIQE	FRI-DoG	3D-IQA	CS-BIQA
BLIIND-II	0	−1	1	−1	1	−1	−1
BRISQUE	1	0	1	1	1	1	−1
NIQE	−1	−1	0	−1	−1	−1	−1
IL-NIQE	1	−1	1	0	1	−1	−1
FRI-DoG	−1	−1	1	−1	0	−1	−1
3D-IQA	1	−1	1	1	1	0	−1
CS-BIQA	1	1	1	1	1	1	0

Fig. 6 The impact of DoG levels

column are statistically indistinguishable. A value of "1" ("-1") indicates that the row is statistically superior (worse) than the column. Table 6 demonstrates that our proposed method is statistically superior to all competing metrics.

4.5 The influence of DoG model

To analyze the influences induced by DoG decomposition, two experiments are conducted. Firstly, we evaluate the influence by the level of DoG model. The performance of CS-BIQA on $2 \sim 5$ level DoG is shown in Fig. 6. N is the level of DoG, and the SRCC values are calculated on TID2013. We can observe that the values of N show little impact on the quality-predicted results. So we choose $N = 2$ in our proposed method.

Secondly, we evaluate the performance of features on each DoG band versus the comprehensive feature. The performances on two DoG bands are tabulated in Table 7. We can observe that the features extracted on high-frequency band dominant the performance on most databases. However, the performance becomes more robustness by combining the features on all DoG bands.

Table 7 Performance on each DoG band versus the comprehensive scores of DoG bands

Database	DOG_h	DOG_l	CS-BIQA
TID2013	0.6546	0.5287	0.6628
TID	0.6734	0.4784	0.6785
CSIQ	0.8299	0.6676	0.8338
LIVE	0.9223	0.8824	0.9247
LIVE MD1	0.8843	0.8467	0.8930
LIVE MD2	0.8546	0.8644	0.8728
BID	0.6175	0.5899	0.6420

Table 8 Computation times of all metrics on LIVE database

Metric	Time Cost (seconds)
BIQI	1.1337
BLIIND-II	62.9794
BRISQUE	0.2671
NIQE	0.4535
IL-NIQE	4.9286
FRIQUEE	35.3601
FRI-DoG	0.1370
3D-IQA	13.6315
CS-BIQA	0.2751

4.6 Computation time

The computation time is a crucial attribute which decides the practicability of IQA algorithms. We compare the computation times of all competing metrics in Table 8. The experiments are conducted under a single PC with an Intel(R) Core(TM) i5-4570 CPU of 3.20GHz and 8GB memory. All metrics are implemented with MATLAB 2014a. We record the averaged computation times of all images in LIVE database. From Table 8, we can see that the CS-BIQA has a competitive computation time. This guarantees the practicability of our proposed method. Also, we can observe that the computation time of 3D-IQA metric and entire FRIQUEE metric is quite long for practical applications.

5 Conclusion

In this paper, we proposed a novel NR-IQA metric leveraging the DoG model. We firstly analyzed the frequency properties of image distortions which affect the prediction of image perceptual quality. Then the statistical features were extracted on each DoG band to represent the properties of distortions. Subsequently, SVR was employed to train the regression models between the integrated features and subjective opinion scores and predict the objective quality scores of test images. Experiments on several IQA databases demonstrated that our proposed method has competing performance with state-of-the-art NR-IQA metrics. In the proposed method, we adopted commonly used GGD and AGGD methods to fit MSCN coefficients of DoG bands. In future work, we plan to use other statistical models to fit the coefficients of DoG bands, which can fit the empirical distributions more appropriately.

