

Blind image quality assessment based on joint log-contrast statistics

Qiaohong Li^a, Weisi Lin^{a,*}, Ke Gu^b, Yabin Zhang^a, Yuming Fang^c

^aSchool of Computer Science and Engineering, Nanyang Technological University 639798, Singapore

^bBeijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing University of Technology, Beijing 100124, China

^cSchool of Information Technology, Jiangxi University of Finance and Economics, Nanchang 330013, China

ARTICLE INFO

Article history:

Received 13 March 2018

Revised 13 September 2018

Accepted 9 November 2018

Available online 15 November 2018

Communicated by Dr Jianjun Lei

Keywords:

Blind image quality assessment (BIQA)

No-reference (NR)

Natural scene statistics

Partial least square

ABSTRACT

During recent years, quality-aware features extracted from natural scene statistics (NSS) models have been used in development of blind image quality assessment (BIQA) algorithms. Generally, the univariate distributions of bandpass coefficients are used to fit a parametric probabilistic model and the model parameters serve as the quality-aware features. However, the inter-location, inter-direction and inter-scale correlations of natural images cannot be well exploited by such NSS models, as it is hard to capture such dependencies using univariate marginal distributions. In this paper, we build a novel NSS model of joint log-contrast distribution to take into account the across space and direction correlations of natural images (inter-scale correlation to be explored as the next step). Furthermore, we provide a new efficient approach to extract quality-aware features as the gradient of log-likelihood on the NSS model, instead of using model parameters directly. Finally, we develop an effective joint-NSS model based BIQA metric called BJLC (BIQA based on joint log-contrast statistics). Extensive experiments on four public large-scale image databases have validated that objective quality scores predicted by the proposed BIQA method are in higher accordance with subjective ratings generated by human observers compared with existing methods.

© 2018 Published by Elsevier B.V.

1. Introduction

Blind or no-reference image quality assessment has recently become an active research topic, whose objective is to estimate the human perception of image quality without any access to the original reference image [1,2]. It can be deployed in every location of an image communication system whenever the distorted image is available for evaluation. Generally, blind image quality assessment (BIQA) methods are divided into two groups, distortion-specific and general-purpose methods. Distortion-specific methods focus on specific applications where distortion types are known in advance [3–7]. General-purpose methods are designed to evaluate image quality without any cues to distortion types [8–14].

Natural scene statistics (NSS) based methods constitute the majority of existing general-purpose BIQA methods. The underlying philosophy of NSS based methods lies in that natural visual stimuli would possess certain statistical regularities, regardless of different contents, and distortions introduced in different image processing stages would damage such regularities [1,15]. Therefore, the deviation of the statistics of a query image from that of the NSS model

would reveal the distortion degree of that image. A plethora of NSS models in different spatial or transform domains have been incorporated into state-of-the-art BIQA methods [16–26]. BIQI (blind image quality index) [16] is a wavelet NSS based method that extracts generalized Gaussian distribution (GGD) parameters from wavelet subband coefficients. The exponential decay characteristics (EDC) of wavelet energies across scales have been utilized in SRNSS (sparse representation of NSS feature) method [18]. MKL (multiple kernel learning) extracts statistical features of wavelet marginal distribution, mutual information between neighboring coefficients and EDC properties as quality-aware features [19]. In DIIVINE (distortion identification-based image verity and integrity evaluation) [17], the divisively normalized wavelet subbands are used to build GGD models. To account for the correlation across scale and orientation, GGD models are fitted on the stacked subband coefficients with the same orientation or scale. Structural similarity between subbands is also calculated to measure their correlation. C-DIIVINE (Complex DIIVINE) [20], a complex extension of DIIVINE, has adopted three NSS models in complex wavelet domain, i.e., the complex GGD model of wavelet coefficient magnitudes, the GGD model of relative magnitudes and the wrapped Cauchy distribution (WCD) model of relative phases. BLIINDS (blind image integrity notator using DCT statistics) extracts the contrast, structure and anisotropy features from block DCT coefficients [21].

* Corresponding author.

E-mail address: wslin@ntu.edu.sg (W. Lin).

Later, it was extended to BLIINDS2 by using GGD to fit the block DCT coefficients and some partitions of the DCT block [22]. The mean subtracted contrast normalized (MSCN) coefficients are used to build the spatial domain NSS models, including GGD of MSCN coefficients and asymmetric GGD (AGGD) of pairwise MSCN coefficient products [23]. The joint statistics of gradient magnitude and Laplacian of Gaussian (LoG) responses in the form of marginal and dependency histograms are used to build GMLOG method [27]. Distributions of neighbouring pixel differences in four color channels (hue, saturation, opponent angle and spherical angle) are used to build color NSS models to account for spatial correlation in color descriptors [26].

There are also some methods using an ensemble of NSS models from multiple domains to extract a more comprehensive feature set. In DESIQUE (derivative statistics-based image quality evaluator), GGD distributions of pixel-wise and pairwise log-derivative statistics in both spatial and log-Gabor domains are combined to form the quality-aware features [24]. ILNIQE has adopted the GGD of MSCN coefficients, AGGD of MSCN products, GGD of directional gradients, Weibull distribution of gradient magnitudes in luminance, log-Gabor and opponent color space, as well as the Gaussian distribution of logarithmic-scale opponent color space [28]. In FRIQUEE (feature maps based referenceless image quality evaluation engine), dozens of feature maps over several transform domains and three color spaces are used to fit the GGD, AGGD and WCD models [25].

