

Blind image quality assessment in the contourlet domain[☆]Chaofeng Li^{a,*}, Tuxin Guan^a, Yuhui Zheng^b, Xiaochun Zhong^c, Xiaojun Wu^c, Alan Bovik^d^a Institute of Logistics Science and Engineering, Shanghai Maritime University, Shanghai, 200135, China^b College of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, China^c School of Internet of Things Engineering, Jiangnan University, Wuxi, 214122, China^d Laboratory for Image and Video Engineering (LIVE), The University of Texas at Austin, Austin, TX 78712, USA

ARTICLE INFO

Keywords:

No-reference image quality assessment

Contourlet transformation

CIELAB color space

Gradient magnitude image

Support vector regression

ABSTRACT

No-reference/blind image quality assessment (NR-IQA/BIQA) algorithms play an important role in image evaluation, as they can assess the quality of an image automatically, only using the distorted image whose quality is being assessed. Among the existing NR-IQA/BIQA methods, natural scene statistic (NSS) models which can be expressed in different bandpass domains show good consistency with human subjective judgments of quality.

In this paper, we create new ‘quality-aware’ features: the energy differences of the sub-band coefficients across scales via contourlet transform, and propose a new NR-IQA/BIQA model that operates on natural scene statistics in the contourlet domain. Prior to applying the contourlet transform, we apply two preprocessing steps that help to create more information-dense, low-entropy representations. Specifically, we transform the picture into the CIELAB color space and gradient magnitude map. Then, a number of ‘quality-aware’ features are discovered in the contourlet transform domain: the energy of the sub-band coefficients within scales, and the energy differences between scales, as well as measurements of the statistical relationships of pixels across scales. A detailed analysis is conducted to show how different distortions affect the statistical characteristics of these features, and then features are fed to a support vector regression (SVR) model which learns to predict image quality. Experimental results show that the proposed method has high linearity against human subjective perception, and outperforms the state-of-the-art NR-IQA models.

1. Introduction

In recent years, massive numbers of digital images are captured, transmitted, stored and compressed by pervasive mobile/camera devices. All of these processes modify the obtained images and may introduce annoying visual distortions that reduce user satisfaction. Therefore, assessing image quality automatically to predict, preclude or ameliorate visual distortions has become a hot issue. Generally, image quality assessment can be divided into subjective and objective quality evaluation. As the ultimate receivers, humans are the final arbiters of visual quality. However, it is impractical to solicit image quality opinions in applications. As a result, objective quality evaluation methods are now an intensive research focus.

Objective image quality assessment methods generally can be classified into three kinds depending on the amount of information provided to the algorithm: full-reference (FR), reduced-reference (RR) and no-reference (NR) algorithms. NR IQA refers to automatic quality assessment of an image only using the distorted image whose quality is being assessed. FR IQA and RR IQA algorithms require full or partial

information regarding reference image. In most cases, it is hard to obtain any reference image information. In particular, it is desirable that digital cameras can automatically control perceptual image quality during image capture. Hence NR IQA models and algorithms are in high demand.

Current NR IQA models generally fall in one of two categories: distortion-specific NR IQA and ‘universal’ NR IQA. The first category aims to for IQA metrics that are ‘expert’ at predicting the perceptual severity of specific types of distortions [1], while the latter are methods suitable for distortion-agnostic scenarios.

In recent years, research on NR IQA algorithms has made great progress, and most of which are learning-based methods. Generally speaking, all learning-based methods follow a two-step process of ‘quality-aware’ feature extraction and model regression on human subjective scores. Both BIQI [2] and DIIVINE [3] methods used a 2-stage framework of distortion identification and distortion-specific NR quality assessment, first using a support vector machine (SVM) to predict the image distortion type, then a SVR specified to that

[☆] This work was supported by the National Natural Science Foundation of China (No. 61771223).

* Corresponding author.

E-mail address: wxlichao@126.com (C. Li).

distortion to conduct NR IQA. The authors of [4] developed a DIIVINE-GGSM model for better modeling the wavelet coefficients of both reference and distorted images and applying in NR IQA. The LBIQ metric [5] trained three regression model for three types of features, namely natural scene statistics, distorted textures and blur/noise, and finally the image quality is estimated via a weighted method. The authors of [6] developed a IQA index BLIINDS-II, which analyzed DCT coefficients of local image, and used a generalized parametric model of the natural statistics for predicting image quality. The authors of [7] proposed a generalized regression neural network-based NR IQA algorithm, which four types of perceptually relevant features, including the entropy of the natural image, the gradient of the natural image, the mean gray value of the phase congruency image, and the entropy of the phase congruency image, were extracted to construct image quality assessment model. The authors of BRISQUE [8] index utilized the empirical distribution of locally normalized luminance, extracted spatial natural scene statistic (NSS) via locally normalized luminance, and used a SVR model to train and predict image quality. In [9], Gao et al. established a NR image quality evaluation model NSS-GS using non-Gaussianity, local dependency and exponential decay characteristic via multiple kernel learning. Liu et al. [10] used features of an NSS model in curvelet domain within a 2-stage framework, calculate the maxima of the log-histograms and energy distribution of curvelet coefficients, then use a SVR for quality prediction. More recently, Liu et al. [11] deployed a relative gradient magnitude feature which accounts for perceptual masking, and used an adaboosting BP neural network to predict image quality scores. The authors of [12] used the speeded-up robust feature to detect and describe keypoints of images for characterizing the distortion of natural image, and statistics of these feature descriptors to obtain the final score. Gao et al. [13] employed statistical features of the image to describe the image local contrast, structure, multi-scale and multi-direction properties, include and mean subtracted and contrast normalized (MSCN) features, gradient magnitude (GM) features and Laplacian of Gaussian (LOG) features, and used a SVR to map between the statistical features and the human subjective perception. Gu et al. [14] used a vector regression method to obtain a vector of image belief scores, then used an object-oriented pooling method to produce quality predictions. Liu et al. [15] used pre-attention theory to simulate human visual perception, by extracting spatial-dependency features from textures on both gray and color channels, and using them to learn and predict quality scores. In the NR-IQA methods, FR-IQA models can be successfully used as weak supervised indicators to make the regressor more robust, e.g. [16,17] and [18], the authors of which have utilized big data technology for natural, screen content and enhanced image to extract perceptual features followed by using FR IQMs to obtain the label of training data for conducting more reliable quality regression.

With the rapid development of deep learning, it is inevitable that a variety of deep learning-based NR IQA methods have been proposed. For example, Hou et al. [19] proposed a NR IQA model that learns NSS features using a deep learning model. The authors of [20] extracted deep perceptual features using a convolution neural network to conduct quality prediction. The authors of [21] proposed a deep-learning approach (DIQM) to learn global image quality. Yan et al. [22] designed a two-stream convolution network consisting of an image and a gradient stream, conveying complementary information to a region-based fully convolutional layer which predicts quality scores. Bosse et al. [23] proposed a novel network structure that can conduct both FR IQA and NR IQA tasks with slight adaptations, by joint learning of local quality and local weights. Lin et al. [24] proposed a deep-learning network, which generates a hallucinated reference image, then learns the perceptual difference between the hallucinated image and the distorted image to obtain the quality scores. Pan et al. [25] designed an FCNN architecture to predict a dense quality map, using intermediate similarity maps derived from an FR IQA model. HFD-BIQA [26] method combined the orientation selectivity based low-level visual features and high-level visual features extracted from deep residual network to represent

quality degradation. Although these deep learning based methods have delivered improved performance on the NR IQA problem, the sites of available human-rated datasets have limited the performance of these deep-learning methods. There still remains considerable room for improvement of objective IQA methods.

In previous, several researchers focused on the contourlet transform to establish a computational model to predict the objective scores of image. Liu et al. [27] proposed a FR-IQA method by considering the contrast-masking and entropy-masking effect jointly after the contourlet transform, then using an error pooling method to obtain objective scores. Lu et al. [28] applied CSF masking on each contourlet subband to obtain the sensitive threshold, and then obtained quality scores by pooling the difference between visual sensitive coefficients of original and distortion images. Lu et al. [29] proposed a NR-IQA method which uses the joint statistics distribution of contourlet coefficients, and then an image-dependent threshold is adopted to reduce the effect of content to the statistical model.