References

1. Adelson EH (2000) Lightness perception and lightness illusions, chap 24. MIT, Cambridge, pp 339–351

2. Bovik AC (2013) Automatic prediction of perceptual image and video quality. Proc. IEEE 101(9):2008–2024

3. Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2:27:1–27:27. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

4. Ciancio A, Da CA, Da SE, Said A, Samadani R, Obrador P (2011) No-reference blur assessment of digital pictures based on multifeature classifiers. IEEE Trans Image Process 20(1):64

5. Fu Y, Wang S (2016) A no reference image quality assessment metric based on visual perception. Algorithms 9(4)

6. Ghadiyaram D, Bovik AC (2015) Feature maps driven no-reference image quality prediction of authentically distorted images. Proc SPIE - Int Soc Optl Eng 9394:93,940J–93,940J–14

7. Ghadiyaram D, Bovik AC (2016) Perceptual quality prediction on authentically distorted images using a bag of features approach. J Vis 17(1)

8. Gu K, Liu M, Zhai G, Yang X (2015) Quality assessment considering viewing distance and image resolution. IEEE Trans Broadcast 61(3):520–531

9. Gu K, Zhai G, Yang X, Zhang W (2015) Using free energy principle for blind image quality assessment. IEEE Trans Multimed 17(1):50–63

10. Gu K, Lin W, Zhai G, Yang X, Zhang W, Chang W (2016) No-reference quality metric of contrast-distorted images based on information maximization. IEEE Trans Cybern 1–7

11. Gu K, Zhai G, Lin W, Liu M (2016) The analysis of image contrast: from quality assessment to automatic enhancement. IEEE Trans Cybern 46(1):284

12. Gu K, Li L, Lu H, Xin X, Lin W (2017) A fast reliable image quality predictor by fusing micro- and macro-structures. IEEE Trans Indus Electron

13. Gu K, Tao D, Qiao J, Lin W (2017) Learning a no-reference quality assessment model of enhanced images with big data. IEEE Trans Neural Netw Learn Syst

14. Jayaraman D, Mittal A, Moorthy AK, Bovik AC (2012) Objective quality assessment of multiply distorted images. In: Conference record of the forty sixth asilomar conference on signals, systems and computers, pp 1693–1697

15. Larson EC, Chandler DM (2010) Most apparent distortion: full-reference image quality assessment and the role of strategy. J Electron Imag 19(1):143–153

16. Li L, Zhou Y, Lin W, Wu J, Zhang X, Chen B (2016) No-reference quality assessment of deblocked images. Neurocomputing 177(C):572–584

17. Lin YH, Wu JL (2014) Quality assessment of stereoscopic 3d image compression by binocular integration behaviors. IEEE Trans Image Process 23(4):1527–1542

18. Liu H, Klomp N, Heynderickx I (2010) A no-reference metric for perceived ringing artifacts in images. IEEE Trans Circ Syst Vid Technol 20(4):529–539

19. Lu W, Xu T, Ren Y, He L (2016) Statistical modeling in the shearlet domain for blind image quality assessment. Multimed Tools Appl 75(22):1–15

20. Mittal A, Moorthy AK, Bovik AC (2012) No-reference image quality assessment in the spatial domain. IEEE Trans Image Process 21(12):4695–4708

21. Mittal A, Soundararajan R, Bovik AC (2013) Making a "completely blind" image quality analyzer. IEEE Signal Process Lett 20(3):209–212

22. Moorthy AK, Bovik AC (2010) A two-step framework for constructing blind image quality indices. IEEE Signal Process Lett 17(5):513–516

23. Moorthy AK, Bovik AC (2011) Blind image quality assessment: from natural scene statistics to perceptual quality. IEEE Trans Image Process 20(12):3350–64

24. Pei SC, Chen LH (2015) Image quality assessment using human visual dog model fused with random forest. IEEE Trans Image Process 24(11):3282–92

25. Ponomarenko N, Lukin V, Zelensky A, Egiazarian K, Carli M, Battisti F (2004) TID2008 - a database for evaluation of full-reference visual quality assessment metrics. Adv Modern Radioelectron 10:30–45

26. Ponomarenko N, Leremeiev O, Lukin V, Egiazarian K, Jin L, Astola J, Vozel B, Chehdi K, Carli M, Battisti F (2013) Color image database TID2013: peculiarities and preliminary results. In: European workshop on visual information processing, pp 106–111

27. Ruderman DL (2009) The statistics of natural images. Netw Comput Neural Syst 5(4):517–548

28. Saad MA, Bovik AC, Charrier C (2010) A DCT statistics-based blind image quality index. IEEE Signal Process Lett 17(6):583–586