Despite that NSS models in different domains have been applied in previous BIQA methods, we can observe that the construction and utilization of NSS models mainly follow the same framework. Generally, the marginal distributions of some band-pass coefficients are fitted with a univariate parametric probabilistic model, then the model parameters are adopted as the quality-aware features. The most widely used natural image properties for BIQA include the non-Gaussian distribution of bandpass responses, and the correlation properties of natural images across space, scale and orientation. Univariate probabilistic models (such as 1-D GGD, AGGD, WCD) are usually employed to depict the non-Gaussian properties of bandpass coefficients [16,17,20,22–26]. It has also been used in several methods [23,24,26] to measure the spatial correlation by characterizing the marginal distribution of pairwise coefficients products or differences. For across scale and orientation correlation, existing BIQA methods calculate mutual information [19], structural correlation [17] or energy ratio [22] between neighboring subbands to measure the correlation.

Thus, we summarize the limitations of existing NSS based BIQA methods in three aspects: (1) univariate marginal distributions adopted by previous NSS models have difficulties to capture the correlation properties of natural images; (2) for measuring image correlation, only two neighboring pixels or subbands are considered each time, which is insufficient to reflect the complex degradation patterns as image distortions generally corrupt all pairs simultaneously; (3) the direct adoption of model parameters as quality-aware features is susceptible to the model fitting errors. To address these problems, we build a novel comprehensive joint log-contrast statistics based NSS model to account for the across space and orientation correlation simultaneously by means of multivariate Gaussian mixture model (GMM). Furthermore, we employ a new approach to extract quality-aware features as the gradient of log-likelihood on the NSS model instead of directly using the model parameters. In such a way, we circumvent the fitting error problems as no curve fitting is involved in the feature extraction. Extensive experiments conducted on four large-scale image databases (i.e., LIVE [29], CSIQ [30], TID2013 [31], CLIVE [32]) show that the proposed method highly correlates with human perception of visual quality.

The rest of this paper is structured as follows. In Section 2, the methodology of proposed BJLC quality metric is explained in details. Extensive experiments and analysis are presented in Section 3, followed by conclusions in Section 4.

2. BJLC metric

We propose to measure across space and direction correlations by using the joint log-contrast distribution and build a comprehensive NSS model to approximate it. We show that joint log-contrast response is sensitive to degraded images with different distortion types and quality levels, and NSS model built on it is indicative of image quality. Then we extract quality-aware features as the FV (Fisher Vector) on the generative NSS model, which describe the distance of one degraded image from the NSS model. Finally, PLS (partial Least Square) regression method is adopted to relate the quality-aware features to quality scores.

2.1. Joint directional log-contrast statistics

We propose to use the joint log-contrast response in spatial domain to measure the impact of various distortions on natural images, which is defined as follows. First, for each image pixel, we define a squared symmetric neighbor set of P pixels placed on a square whose sides have the length $(2R + 1)$, as shown in Fig. 1. Here P is the number of neighboring pixels, and R is the spatial resolution of the operator. Although we define the operator for a general case based on a squared symmetric neighbor set, one can also generalize them to a circular one by interpolation [33].

We take advantage of the pairwise pixel differences to account for the spatial correlation. Specifically, we subtract the center pixel value x_c from the P neighboring pixels on the square side (x_1, x_2, \dots, x_p) , yielding the feature vector $\mathbf{y} = (y_1, y_2, \dots, y_p) = (x_1 - x_c, x_2 - x_c, \dots, x_p - x_c)$. \mathbf{y} is the gray-level differences vector which denotes the contrast responses in different directions.

Logarithm function is used to mimic the nonlinear human perception on physical stimuli and narrow the range of contrast magnitudes:

$$z_i = \text{sign}(y_i) \log(|y_i| + 1) \quad (1)$$

The log-contrast response for each pixel x_c is then represented by the vector $\mathbf{z}_c = (z_1, z_2, \dots, z_p)$, and z_i represents the log-contrast response in the i th direction. Unlike previous works using the univariate parametric model to approximate the marginal distributions independently, we propose to jointly model the multivariate distribution by treating the log-contrast vector \mathbf{z}_c as a whole. In such a way, the inter-location and inter-direction correlations are implicitly encoded in the joint distribution.

The proposed method is designed on the assumption that cues to distortion type and degree can be captured by the joint distribution of directional log-contrast. To illustrate that the joint distribution varies with the levels of distortions, we extract a collection set of \mathbf{z}_c from distorted versions of the *painted house* image. All images are from TID2013 database, and sample blocks extracted from the distorted images are shown in Fig. 2. We use settings of $(R = 1, P = 8)$ to calculate the joint log-contrast vector, and one pixel is associated with an 8-by-1 feature vector \mathbf{z}_c . We use the collection of \mathbf{z}_c vectors from the reference image to learn one PCA (principal component analysis) mapping matrix. Then the first principal components associated with \mathbf{z}_c computed from distorted images can be obtained by mapping the collection set using the previous PCA matrix. By fixing the PCA projecting matrix, we make the first principal components of distorted images directly comparable.