In the approach we take here, we study the distributions of the sub-band coefficients of contourlet transforms of image luminances, chrominances and gradient magnitude, and use what we learn to construct and compute three kinds of quality-aware features: the energies of sub-band coefficients within scales and the energy differences between scales, as well as Neighboring energy statistics across scales. These features are then used to predict image quality using a SVR model. We called the new blind IQA model **ContourletQA**. ContourletQA is tested on the LIVE IQA database [30], and the TID2013 database [31], and the CSIQ database [32], and the LIVE wild challenge database [33], and shown to be a highly consistent predictor of human subjective scores, outperform state-of-the-art NR IQA methods.

Our major contributions can be summarized as follows.

(1) We create new 'quality-aware' features: the energy differences of the sub-band coefficients across scales via contourlet transform. A number of 'quality-aware' features are discovered in the contourlet transform domain: the energy of the sub-band coefficients within scales, and the energy differences between scales, as well as measurements of the statistical relationships of pixels across scales.

(2) We propose a new NR-IQA/BIQA model that operates on natural scene statistics in the contourlet domain via the chromatic channel of CIELAB color space and gradient magnitude image.

(3) Experimental results show that the proposed method has high linearity against human subjective perception, and outperforms the state-of-the-art NR-IQA models.

The rest of the paper is organized as follows. In Section 2, quality-aware features constructed from contourlet energy statistics are described. Our image quality assessment model is proposed in Section 3. In Section 4, the performance of algorithm on the LIVE IQA database, the TID database, the CSIQ database and the LIVE challenge database are discussed and analyzed. Conclusions are given in Section 5.

2. Quality-aware feature discovery in the contourlet domain

2.1. CIELAB color space and gradient transform

The CIELAB color space consists of one luminance (L^*) and two chrominance (a^* and b^*) components. The nonlinear relationships between the L^* , a^* and b^* channels mimic the nonlinear responses of the L, M, and S cone cells in the retina, and are designed to uniformly quantify perceptual color differences [34]. Considering that the RGB color space cannot be converted to CIELAB color space directly, and the CIEXYZ color space is introduced to convert RGB to CIELAB color space, which is a linear transformation of RGB space, and has nothing to do with the device. Firstly, the normalized RGB space is converted to CIEXYZ space as follows [35]:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

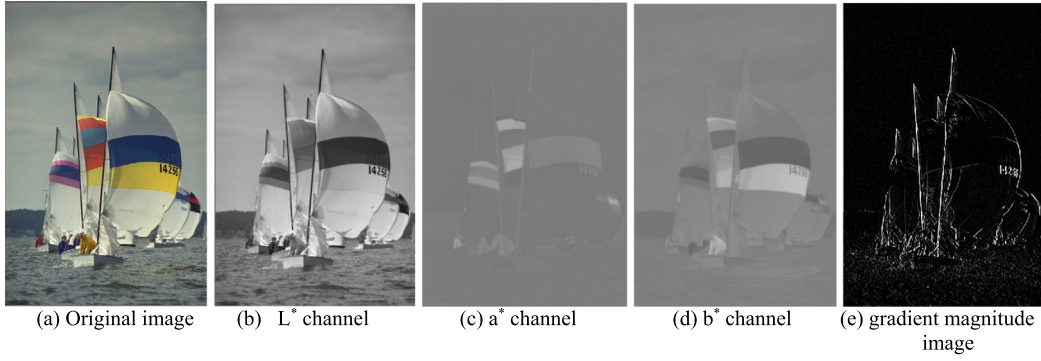


Fig. 1. An example of CIELAB color components and gradient of a natural image.

Then the CIEXYZ space is converted to the target CIELAB color space as following Eq. (2).

$$\begin{aligned} L^* &= 116f\left(\frac{Y}{Y_n}\right) - 16 \\ a^* &= 500 \left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right) \right] \\ b^* &= 200 \left[f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right) \right] \end{aligned} \quad (2)$$

wherein the reference white point (X_n, Y_n, Z_n) is the standard D65 illuminate white point, whose values are given by $X_n = 0.950456$, $Y_n = 1.000000$ and $Z_n = 1.088754$. The function $f(t)$ is calculated by (3).

$$f(t) = \begin{cases} t^{\frac{1}{3}} & t > \left(\frac{6}{29}\right)^3 \\ \frac{1}{3} \left(\frac{29}{6}\right)^2 t + \frac{4}{29} & t \leq \left(\frac{6}{29}\right)^3 \end{cases} \quad (3)$$

The gradient map reflects important structural information in natural images. Here we deploy the stabilizing Gaussian partial derivative filter to obtain the gradient map, and the formula of Gaussian partial derivative filter is shown as following.

$$h_d(x, y | \sigma) = -\frac{1}{2\pi\sigma^2} \frac{d}{\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right), d \in \{x, y\} \quad (4)$$

Then the gradient magnitude map GM is conducted as follows:

$$GM = \sqrt{G_x^2 + G_y^2} \quad (5)$$

where the G_x and G_y represent the gradient map of horizontal and vertical direction, which are calculated as following.

$$G_x = I \otimes h_x, \quad G_y = I \otimes h_y \quad (6)$$

where I represents the original image, and \otimes is the linear convolution operator.

An example of the CIELAB color components of an image and the gradient magnitude of a natural image is given in Fig. 1.

2.2. Quality-aware feature in the contourlet domain

In this paper, we use multi-scale and multi-direction contourlet transform to research the inherent properties of image. Contours and curves projected from surfaces to images are perceptually relevant, and have intrinsic geometrical properties [36]. The directional multi-resolution contourlet transform (CT) efficiently captures and represents singularities along smooth object boundaries in images [37]. As we shall see, we can make use of the nonlinear statistical dependencies that exist between CT coefficients to assess changes in image quality arising from image degradations.

The contourlet transform, or pyramidal directional filter bank (DFB) was proposed by Do and Vetterli [37] in 2002. As an efficient representation of image structure, it decomposes a 2-D image into low-pass

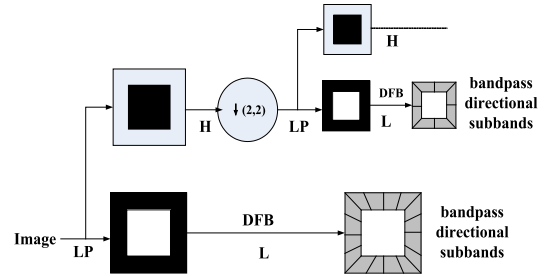


Fig. 2. Illustration of contourlet transform.

and high-pass sub-bands using a Laplacian Pyramid (LP), thereby highlighting singularities. A directional filter bank (DFB) is applied to the highpass sub-bands to capture singular points distributed along like directions. Fig. 2 illustrates the filter structure of the contourlet transform. Fig. 3 shows an example of a contourlet decomposition of luminance image and gradient image using a 3 lowpass and 8 directional bandpass transform. The principle advantage of applying a contourlet transform is that its characteristics can be used to capture image structures that exhibit directional multi-scale dependencies, and conversely, modifications of these dependencies owing to image degradation.

We follow the philosophy that NSS models are regular, and natural images possess an underlying statistical structure that can be modeled. Likewise, once an image is distorted, it becomes ‘unnatural’, and its adherence to NSS models will be affected. We evaluate the impairment in quality of an afflicted image by characterizing its statistical unnaturalness. In this section, we capture ‘quality-aware’ features using the contourlet transform.