29. Saad MA, Bovik AC, Charrier C (2012) Blind image quality assessment: a natural scene statistics approach in the DCT domain. IEEE Trans Image Process 21(8):3339–52

30. Shao F, Li K, Lin W, Jiang G, Yu M (2015) Using binocular feature combination for blind quality assessment of stereoscopic images. IEEE Signal Process Lett 22(10):1548–1551

31. Sharifi K, Leongarcia A (1995) Estimation of shape parameter for generalized gaussian distributions in subband decompositions of video. IEEE Trans Circ Syst Vid Technol 5(1):52–56
32. Tang L, Li Q, Li L, Gu K, Qian J (2017) Training-free referenceless camera image blur assessment via hypercomplex singular value decomposition. Multimed Tools Appl 1–22
33. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612
34. Wang S, Ma K, Yeganeh H, Wang Z (2015) A patch-structure representation method for quality assessment of contrast changed images. IEEE Signal Process Lett 22(12):2387–2390
35. Xu J, Ye P, Li Q, Du H (2016) Blind image quality assessment based on high order statistics aggregation. IEEE Trans Image Process 25(9):4444–4457
36. Zhang L, Zhang L, Bovik AC (2015) A feature-enriched completely blind image quality evaluator. IEEE Trans Image Process 24(8):2579–2591

Jun Wu received the M.S. degree from the School of Electronics and Information, Northwestern Polytechnical University, China, in 2013. He is currently pursuing his Ph.D. degree in Northwestern Polytechnical University. His research interests include image quality assessment and machine learning.

Zhaoqiang Xia is with School of Electronics and Information in Northwestern Polytechnical University. His research interests include multimedia computing, machine learning and computer vision.

Huifang Li was born in 1962. He received the B.S. degree in electronic science and technology from Lanzhou University, Lanzhou, China, in 1984 and M.S. and Ph.D. degrees in electronic science and technology from Northwestern Polytechnical University, Xi'an, China, in, 1990, and 2004, respectively. Since 1990, he has been a faculty member in Northwestern Polyte-chnical University, Xi'an, China, where he is currently a professor of school of electronic and information. He was a visiting scholar with the University of North Carolina at Charlotte, Charlotte, NC, USA, from 2010 to 2011. His current research interests include topics related to image detection and recognition, multimedia information process, quantum information processing and etc.
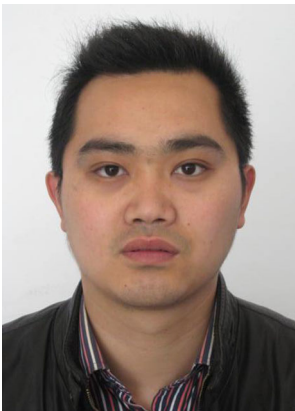
Kezheng Sun received the B.E. degree from Jiangsu Normal University, China, in 2005, and the master's degree with Jiangsu University, China, in 2010. His research interests include image quality assessment and visual perceptual modeling.

Dr. Ke Gu received the B.S. and PhD degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009 and 2015. He is an associated editor for IEEE Access, and the reviewer for IEEE T-NNLS, T-IP, T-MM, T-CSVT, T-CYB, T-IE, T-BC, J-STSP, SPL, Access, Information Sciences, Neurocomputing, MTAP, SPIC, JVCI, DSP, ELL, ICME, ICIP, VCIP, etc. He has reviewed more than 50 journals each year. Dr. Gu has authored or coauthored papers appearing in over 80 international conference and journal publications, including the IEEE T-NNLS, T-IP, T-CSVT, T-MM, T-CYB, T-IE, T-BC, SPL, Access, ICME, ICIP, etc. He received the Best Paper Award at the IEEE International Conference on Multimedia and Expo (ICME) 2016. Dr. Gu is the leading/first special session organizer in VCIP 2016 and ICIP2017.

Hong Lu received the Ph.D. degree in control theory and control engineering from the School of Automation, and the M.S. degree from the School of electrical engineering, Southeast Uni-versity, Nanjing, China, in 2009 and 2003, respectively. She is currently an Associate Professor with the School of Automation, Nanjing Institute of Technology, Nanjing, China. She is also a Visiting Scholar with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. Her current research interests include object detection, object tracking, and image and video data mining.