As shown in Fig. 3, one can observe that the first principal components of joint log-contrast response are quite informative of subjective image quality. First, it is informative on the distortion type

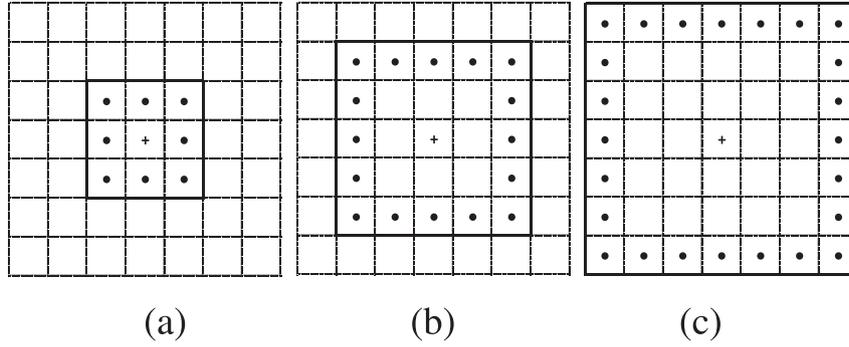


Fig. 1. Squared symmetric neighborhood for different (P, R) .

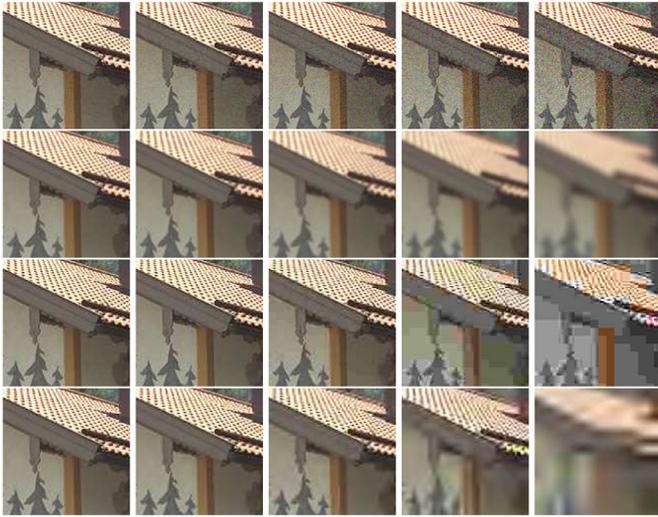


Fig. 2. Examples of image blocks with different types and levels of distortions. The distortion types for the four rows are: Gaussian white noise, Gaussian blur, JPEG compression, JPEG 2000 compression. The distortion levels increase from left to right. (Block size = 128×128).

and images with different distortions have characteristic variation. Generally, Gaussian white noise (WN) changes the distribution to a more uniform one with larger standard deviation. Gaussian blur (GB), JPEG and JPEG 2000 (JP2K) generate a higher peaked distribution. GB and JP2K exhibit a smooth distribution, while JPEG has a jaggy distribution which may result from the blocking artifacts. Second, it is informative on distortion degree. For the four distortion types considered here, we observe that the change of the distribution is a monotonic function of the subjective scores. For example, WN increases the standard deviation of the distribution. The larger the standard deviation, the lower the subjective quality

score. Similar observations can be made on other three distortion types. Considering the effectiveness of the first principal components of joint log-contrast response to represent image quality, we assume that more quality-awareness can be achieved if we model the joint log-contrast response as a whole.

Therefore, we propose one comprehensive NSS model, to represent the joint distribution of directional log-contrast response of natural images. Specifically, we use the GMM to approximate such joint distribution, and learn the GMM model using high-quality natural images, which represents the average joint distribution of pristine natural images.

Given a feature space $\mathcal{F} \in \mathcal{R}^D$, a GMM is defined as

$$p(\mathbf{x}|\theta) = \sum_{k=1}^K \omega_k p_k(\mathbf{x}|\theta) \tag{2}$$

$$p_k(\mathbf{x}|\theta) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_k|}} e^{-\frac{1}{2}(\mathbf{x}-\mu_k)\Sigma_k^{-1}(\mathbf{x}-\mu_k)^T} \tag{3}$$

with prior probabilities $\omega_k \in \mathcal{R}$, means $\mu_k \in \mathcal{R}^D$ and covariance matrices $\Sigma_k \in \mathcal{R}^{D \times D}$ ($\theta = \{\mu, \Sigma\}$). The GMM distribution is a linear combination of Gaussian densities, with ω_k as the probabilistic contribution of the k th component.

We extract a large set of log-contrast vector \mathbf{z}_c from pristine natural images, then the GMM parameters are learned in an unsupervised manner using the Maximum Likelihood (MLP) criterion. For computational simplicity, we use GMM with diagonal covariances, and apply PCA to the feature vector before GMM learning.

2.2. Quality-aware feature extraction

The learned joint NSS model represents the statistical regularities of pristine natural images, and based on this NSS model, we can extract the quality-aware features for each image. The traditional way to extracting features from NSS model is to fit the NSS model for each distorted image and use the model parameters as

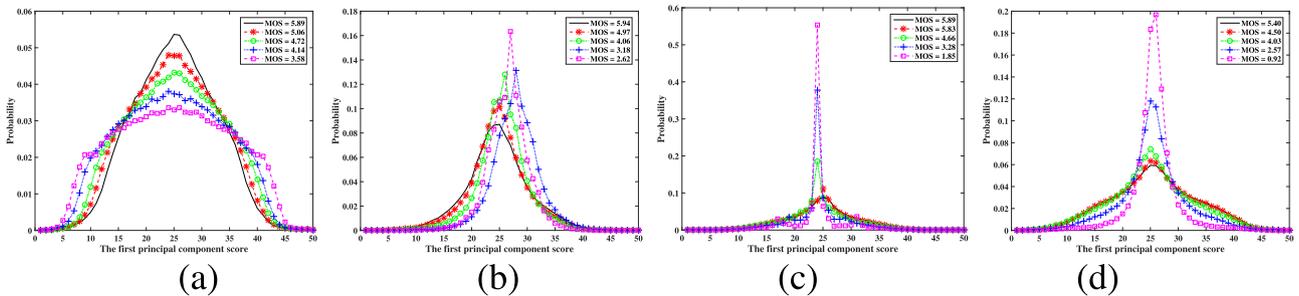


Fig. 3. The first principal components of joint log-contrast responses of degraded images. (a) WN. (b) GB. (c) JPEG. (d) JP2K. MOS is the subjective quality score of each image.