2.2.1. Energy statistics within scales (ESS)

As observed in [38], the energy of a natural image falls off by a factor of roughly $1/f$, wherein f denotes radial spatial frequency. Field [39] pointed out that the spectra of images of natural scenes are linear on a logarithmic scale and that the responses of visual cortical cells are matched to this characteristic. In the contourlet domain, this property indicates energy decay as the scale becomes finer, and the relationship becomes linear in the log-domain. Accordingly, we employ statistical features derived from logarithms of the contourlet transform. The logarithmic energy of the sub-band coefficients are computed by the following formula (7).

$$E_{s,o} = \frac{1}{N} \sum \log_2 |C_{s,o}| \quad (7)$$

wherein E is the energy of each sub-band of scale index S and orientation index O , and C represents the value of the coefficient of each sub-band, and N is the number of sub-band coefficients. To capture the change of logarithmic energy of the sub-band coefficients usefully, we

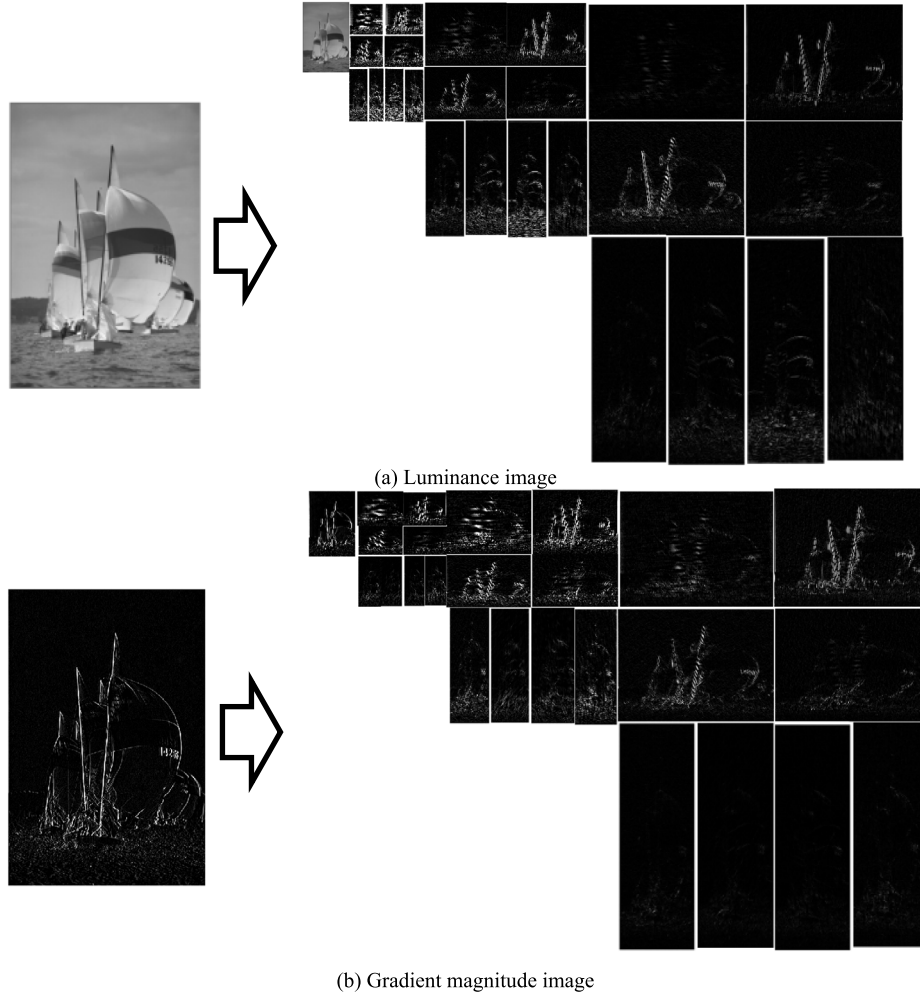


Fig. 3. An example of the contourlet transform.

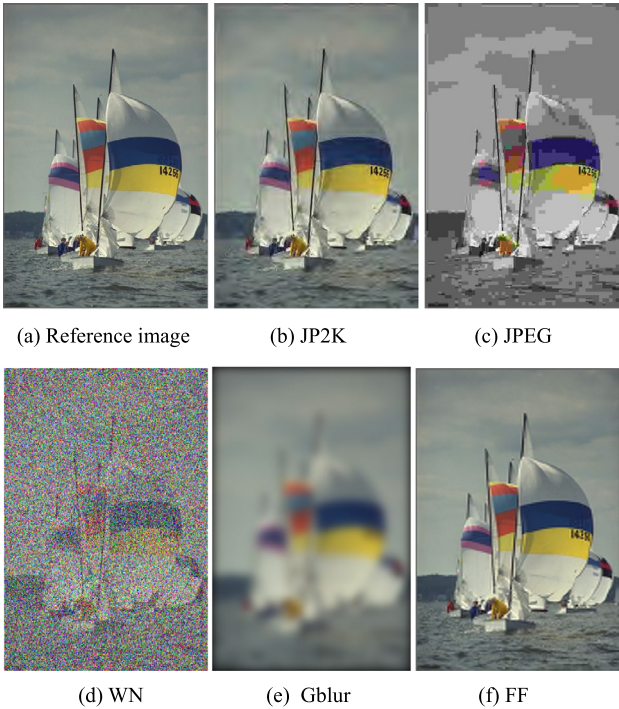


Fig. 4. Illustration of reference image and its five distorted versions.

choose a 3-scale and 8-orientation contourlet transform for extracting quality-aware features in this paper after considering the effectiveness and complexity jointly.

Fig. 4 shows an original image and five types of distorted images from the LIVE IQA database. Their energy distribution of CIELAB space and gradient magnitude images are shown in Fig. 5, noting where eight sub-band orientation energies are reordered in the 7th, 2nd, 6th, 3rd, 8th, 1st, 5th, 4th at each scale. The horizontal axis of each plot is the sub-band index, while the vertical axis represents the sub-band energy values.

As may be seen from Fig. 5, the energy curves of the reference images decrease almost linearly as the scale become finer. However, for the JP2K distortion category, the energy curves of L^* , a^* , b^* are linear at the coarse scale, but are destroyed at the finer scale. For the WN distortion category, the energy curves of L^* , a^* , b^* rise unnaturally at finer scales owing to the added energy. For Gblur and FF distortion categories, the energy attenuation of L^* , a^* , b^* is quite significant at medium to fine scales. All the energy attenuation of gradient image of JP2K, WN, Gblur and FF distortion categories are choppy. The exception is the JPEG distortion category, which although it does lead to high-frequency information loss, also produces high-energy DCT blocking effects. Since the contourlet energies of distorted image types exhibit significant variance, we use them as the first type of quality-aware image features.

2.2.2. Energy differences across scales (EDAS)

The energy distribution across scales will also be modified by image distortions. It can be clearly seen from Fig. 6 that energies of different

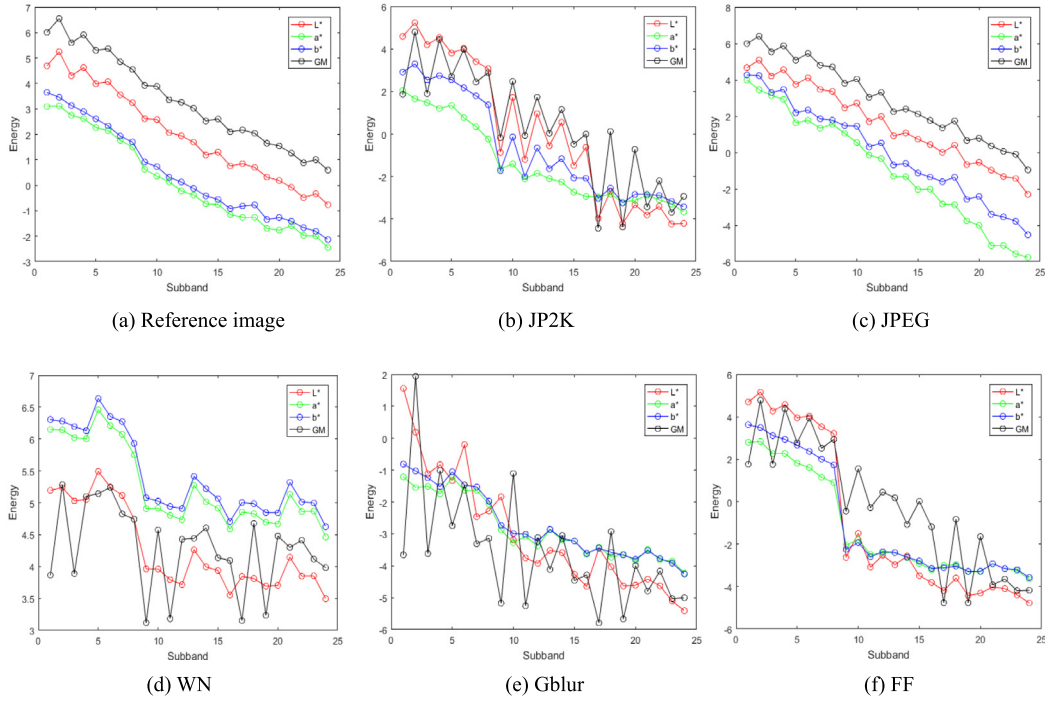


Fig. 5. Energy distribution of CIELAB components and gradient magnitude of reference image and five distorted versions.