features for quality assessment. It has the advantage of compactness with only several parameters, but it may produce an inaccurate representation due to the fitting errors, especially in the case where the NSS model is not good enough for distorted images. When things come to multivariate distribution, two additional problems are raised. First, fitting a multivariate distribution would require more samples than univariate distribution as the curse of dimensionality. Feature samples extracted from one image may not be sufficient to produce a stable model. Second, training the GMM model for each image would be quite time-consuming, especially when the components of GMM increase. Inspired from the great success of FV in image classification, retrieval or matching applications [34,35], we have introduced it in BIQA framework to describe the deviation of distorted images from NSS model.

The idea of FV encoding is to characterize an image with a gradient vector derived from a generative probability model. Let $\mathbf{X} = \{\mathbf{x}_i, i = 1, \dots, N\}$ be a set of log-contrast features extracted from an image, $\mathbf{x}_i \in \mathcal{R}^D$, where N is the number of features and D is the feature dimension. The log-likelihood of \mathbf{x}_i generated from the GMM is represented as:

$$\mathcal{L}(\mathbf{x}_i) = \log \sum_{k=1}^K \omega_k p_k(\mathbf{x}_i | \theta). \quad (4)$$

Assuming that N features from one image are independent, the log-likelihood of \mathbf{X} is then given by

$$\mathcal{L}(\mathbf{X}) = \sum_{i=1}^N \log \left(\sum_{k=1}^K \omega_k p_k(\mathbf{x}_i | \theta) \right). \quad (5)$$

Fisher encoding [34] computes the gradient of the log-likelihood of the N feature vectors from one image with respect to the parameters of GMM model

$$\nabla \mathcal{L}(\mathbf{X}) = \frac{1}{N} \sum_{i=1}^N \nabla \log \left(\sum_{k=1}^K \omega_k p_k(\mathbf{x}_i | \theta) \right). \quad (6)$$

Previous works observed that the gradient with respect to the weight parameter ω_k does not provide significant information [35]. Thus, the gradient is calculated with respect to the mean μ_k and the covariance Σ_k (variance vector σ_k^2) of GMM. These gradients can be derived as follows [34]:

$$\mathcal{G}_k^{\mu,d} = \frac{1}{N\sqrt{\omega_k}} \sum_{i=1}^N \gamma_{ik} \left(\frac{x_i^d - \mu_k^d}{\sigma_k^d} \right) \quad (7)$$

$$\mathcal{G}_k^{\sigma,d} = \frac{1}{N\sqrt{2\omega_k}} \sum_{i=1}^N \gamma_{ik} \left(\frac{(x_i^d - \mu_k^d)^2}{(\sigma_k^d)^2} - 1 \right) \quad (8)$$

where γ_{ik} is the chance that local descriptor \mathbf{x}_i is generated by the k th Gaussian component using Bayes' formula

$$\gamma_{ik} = \frac{\omega_k p_k(\mathbf{x}_i | \theta)}{\sum_{j=1}^K \omega_j p_j(\mathbf{x}_i | \theta)}. \quad (9)$$

The final FV representation for one image \mathbf{X} is then the concatenation of $\mathcal{G}_k^{\mu,d}$ and $\mathcal{G}_k^{\sigma,d}$.

Given the generative GMM model that represents the average distribution of natural images, FV describes how the set of descriptors extracted from one image deviate from the NSS model. It represents how the NSS model parameters should be modified to fit the features from the image. Therefore, it measures the statistical differences between one image and the universal NSS model.

With the increase of GMM components, Fisher vectors become sparser. [35] uses the dimension wise power normalization to unsparsify the features and improve the performance:

$$f(v) = \text{sign}(v)|v|^\alpha \quad (10)$$

where $0 \leq \alpha \leq 1$ is the normalization parameter, and v is one dimension of the feature vector. Similar processing is used in this work and the parameter α is determined by experimental validation.

2.3. Regression module

FV encoding generates a representation of length $2KD$, where K is the number of GMM components and D is the dimension of local features. We adopt PLS method as the regression module to relate extracted quality-aware features to objective scores. PLS models relationships between independent variables and response variables by means of latent components. It was originally developed in chemometrics and has received much attention in other scientific areas, such as bioinformatics, neuroscience, social science, and computer vision [36,37]. It is known to be particularly suited for highly correlated, noisy data with high dimensionality [38].