distortion categories among different scales along the same orientation vary relative to each other. In order to capture energy changes across scales, we calculate the energy differences between scales along the same orientation as follows:

$$\begin{cases} D_{1i} = E_{1i} - E_{2i} \\ D_{2i} = E_{2i} - E_{3i} \\ D_{3i} = E_{1i} - E_{3i} \end{cases} \quad (8)$$

wherein E_{1i}, E_{2i}, E_{3i} represent the energies along different orientations at the first, second and third scale respectively, and where D_{1i}, D_{2i}, D_{3i} are the energy differences across scales.

In order to capture the relationships between distortion category and energy differences, we show the distribution of the energy differences of D_{1i}, D_{2i}, D_{3i} corresponding to the five types of distortion of the L^*, a^*, b^* components and the gradient magnitude image in Fig. 6. As can be seen from Fig. 6, the distributions of the energy differences for different distortion categories vary widely. For example, the energy differences on WN distortion category are relatively quite low, since the noise contribution caused the overall distorted image energies to remain high across scales. The distributions of energy differences are changed by distortion, so we also deploy energy differences across scales as the second type of quality-aware image features.

2.2.3. Neighboring energy statistics across scales (NESS)

In [8], Mittal et al. proposed that the empirical distributions of the products of adjoining bandpass pixels can be statistically modeled. Motivated by the work in [8], Zhang et al. [40] computed five orientations of neighboring pixels and found that the empirical distributions can be similarly modeled. We have found that the statistics of contourlet coefficients are strongly non-Gaussian, and we use the method of [40] to compute neighboring statistics across scales. To better model the statistical characteristics of pairs of pixels, we compute the logarithm of the sub-band coefficients:

$$J(i, j) = \ln(|I(i, j) + K|) \quad (9)$$

wherein $I(i, j)$ denotes the distorted image, and $K = 0.01$ is a small stabilizing constant to prevent instabilities from occurring when the

logarithm tends to zero. Then at each spatial coordinate, compute the empirical distributions of five neighboring pairs of pixels as the logarithm of sub-band coefficients:

$$\begin{cases} R_1 = J(i, j+1) - J(i, j) \\ R_2 = J(i+1, j) - J(i, j) \\ R_3 = J(i+1, j+1) - J(i, j) \\ R_4 = J(i+1, j-1) - J(i, j) \\ R_5 = J(i, j) + J(i+1, j+1) - J(i, j+1) - J(i+1, j) \end{cases} \quad (10)$$

Images are naturally multi-scale, and their sub-band coefficients exhibit relationships at each orientation but across scales, which relationships are also modified by distortions. In order to measure these relationships, merge the sub-bands across different scales along the each orientation form a 1-D vector, then calculate R_1 – R_5 using (9) and (10). We found that R_2 contains little useful information, and that R_3 and R_4 are similar to R_1 , so we only used R_1 and R_5 , which are affected by distortion.

Fig. 7 shows the empirical distribution of the coefficients of R_1 and R_5 . As can be seen from Fig. 7, the histograms of coefficients across scales can be modeled as following a generalized Gaussian distribution (GGD) with zero mean [41]. From the fit to this model, we calculate the shape and scale parameter, which controls the attenuation and the width of the generalized Gaussian density function respectively. The density of the GGD is given by [42]:

$$f(x; \alpha, \beta) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp\left(-\left(\frac{|x|}{\beta}\right)^\alpha\right) \quad (11)$$

wherein α is the shape parameter, β denotes the scale parameter, and $\Gamma(\cdot)$ is the gamma function. In our experiments, we found that the distribution over all three scales follow this model. In order to reduce redundancy, we use the sub-bands between the second and third scales along orientations as the third kind of quality-aware image feature.

3. Image quality assessment model using contourlet energy statistics

Our proposed algorithm is depicted in Fig. 8. First, the CIELAB components and gradient image are computed. Next, the contourlet

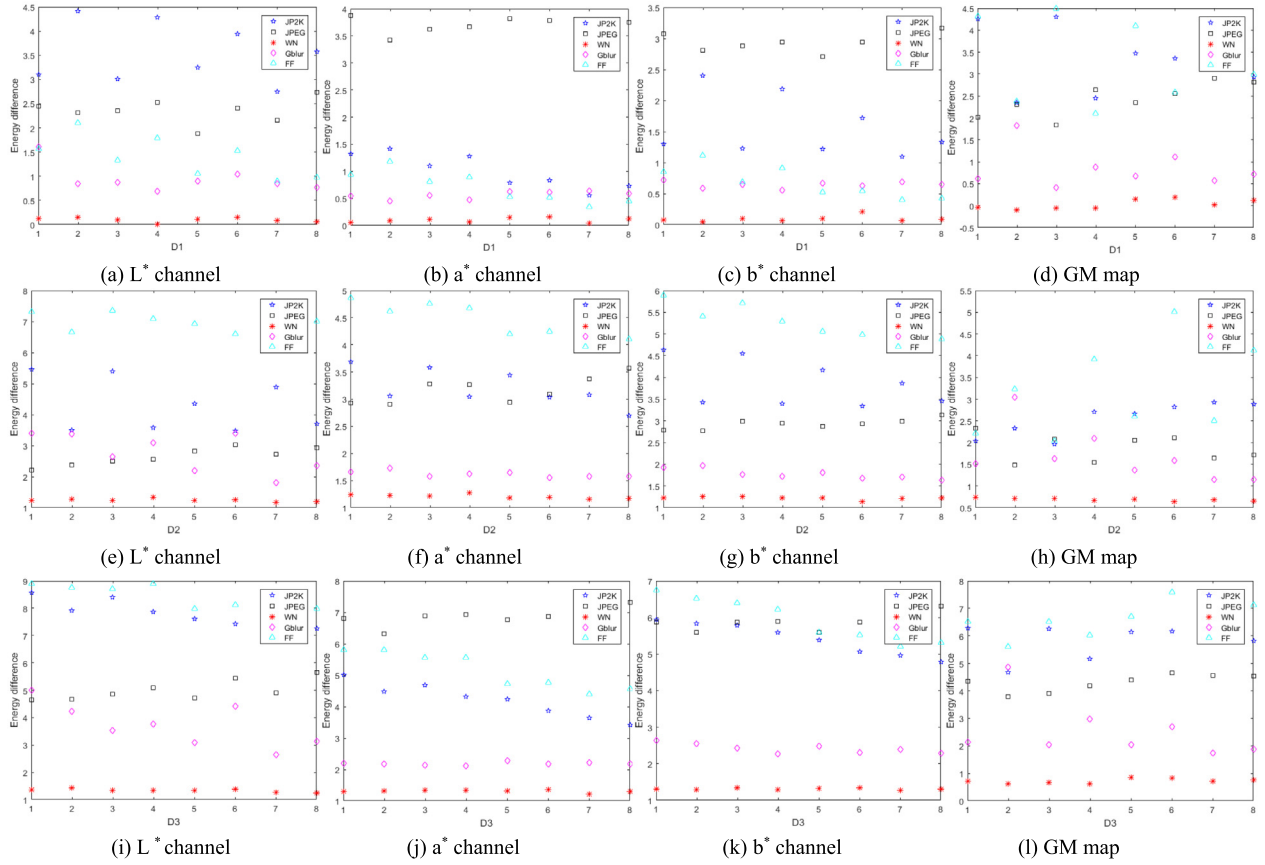


Fig. 6. Distributions of energy differences across scales for five types of distortions. The first to the last columns show the L^* channel, a^* channel, b^* channel and gradient magnitude map, respectively.