Let $\mathcal{X} \in \mathcal{R}^m$ denotes an m -dimensional space of input variables and $\mathcal{Y} \in \mathcal{R}^n$ represents an n -dimensional space of response variables. PLS models the relationship between these two data sets by means of latent components. Let the number of samples in the data sets be N , PLS decomposes the zero-mean variables $\mathbf{X} \in \mathcal{R}^{N \times m}$ and $\mathbf{Y} \in \mathcal{R}^{N \times n}$ into

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \quad (11)$$

$$\mathbf{Y} = \mathbf{U}\mathbf{Q}^T + \mathbf{F} \quad (12)$$

where $\mathbf{T} \in \mathcal{R}^{N \times p}$ and $\mathbf{U} \in \mathcal{R}^{N \times p}$ are matrices of the p score vectors, $\mathbf{P} \in \mathcal{R}^{m \times p}$ and $\mathbf{Q} \in \mathcal{R}^{n \times p}$ are matrices of loadings, and $\mathbf{E} \in \mathcal{R}^{N \times m}$ and $\mathbf{F} \in \mathcal{R}^{N \times n}$ are matrices of residuals. The classical form of PLS method uses the nonlinear iterative partial least squares (NIPALS) algorithm to iteratively decompose \mathbf{X} and \mathbf{Y} to find the weight vectors \mathbf{w} , \mathbf{c} such that

$$\begin{aligned} [\text{cov}(\mathbf{t}, \mathbf{u})]^2 &= [\text{cov}(\mathbf{X}\mathbf{w}, \mathbf{Y}\mathbf{c})]^2 \\ &= \max_{|\mathbf{r}|=|\mathbf{s}|=1} [\text{cov}(\mathbf{X}\mathbf{r}, \mathbf{Y}\mathbf{s})]^2 \end{aligned} \quad (13)$$

where \mathbf{t} and \mathbf{u} are the column vectors of matrices \mathbf{T} and \mathbf{U} , and $\text{cov}(\mathbf{t}, \mathbf{u}) = \mathbf{t}^T \mathbf{u} / N$ is the sample covariance between latent vectors \mathbf{t} and \mathbf{u} . After the extraction of the latent vectors \mathbf{t}_1 and \mathbf{u}_1 , the matrices \mathbf{X} and \mathbf{Y} are deflated by subtracting their rank-one approximations based on \mathbf{t}_1 and \mathbf{u}_1 . This process is repeated until the residuals are small enough or the desired number of latent vectors is obtained. More details on PLS analysis can be referred to [38].

We use $\mathcal{X} \in \mathcal{R}^m$ to represent the feature space of extracted quality-aware features (i.e., FV representation), and $\mathcal{Y} \in \mathcal{R}$ to denote the quality score space of images.

3. Experiments

3.1. Experiment protocol

3.1.1. Database description

The proposed method is examined on four large-scale image databases (i.e., LIVE [29], CSIQ [30], TID2013 [31], CLIVE [32]). LIVE, CSIQ and TID2013 databases include images degraded by simulated distortions, while CLIVE database contains images degraded by realistic distortions. These databases are summarized as follows:

- The LIVE image database consists of 779 distorted images with five distortion types: JP2K, JPEG, WN, GB, and FF (simulated fast fading Rayleigh channel). There are 29 reference images. Subjective ratings are DMOS (Differential Mean Opinion Score) in the range of 0 to 100.

- The CSIQ image database consists of 866 distorted images with six distortion types: JPEG, JP2K, WN, GB, pink Gaussian noise (PGN), and global contrast decrements (CTD). There are 30 reference images. Subjective ratings are MOS (Mean Opinion Score) in the range of 0 to 1.
- The TID2013 database consists of 3000 distorted images with 24 distortion types: #01 WN, #02 WN in color components, #03 spatially correlated WN, #04 masked noise, #05 high-frequency noise, #06 impulse noise, #07 quantization noise, #08 GB, #09 image denoising, #10 JPEG, #11 JP2K, #12 JPEG transmission errors, #13 JP2K transmission errors, #14 non eccentricity pattern noise, #15 local blockwise distortion of different intensity, #16 mean shift, #17 contrast change, #18 change of color saturation, #19 multiplicative Gaussian noise, #20 comfort noise, #21 lossy compression of noisy images, #22 image color quantization with dither, #23 chromatic aberrations and #24 sparse sampling and reconstruction. There are 25 reference images. Subjective ratings are MOS in the range of 0 to 9.
- The CLIVE (LIVE in the wild image quality challenge database) image database consists of 1162 distorted images with realistic distortions, such as low-light blur and noise, motion blur, overexposure, underexposure, compression errors and their combination. Subjective ratings are MOS in the range of 0 to 100.

3.1.2. Performance evaluation criteria

The performance of BIQA methods can be evaluated by four criteria computed between predicted scores and subjective scores: Spearman rank order correlation coefficient (SRCC) and Kendall rank order correlation coefficient (KRCC) for prediction monotonicity, Pearson linear correlation coefficient (PLCC) for prediction accuracy and root mean squared error (RMSE) for prediction accuracy. The latter two are computed after the monotonic logistic mapping between objective and subjective scores [39].

$$f(x) = \beta_1 \left(\frac{1}{2} - \frac{1}{\exp(\beta_2(x - \beta_3))} \right) + \beta_4 x + \beta_5 \quad (14)$$

where x is the score predicted by IQA metric, $f(x)$ is the fitted score, $\beta_j (j = 1, 2, \dots, 5)$ are regression parameters determined by minimizing the errors between objective and subjective scores.

3.2. Implementation details

Since image correlation is related to image scale, we resize each image before feature extraction, to make the larger image side equals 512. Similar operation has also been adopted by previous method [28]. We use the neighborhood setting ($R = 1, P = 8$) to calculate the log-contrast vectors, as shown in Fig. 1 (a). We fix the number of GMM components as 512, and the power normalization parameter α as 0.25. To construct the NSS model that represents the statistical regularities of natural images, we use the collection of pristine naturalistic images released by [28] to train the GMM model.

3.3. Parameter choice

There are a few parameters in the proposed BJLC method. Here, we examine how the performance varies with different parameters. Specifically, we examine three parameters, the neighborhood setting (R, P), the number of GMM components K , and the power normalization parameter α . We report the median SRCC values with different parameters on each database, with 80% of the distorted images are used for training, and the remaining 20% for testing. The train and test sets are content disjoint. And the train-test split is repeated 1000 times.

3.3.1. The neighborhood setting

First, we examine how the performance of BJLC varies with different neighborhood settings. Three square neighborhood settings are tested, as shown in Fig. 1. The first setting is ($R = 1, P = 8$), where 8 neighbouring pixels on the border side of 3×3 block are employed. The second setting is ($R = 2, P = 16$), where 16 neighbouring pixels on the border side of 5×5 block are employed. The third setting is ($R = 3, P = 24$), where 24 neighbouring pixels on the border side of 7×7 block are employed. From Fig. 4 (a), we observe that increasing the spatial resolution of the square neighborhood does not provide significant performance improvement. With a larger neighborhood, the directional resolution increases as more directions are considered in the log-contrast feature vector. However, the spatial correlation becomes weak as the spatial distance for calculating the grey-level differences increases. It may counteract the benefits of increased directional resolution. As larger neighborhood induces the computational overhead, we adopt ($R = 1, P = 8$) in the proposed method for computational simplicity and performance accuracy.

3.3.2. The number of GMM components

The number of GMM components is a key factor to the performance. As the number of GMM components increases, the feature dimension increases and more discriminative information is retained in the extracted features. However, it would also induce more memory and computational costs. In Fig. 4 (b), we show the performance measured by SRCC on four benchmark databases with different GMM sizes. From this Figure, we observe that the performance of the proposed method increases with the number of GMM components on all the four databases, but the performance gradually saturates as the increase of GMM components. Therefore, to achieve a tradeoff between performance and computational complexity, we fix the number of GMM components as 512 in the following experiments.

3.3.3. The power normalization parameter

The power normalization parameter $0 \leq \alpha \leq 1$ is used to unsparify the FV feature and improve the regression performance. When α equals 0, only the sign information is preserved; when α is 1, there would be no processing on the FV feature. From Fig. 4 (c), we choose $\alpha = 0.25$ as a reasonable value for all the databases.

3.4. Overall performance on individual database

We examine the performance of BIQA methods on overall database. For each database, We split it into train and test sets with 80% for training and 20% for testing. For LIVE, CSIQ and TID2013, the split is conducted according to reference images to make sure that the two sets are content disjoint. For CLIVE, the database is split randomly to the train and test sets. The split is repeated 1000 times and the median performance is reported. The 25th reference image and its distorted versions are excluded from TID2013 as it is not a natural image.

BJLC is compared with ten state-of-the-art BIQA models, including NIQE [40], ILNIQE [28], DIIVINE [17], BLIINDS2 [22], CORNIA [41], BRISQUE [23], GMLOG [27], NFERM [42], NRSL [43] and FRIQUEE [25]. Among them, ILNIQE and FRIQUEE are color metrics, and other methods work on the luminance channel only. Both NIQE and ILNIQE are opinion-free BIQA models, which requires no training on the database, and we report the median performance of 1000 trails on the test set for consistent comparison.

Table 1 lists the SRCC and KRCC values of examined BIQA methods on four databases to show their prediction monotonicity. Table 2 lists the PLCC and RMSE values to show their prediction accuracy. These two tables show that BJLC achieves better quality

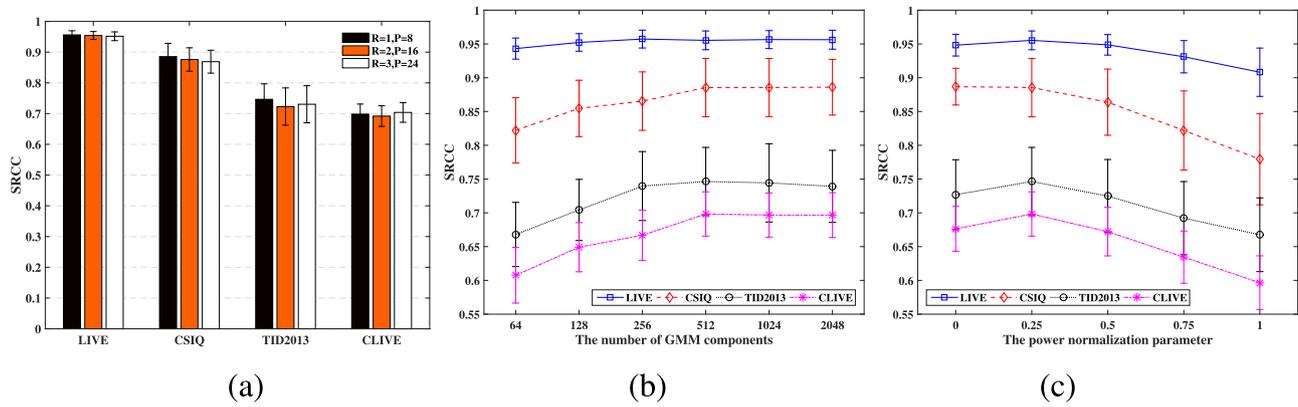


Fig. 4. The SRCC performance with different parameter settings. (a) SRCC values with different square neighborhood settings. (b) SRCC values with different number of GMM components. (c) SRCC values with different power normalization parameters.