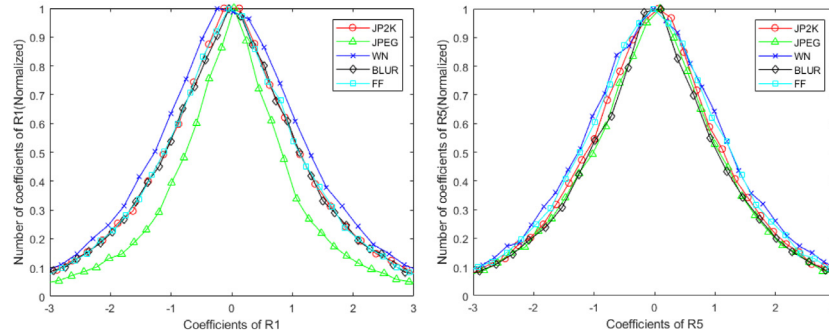


Fig. 7. Histograms of coefficients of R_1 and R_5 across scales.

transform is computed on the CIELAB color components and the gradient magnitude map, from which the following features are extracted: the energies of sub-band coefficients within scales, the energy differences across scales, and the neighboring energy statistics across scales. These quality-aware features are summarized in Table 1.

To obtain an image quality assessment model, a mapping must be established from feature space to subjective scores using a regression module. We used the SVR implemented by the LIBSVM package [43]. The kernel used is the radial basis function (RBF) kernel.

4. Experiments and discussions

To evaluate the performance of the algorithm, three indices are used as the criteria: the Spearman rank order correlation coefficient (SROCC), which reflects prediction monotonicity, the linear correlation coefficient (PLCC) after non-linear regression, the root mean square

Table 1

Summary of the features and their computation method.

Feature	Feature description	Computation procedure
CES (F_1 – F_{24})	Energy within scales	Sub-band image coefficients by contourlet transform
EDAS (F_{25} – F_{48})	Energy differences across scales	Compute the energy differences across scales along 8 orientations
NESS (F_{49} – F_{80})	Neighboring energy statistics across scales	Fit GGD to R_1 and R_5 after merging all orientations at the second and third scales

error (RMSE), which is used to measure the deviation between the predicted and the subjective scores. For the second and third criteria, the logistic function specified in [44] is employed to fit the model predictions to the subjective scores. Values of SROCC and PLCC that

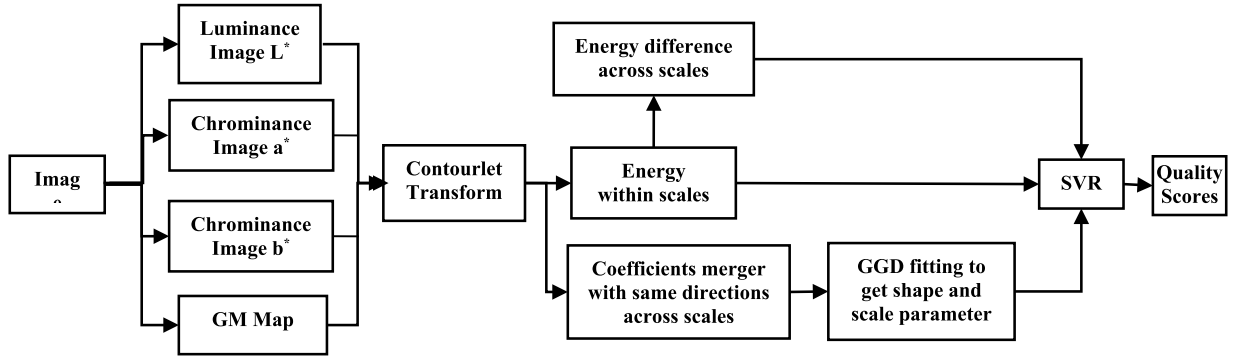


Fig. 8. Overall ContourletQA process.

are close to 1 indicate good performance in terms of correlation with human perception, but the value of RMSE that is close to 0 indicates good performance.

4.1. Comparison on synthetic distortion databases

Firstly we test our proposed ContourletQA on the synthetic LIVE, CSIQ and TID2013 IQA database. The LIVE database contains five types of distorted categories: JP2K, JPEG, WN, Gblur, and FF, including 29 reference images and 779 distorted images. Each distorted image has a difference mean opinion score (DMOS) which represents the subjective score of the image. The CSIQ database contains 6 types of distortion: JP2K, JPEG, WN, Gblur, fnoise and contrast distortion, including 30 reference images and 866 distorted images. The TID2013 database contains 25 reference images and 3000 distorted images over 24 types of distortion.

Since learning-based approaches require a training process to calibrate the regressor module, we randomly chose 80% reference images and their associated distorted images for training, and the rest for testing. No overlap was allowed between the train and test content. In order to verify robustness, we repeated the above random procedure 1000 times and report the median of the correlation results over the 1000 iterations.

We compare the performance of ContourletQA with several reported top NR-IQA indices BRISQUE [8], DIIIVINE [3], CurveletQA [10], DIIIVINE-GGSM [4], HFD-BIQA [26], and PSQA-I [15]. The median SROCC, PLCC and RMSE across 1000 trials of each metric are tabulated in Table 2. Avg. is the direct average results on the three databases. From Table 2, it can be seen PSQA-I performed the best on the LIVE database, but ContourletQA did better on the TID2013 and CSIQ databases overall, ContourletQA delivered excellent consistency relative to subjective evaluation, and outperformed the compared state-of-the-art NR IQA models.

For a better explanation, we also plot the correlations between objective scores and subjective scores on several synthetic distorted databases as shown in Fig. 9, where three classical NR-IQA methods BRISQUE, DIIIVINE and CurveletQA are also listed for comparison.

From Fig. 9, it can be seen that predicted scores of the proposed method correlates well with the human subjective judgments from the perspective of monotonicity and linearity. These results further confirm that gradient information and color channels both have the ability to characterize image distortions via multi-scale and multi-orientation transform.

We also tabulate the SROCC results on each distortion category in Tables 3–5. For brevity, we do not include PLCC results, but they are identical. Since Lu et al. [29] only report their results on LIVE IQA database, we only list their results on Table 3. From Tables 3–5, it may be observed that ContourletQA performed the best on JPEG distortion in the LIVE IQA database and on almost all distortion types in the TID2013 and CSIQ databases.

4.2. Database independence

Since our method is based on a training procedure, it is necessary to demonstrate whether the prediction model generalizes across databases. We trained the model on the LIVE IQA database, the CSIQ database and the TID2013 database respectively, and tested on the other databases. Since both the TID2013 database and the CSIQ database do not contain FF distortion, we only tested on the corresponding distortions that it was trained on: JPEG, JP2K, WN and Gblur. The cross-validation results are listed in Tables 6 and 7, respectively.

From Tables 6 and 7, it may be seen that ContourletQA achieved the best overall performance, outperforming the compared state-of-the-art NR IQA indices, suggesting it is not database dependent.

4.3. Comparison on LIVE wild challenge database

Images captured by mobile camera devices in the real world suffer from complex mixtures of (often) multiple distortions. The LIVE Wild Challenge (CLIVE) Database [15] consists of 1162 naturally distortion images. These images are affected by unknown mixtures of numerous possibly interacting authentic distortions of diverse severities. Among these, there are 149 images that were captured under difficult, low-light conditions. As shown in Fig. 10, we choose some examples suffered low-light distortions.

We trained several NR-IQA models from scratch using random samples of 80% of the images, and then tested them on the remaining non-overlapping 20% of the data. This process of randomly splitting the dataset was repeated 1000 times to eliminate any bias introduced by the division of data. The median SROCC, PLCC and RMSE scores across 1000 trials of each model are tabulated in Table 8. From Table 8, it can be seen that our proposed ContourletQA is highly competitive on the LIVE wild database, suggesting its efficiency on real distortions.