Table 1

SRCC and KRCC comparison on four benchmark databases. The best method is shown in boldface.

BIQA model	LIVE (779)		CSIQ (866)		TID2013 (2880)		CLIVE (1162)		Weighted average	
	SRCC	KRCC	SRCC	KRCC	SRCC	KRCC	SRCC	KRCC	SRCC	KRCC
NIQE [40]	0.9084	0.7333	0.6299	0.4665	0.3238	0.2208	0.4496	0.3066	0.4777	0.3473
ILNIQE [28]	0.9024	0.7224	0.8209	0.6323	0.5251	0.3730	0.4393	0.2990	0.6053	0.4462
DIIVINE [17]	0.9120	0.7487	0.7594	0.5718	0.6735	0.4947	0.5969	0.4213	0.7043	0.5270
BLIINDS2 [22]	0.9298	0.7754	0.7528	0.5652	0.5723	0.4137	0.4626	0.3227	0.6274	0.4688
CORNIA [41]	0.9452	0.7953	0.7325	0.5464	0.6542	0.4770	0.6173	0.4400	0.6993	0.5246
BRISQUE [23]	0.9436	0.8005	0.7403	0.5590	0.5739	0.4149	0.6072	0.4291	0.6475	0.4859
GMLOG [27]	0.9498	0.8105	0.8035	0.6189	0.7100	0.5256	0.5972	0.4199	0.7372	0.5596
NFERM [42]	0.9427	0.8063	0.8213	0.6394	0.6747	0.4976	0.5404	0.3734	0.7071	0.5371
NRSL [43]	0.9517	0.8186	0.8454	0.6681	0.6706	0.4951	0.6290	0.4472	0.7281	0.5569
FRIQUEE [25]	0.9347	0.7817	0.8815	0.7077	0.6926	0.5161	0.6909	0.5012	0.7548	0.5793
BJLC	0.9561	0.8203	0.8855	0.7122	0.7490	0.5606	0.6996	0.5074	0.7886	0.6091

Table 2

PLCC and RMSE comparison on four benchmark databases. The best method is shown in boldface.

BIQA model	LIVE (779)		CSIQ (866)		TID2013 (2880)		CLIVE (1162)		Weighted average PLCC
	PLCC	RMSE	PLCC	RMSE	PLCC	RMSE	PLCC	RMSE	
NIQE [40]	0.9088	11.3756	0.7253	0.1787	0.4200	1.1249	0.5084	17.4015	0.5528
ILNIQE [28]	0.9086	11.4064	0.8160	0.1482	0.6475	0.9461	0.5126	17.3462	0.6821
DIIVINE [17]	0.9134	11.0959	0.8077	0.1546	0.7294	0.8504	0.6271	15.7668	0.7462
BLIINDS2 [22]	0.9370	9.5072	0.8134	0.1522	0.6511	0.9403	0.5072	17.4188	0.6865
CORNIA [41]	0.9473	8.7478	0.8044	0.1554	0.7451	0.8247	0.6642	15.1365	0.7660
BRISQUE [23]	0.9482	8.6605	0.8311	0.1442	0.6213	0.9668	0.6445	15.4497	0.7037
GMLOG [27]	0.9574	7.8943	0.8583	0.1344	0.7687	0.7979	0.6205	15.8003	0.7785
NFERM [42]	0.9463	8.8021	0.8658	0.1298	0.7465	0.8301	0.5699	16.6413	0.7566
NRSL [43]	0.9570	7.8790	0.8849	0.1199	0.7548	0.8185	0.6510	15.3444	0.7817
FRIQUEE [25]	0.9411	9.2061	0.9069	0.1113	0.7688	0.7965	0.7227	13.9919	0.8045
BJLC	0.9595	7.7081	0.9182	0.1041	0.8083	0.7341	0.7320	13.7830	0.8306

prediction results than other representative BIQA methods. Among them, BJLC, FRIQUEE and NRSL deliver consistent better performance on all the four databases. NFERM achieves encouraging performance on CSIQ and TID2013 databases, while its performance on CLIVE is suboptimal. By contrast, CORNIA and BRISQUE perform well on LIVE and CLIVE databases, but they perform poorly on CSIQ and TID2013 databases. ILNIQE is an effective method on CSIQ database, however, its performance on other three databases is unsatisfactory.

Fig. 5 illustrates the box plots of SRCC values across 1000 train-test trails for all the examined BIQA methods. BJLC has more compact SRCC distributions on four databases, with higher median SRCC values and smaller standard deviation, which suggests that BJLC performs more accurately and consistently than other methods. To testify the performance differences between any two methods are statistically significant, the two sample T-test is performed between SRCC obtained by competing BIQA methods from 1000

train-test trails. The null hypothesis is that the SRCC values of the methods in comparison are drawn from populations with equal means at 95% confidence level. Table 3 shows that BJLC is statistically superior to all the competing methods on four databases with only one exception, where BJLC and GMLOG performs statistically equivalent on the LIVE database.