For further analysis if the trained set of parameters is database specific, we test the proposed model in the combination of different image databases which consist of the LIVE IQA database and the LIVE Wild Challenge (CLIVE) Database. During the training and testing procedure, we choose 80% of the randomly images from the combined database as the training set and the remained 20% images are chosen as the testing set. The results predicted by each algorithm are compared in Fig. 11 and Table 9. Among Fig. 11, the symbol DIIIVINE-II represents the DIIIVINE-GGSM model of reference [4]. As shown in Fig. 11, the proposed method computes the most accurate estimate. Moreover, it indicates that the performance of ContourletQA is not database dependent and that once trained, it is capable of accurately assessing the quality of images across the distortions (both single and multiple, or synthetic and authentic) that it is trained for.

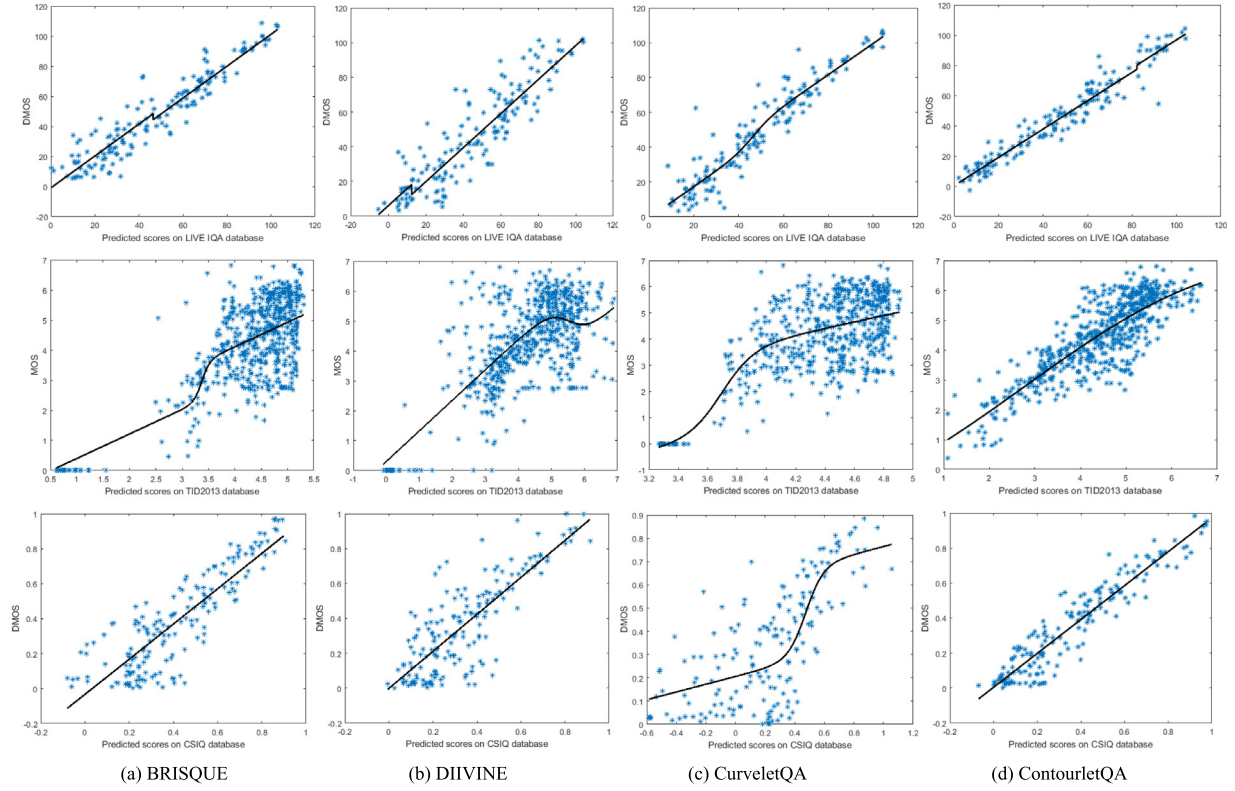


Fig. 9. Scatter plots between objective and subjective scores.

Table 2
Performance comparison on several IQA databases. (The top 2 are highlighted in bold type.)

Database		BRISQUE	DIIVINE	CurveletQA	DIIVINE-GGSM	HFD-BIQA	PSQA-I	ContourletQA
LIVE	SROCC	0.937	0.916	0.930	0.931	0.948	0.960	0.952
	PLCC	0.942	0.917	0.933	0.930	0.951	0.960	0.953
	RMSE	8.971	9.435	9.131	9.125	8.437	7.852	8.058
TID2013	SROCC	0.520	0.583	0.445	0.611	0.681	0.637	0.730
	PLCC	0.626	0.672	0.522	0.679	0.764	0.673	0.766
	RMSE	1.557	0.933	1.996	0.885	0.797	0.876	0.722
CSIQ	SROCC	0.768	0.769	0.661	0.792	0.842	0.834	0.889
	PLCC	0.831	0.800	0.695	0.825	0.890	0.880	0.906
	RMSE	0.225	0.210	0.294	0.187	0.120	0.155	0.084
Avg.	SROCC	0.742	0.756	0.679	0.779	0.824	0.810	0.857
	PLCC	0.800	0.796	0.717	0.812	0.868	0.838	0.875
	RMSE	3.584	3.526	3.807	3.399	3.118	2.961	2.954

Table 3
Median SROCC results on individual distortion of LIVE IQA database.

Database	Distortion type	BRISQUE	DIIVINE	CurveletQA	Lu	DIIVINE-GGSM	HFD-BIQA	PSQA-I	ContourletQA
LIVE	JP2K	0.902	0.913	0.938	0.824	0.911	0.943	0.957	0.926
	JPEG	0.942	0.910	0.912	0.562	0.927	0.951	0.960	0.965
	WN	0.964	0.984	0.988	0.601	0.982	0.972	0.988	0.985
	Gblur	0.951	0.921	0.965	0.856	0.927	0.919	0.955	0.944
	FF	0.861	0.863	0.901	0.823	0.900	0.905	0.912	0.897

4.4. Comparison of each component in the proposed method

To verify impact of the CIELAB components and the gradient magnitude image on the proposed metric, we test its partial versions in LIVE IQA database, CSIQ database and TID2013 database. We plot the result comparison of each component in the proposed method in Fig. 12.

From Fig. 12, we can see that the four partial scores are highly correlated with MOSs/DMOSs, but the combinations of these components are raising the accuracy of quality prediction result as shown in Table 2. It supports our assumption that the gradient magnitude information

and the multi-channel information are complementary for determining image quality.

4.5. Discussion and future plan

In future, we plan to further improve the model by including content-sensitive cues, as in [45]. Furthermore, considering that the amount of perceptual features in this model is excessive, some methods can be adopted to remove some irrelevant ones for improving the performance, such as recently proposed ensemble technologies [46–48]. In [47], the stacked selective ensemble method, which effectively

Table 4
Median SROCC results on individual distortion of TID2013 database.

Database	Distortion type	BRISQUE	DIIVINE	CurveletQA	DIIVINE-GGSM	PSQA-I	ContourletQA
TID2013	AGN	0.852	0.855	0.829	0.877	0.881	0.935
	AGNC	0.709	0.712	0.746	0.715	0.764	0.845
	SCN	0.491	0.463	0.742	0.842	0.923	0.945
	MN	0.575	0.675	0.693	0.687	0.482	0.786
	HFN	0.753	0.878	0.850	0.921	0.896	0.942
	IN	0.630	0.806	0.735	0.830	0.831	0.918
	QN	0.798	0.165	0.704	0.580	0.780	0.888
	GB	0.813	0.834	0.820	0.830	0.874	0.963
	ID	0.586	0.723	0.737	0.806	0.762	0.892
	JPEG	0.852	0.629	0.735	0.887	0.770	0.913
	JP2K	0.893	0.853	0.821	0.866	0.893	0.910
	JPEGt	0.315	0.239	0.091	0.284	0.061	0.517
	JP2Kt	0.359	0.061	0.527	0.341	0.385	0.706
	NEPN	0.145	0.060	0.010	0.184	0.055	0.391
	LBD	0.224	0.093	0.024	0.126	0.266	0.085
	MS	0.124	0.010	0.055	0.066	0.295	0.293
	CC	0.040	0.460	0.453	0.457	0.020	0.656
	CS	0.109	0.068	0.243	0.287	0.323	0.459
	MGN	0.724	0.787	0.789	0.823	0.864	0.930
	CN	0.008	0.116	0.014	0.146	0.074	0.310
	LC	0.685	0.633	0.642	0.677	0.682	0.859
	ICQ	0.764	0.436	0.783	0.828	0.820	0.903
	CA	0.616	0.661	0.755	0.807	0.784	0.922
	SSR	0.784	0.833	0.818	0.825	0.862	0.929

Table 5
Median SROCC results on individual distortion of CSIQ database.

Database	Distortion type	BRISQUE	DIIVINE	CurveletQA	DIIVINE-GGSM	PSQA-I	ContourletQA
CSIQ	JP2K	0.853	0.831	0.837	0.851	0.839	0.937
	JPEG	0.854	0.800	0.798	0.880	0.887	0.910
	fnoise	0.824	0.817	0.778	0.857	0.920	0.924
	Gblur	0.849	0.872	0.817	0.840	0.859	0.918
	contrast	0.357	0.396	0.264	0.413	0.500	0.705
	AGN	0.773	0.866	0.802	0.898	0.934	0.944

Table 6
Median SROCC results on cross database.

Train database	Test database	BRISQUE	DIIVINE	CurveletQA	DIIVINE-GGSM	HFD-BIQA	PSQA-I	ContourletQA
LIVE	TID2013	0.726	0.849	0.881	0.872	0.899	0.911	0.914
	CSIQ	0.827	0.854	0.789	0.871	0.843	0.899	0.903
TID2013	LIVE	0.795	0.621	0.878	0.821	0.890	0.860	0.876
	CSIQ	0.808	0.641	0.737	0.723	0.821	0.872	0.867
CSIQ	LIVE	0.633	0.521	0.851	0.859	0.918	0.920	0.928
	TID2013	0.571	0.764	0.858	0.791	0.888	0.882	0.899

Table 7
Median PLCC results on cross database.

Train database	Test database	BRISQUE	DIIVINE	CurveletQA	DIIVINE-GGSM	HFD-BIQA	PSQA-I	ContourletQA
LIVE	TID2013	0.721	0.860	0.877	0.893	0.921	0.905	0.923
	CSIQ	0.840	0.875	0.810	0.891	0.900	0.903	0.917
TID2013	LIVE	0.790	0.627	0.866	0.829	0.874	0.858	0.861
	CSIQ	0.839	0.659	0.779	0.738	0.877	0.888	0.889
CSIQ	LIVE	0.644	0.522	0.835	0.842	0.910	0.910	0.920
	TID2013	0.583	0.813	0.868	0.835	0.917	0.874	0.903

Table 8
Performance comparison on LIVE wild challenge database.

Crit.	BRISQUE	DIIVINE	CurveletQA	DIIVINE-GGSM	HFD-BIQA	PSQA-I	ContourletQA
SROCC	0.60	0.58	0.59	0.60	0.760	0.57	0.64
PLCC	0.61	0.59	0.62	0.61	0.776	0.58	0.65
RMSE	15.88	16.24	15.91	15.83	–	16.44	15.28

aggregates selected samples, environmental factors, and temporal factors, is very adept in PM2.5 prediction. In [48], the authors skillfully combined texture information and contour information extracted with dual deep subnetworks to achieve the leading level of the world for smoke detection.

5. Conclusions

We analyzed the distributions of the sub-band contourlet coefficients of distorted images, and constructed and extracted three kinds of quality-aware contourlet features: the energy of sub-band coefficients



Fig. 10. Examples of night-time images from the LIVE wild challenge database.

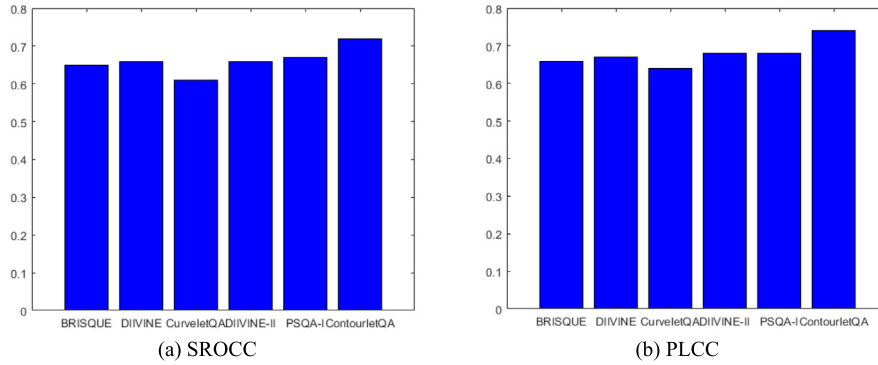


Fig. 11. Performance comparison on combination of different databases.

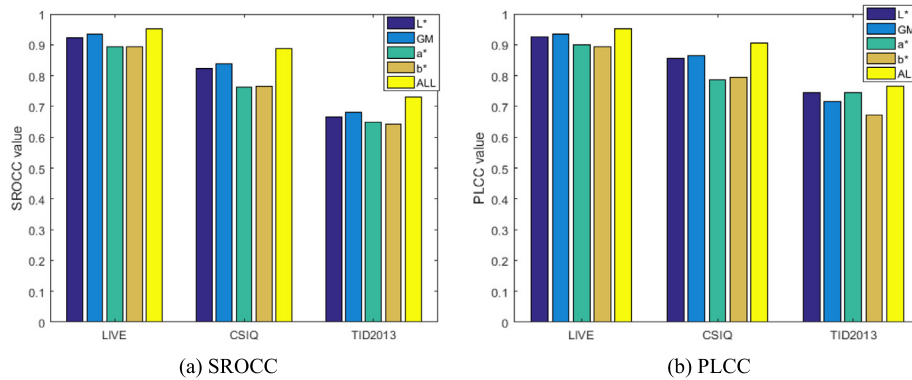


Fig. 12. Comparison of the quality prediction of each feature map.

Table 9

Performance comparison on combination of synthetic and authentic distortion databases.

Crit.	BRISQUE	DIIVINE	CurvetQA	DIIVINE-GGSM	PSQA-I	ContourletQA
SROCC	0.65	0.66	0.61	0.66	0.67	0.72
PLCC	0.66	0.67	0.64	0.68	0.68	0.74
RMSE	17.53	16.94	18.18	16.90	16.82	15.66

within scales and the energy differences between scales as well as the statistical relationships of pixels across scales, and demonstrated how these features correlate with human judgments of image quality. A new NR IQA model was constructed by learning the responses of these quality-aware features on human judgments of distorted images using a support vector regression (SVR) model.

We undertook a thorough evaluation of the new model on the LIVE, TID2013, CSIQ and LIVE Wild Challenge databases. The experimental results suggest that our proposed model is independent of database

content, and outperforms current top NR IQA algorithms, making it an attractive solution for practical NR IQA applications.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] M.J. Chen, A.C. Bovik, No-reference image blur assessment using multiscale gradient, *EURASIP J. Image Video Process.* 1 (2011) 1–11.
- [2] A. Moorthy, A. Bovik, A two-step framework for constructing blind image quality indices, *IEEE Signal Process. Lett.* 17 (5) (2010) 513–516.
- [3] A. Moorthy, A.C. Bovik, Blind image quality assessment: From natural scene statistics to perceptual quality, *IEEE Trans. Image Process.* 20 (12) (2011) 3350–3364.
- [4] P. Gupta, A.K. Moorthy, R. Soundararajan, A.C. Bovik, Generalized Gaussian scale mixtures: A model for wavelet coefficients of natural images, *Signal Process., Image Commun.* 66 (2018) 87–94.

- [5] H. Tang, N. Joshi, A. Kapoor, Learning a blind measure of perceptual image quality, in: *IEEE Conf. Comput. Vis. Pattern Recogn.*, 2011, pp. 305–312.
- [6] M. Saad, A.C. Bovik, C. Charrier, Blind image quality assessment: A natural scene statistics approach in the DCT domain, *IEEE Trans. Image Process.* 21 (8) (2012) 3339–3352.
- [7] C. Li, A.C. Bovik, X. Wu, Blind image quality assessment using a general regression neural network, *IEEE Trans. Neural Netw. Learn. Syst.* 22 (5) (2011) 793–799.
- [8] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, *IEEE Trans. Image Process.* 20 (12) (2012) 4695–4708.
- [9] X. Gao, F. Gao, D.C. Tao, X. Li, Universal blind image quality assessment metrics via natural scene statistics and multiple kernel learning, *IEEE Trans. Neural Netw. Learn. Syst.* 24 (2013) 2013–2026.
- [10] L. Liu, H. Dong, H. Huang, A.C. Bovik, No-reference image quality assessment in curvelet domain, *Signal Process., Image Commun.* 29 (4) (2014) 494–505.
- [11] L. Liu, Y. Hua, Q. Zhao, H. Huang, A.C. Bovik, Blind image quality assessment by relative gradient statistics and adaboosting neural network, *Signal Process., Image Commun.* 40 (2016) 1–15.
- [12] M. Oszust, No-reference image quality assessment using image statistics and robust feature descriptors, *IEEE Signal Process. Lett.* 24 (11) (2017) 1656–1660.
- [13] H. Gao, Q. Miao, J. Yang, Z. Ma, Image quality assessment using image description in information theory, *IEEE Access* 6 (2018) 47181–47188.
- [14] J. Gu, G. Meng, J.A. Redi, S. Xiang, C. Pan, Blind image quality assessment via vector regression and object oriented pooling, *IEEE Trans. Multimed.* 20 (5) (2018) 1140–1153.
- [15] L. Liu, T. Wang, H. Huang, Pre-attention and spatial dependency driven no-reference image quality assessment, *IEEE Trans. Multimed.* 21 (9) (2019) 2305–2318.
- [16] K. Gu, D. Tao, J. Qiao, W. Lin, Learning a no-reference quality assessment model of enhanced images with big data, *IEEE Trans. Neural Netw. Learn. Syst.* 29 (4) (2018) 1301–1313.
- [17] K. Gu, X. Xu, J. Qiao, Q. Jiang, W. Lin, D. Thalmann, Learning a unified blind image quality metric via on-line and off-line big training instances, *IEEE Trans. Big Data* (2019) <http://dx.doi.org/10.1109/TBDATA.2019.2895605>.
- [18] K. Gu, G. Zhai, W. Lin, X. Yang, W. Zhang, No-reference image sharpness assessment in autoregressive parameter space, *IEEE Trans. Image Process.* 24 (10) (2015) 3218–3231.
- [19] W. Hou, X. Gao, D. Tao, X. Li, Blind image quality assessment via deep learning, *IEEE Trans. Neural Netw. Learn. Syst.* 26 (6) (2015) 1275–1286.
- [20] L. Kang, P. Ye, Y. Li, D. Doermann, Convolutional neural networks for no-reference image quality assessment, in: *IEEE Conf. Comput. Vis. Pattern Recogn.*, 2014, pp. 1733–1740.
- [21] A. Artusi, F. Banterle, F. Carrara, A. Moreo, Efficient evaluation of image quality via deep-learning approximation of perceptual metrics, *IEEE Trans. Image Process.* 29 (2020) 1843–1855.
- [22] Q. Yan, D. Gong, Y. Zhang, Two-stream convolutional networks for blind image quality assessment, *IEEE Trans. Image Process.* 28 (5) (2019) 2200–2211.
- [23] S. Bosse, D. Maniry, K. Müller, T. Wiegand, W. Samek, Deep neural networks for no-reference and full-reference image quality assessment, *IEEE Trans. Image Process.* 27 (1) (2018) 206–219.
- [24] K. Lin, G. Wang, Hallucinated-iqa: no-reference image quality assessment via adversarial learning, in: *IEEE Conf. Comput. Vis. Pattern Recogn.*, 2018, pp. 732–741.
- [25] D. Pan, P. Shi, M. Hou, Z. Ying, S. Fu, Y. Zhang, Blind predicting similar quality map for image quality assessment, in: *IEEE Conf. Comput. Vis. Pattern Recogn.*, 2018, pp. 6373–6382.
- [26] J. Wu, J. Zeng, W. Dong, G. Shi, W. Lin, Blind image quality assessment with hierarchy: Degradation from local structure to deep semantics, *J. Vis. Commun. Image Represent.* 58 (2019) 353–362.
- [27] M. Liu, X. Yang, Image quality assessment using contourlet transform, *Opt. Eng.* 48 (10) (2009) 107201.
- [28] W. Lu, X. Gao, X. Li, D. Tao, An image quality assessment metric based contourlet, in: *IEEE Int. Conf. Image Process.*, 2008, pp. 1172–1175.
- [29] W. Lu, K. Zeng, D. Tao, Y. Yuan, X. Gao, No-reference image quality assessment in contourlet domain, *Neurocomputing* 73 (4–6) (2010) 784–794.
- [30] H.R. Sheikh, Z. Wang, L. Cormack, A.C. Bovik, LIVE Image Quality Assessment Database Release 2, <<http://live.ece.utexas.edu/research/quality>>.
- [31] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazari-an, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, Image database tid2013: Peculiarities results and perspectives, *Signal Process., Image Commun.* 30 (2015) (2013) 57–77.
- [32] E.C. Larson, D.M. Chandler, Categorical image quality (CSIQ) database [DB/OL]. <http://vision.okstate.edu/csiq>.
- [33] D. Ghadiyaram, A.C. Bovik, Massive online crowdsourced study of subjective and objective picture quality, *IEEE Trans. Image Process.* 25 (1) (2016) 372–387.
- [34] U. Rajashekar, Z. Wang, E. Simoncelli, Perceptual quality assessment of color images using adaptive signal representation, *Proc. SPIE* 31 (4) (2010) 75271L.
- [35] R.C. Gonzalez, R.E. Woods, *Digital Image Processing*, Prentice-Hall, New Jersey, 2008.
- [36] K.A. Panetta, E.J. Wharton, S.S. Agaian, Human visual system-based image enhancement and logarithmic contrast measure, *IEEE Trans. Syst. Man Cybern.* B 38 (1) (2008) 174–188.
- [37] M.N. D.O., M. Vetterli, The contourlet transform: an efficient directional multiresolution image representation, *IEEE Trans. Image Process.* 14 (12) (2005) 2091–2106.
- [38] D.L. Ruderman, The origins of scaling in natural images, *Vis. Res.* 37 (23) (1997) 3385–3398.
- [39] D.J. Field, Relations between the statistics of natural images and the response properties of cortical cells, *J. Opt. Soc. Am. A* 4 (12) (1987) 2379–2394.
- [40] Y. Zhang, D.M. Chandler, No-reference image quality assessment based on log-derivative statistics of natural scenes, *J. Electron. Imaging* 22 (4) (2013) 043025.
- [41] D.L. Ruderman, The statistics of natural images, *Netw.: Comput. Neural Syst.* 5 (4) (1994) 517–548.
- [42] K. Sharifi, A.L. Garcia, Estimation of shape parameter for generalized gaussian distributions in subband decompositions of video, *IEEE Trans. Circuits Syst. Video Technol.* 5 (1) (1995) 52–56.
- [43] C.-C. Chang, C.-J. Lin, Libsvm: a library for support vector machines, *ACM Trans. Intell. Syst. Technol.* 2 (3) (2011) 27, [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [44] Video Quality Experts Group, Video quality experts group final report from the video quality experts group on the validation of objective models of video quality assessment, in: VQEG, 2000.
- [45] C. Li, A.C. Bovik, Content-partitioned structural similarity index for image quality assessment, *Signal Process., Image Commun.* 25 (7) (2010) 517–526.
- [46] Z. Zhou, J. Wu, W. Tang, Ensembling neural networks: Many could be better than all, *Artificial Intelligence* 137 (1–2) (2002) 239–263.
- [47] K. Gu, Z. Xia, J. Qiao, Stacked selective ensemble for PM2.5 forecast, *IEEE Trans. Instrum. Meas.* 69 (3) (2020) 660–671.
- [48] K. Gu, Z. Xia, J. Qiao, W. Lin, Deep dual-channel neural network for image-based smoke detection, *IEEE Trans. Multimedia* 22 (2) (2020) 311–323.