3.5. Performance on individual distortion type

We compare the performance of examined methods on individual distortion types. The BIQA models are trained with 80% of the distorted images with specific distortion, and tested on the remaining 20% distorted images with the same distortion type. We only list the SRCC results for brevity. KRCC, PLCC and RMSE would follow similar trend. Table 4 shows the performance on LIVE and CSIQ databases, and Table 5 shows the performance on TID2013 database. Since images degraded by realistic distortions have no

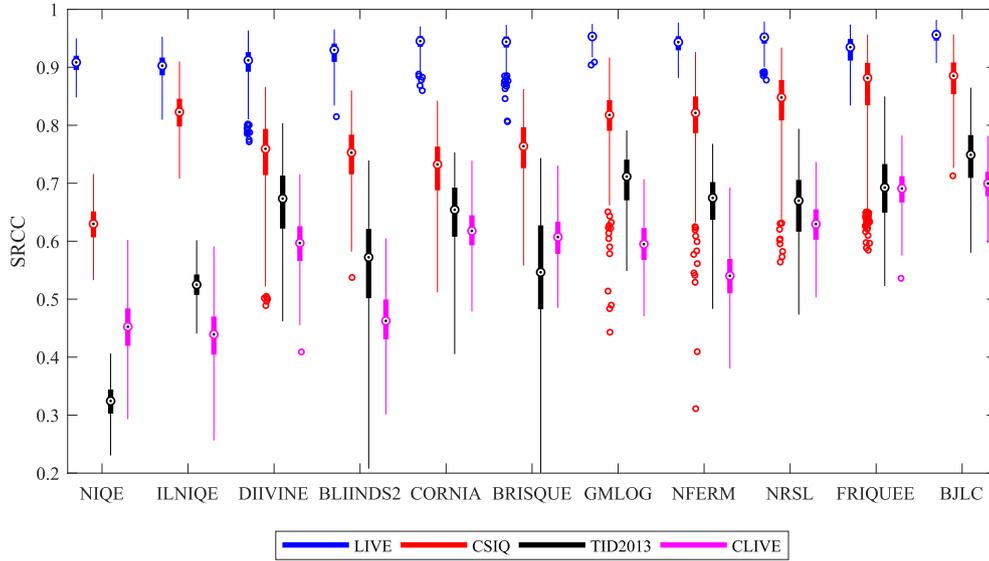


Fig. 5. The box plots of SRCC across 1000 train-test trails on four databases. The upper and lower edges of the rectangles indicate the first and third quartiles. The black dot inside the circle is the median SRCC of 1000 trials.

Table 3

Statistical significance test (t-test) on four databases. \uparrow (\downarrow) indicates the row method is statistically better (worse) than the column method at 95% confidence level. 0 indicates that the row and column methods are statistically equivalent. The four symbols in each table cell represent the results on four databases, i.e., LIVE, CSIQ, TID2013, CLIVE.

	NIQE	ILNIQE	DIIVINE	BLIINDS2	CORNIA	BRISQUE	GMLOG	NFERM	NRSL	FRIQUEE	BJLC
NIQE	0000	$\uparrow\downarrow\downarrow$	$0\downarrow\downarrow$	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$
ILNIQE	$\downarrow\downarrow\uparrow$	0000	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$
DIIVINE	$0\uparrow\uparrow\uparrow$	$\uparrow\downarrow\uparrow\uparrow$	0000	$\downarrow\downarrow\uparrow$	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$
BLIINDS2	$\uparrow\uparrow\uparrow\uparrow$	$\uparrow\downarrow\uparrow\uparrow$	$\uparrow\downarrow\downarrow$	0000	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$
CORNIA	$\uparrow\uparrow\uparrow\uparrow$	$\uparrow\downarrow\uparrow\uparrow$	$\uparrow\downarrow\uparrow$	$\uparrow\downarrow\uparrow\uparrow$	0000	$0\downarrow\uparrow\uparrow$	$\downarrow\downarrow\uparrow$	$\uparrow\downarrow\uparrow$	$\downarrow\downarrow\downarrow$	$\uparrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$
BRISQUE	$\uparrow\uparrow\uparrow\uparrow$	$\uparrow\downarrow\uparrow\uparrow$	$\uparrow\uparrow\downarrow$	$\uparrow\uparrow\downarrow$	$0\uparrow\downarrow$	0000	$\downarrow\downarrow\uparrow$	$0\downarrow\downarrow$	$\downarrow\downarrow\downarrow$	$\uparrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$
GMLOG	$\uparrow\uparrow\uparrow\uparrow$	$\uparrow\downarrow\uparrow\uparrow$	$\uparrow\uparrow\downarrow$	$\uparrow\uparrow\uparrow\uparrow$	$\uparrow\uparrow\downarrow$	$\uparrow\uparrow\downarrow$	0000	$\uparrow\downarrow\uparrow$	$\uparrow\downarrow\downarrow$	$\uparrow\downarrow\downarrow$	$0\downarrow\downarrow$
NFERM	$\uparrow\uparrow\uparrow\uparrow$	$\uparrow\downarrow\uparrow\uparrow$	$\uparrow\uparrow\downarrow$	$\uparrow\uparrow\uparrow\uparrow$	$\uparrow\uparrow\downarrow$	$0\uparrow\downarrow$	$\downarrow\downarrow\downarrow$	0000	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$
NRSL	$\uparrow\uparrow\uparrow\uparrow$	$\uparrow\uparrow\uparrow\uparrow$	$\uparrow\uparrow\downarrow$	$\uparrow\uparrow\uparrow\uparrow$	$\uparrow\uparrow\uparrow\uparrow$	$\uparrow\uparrow\uparrow\uparrow$	$\downarrow\downarrow\uparrow$	$\uparrow\downarrow\uparrow$	0000	$\uparrow\downarrow\downarrow$	$\downarrow\downarrow\downarrow$
FRIQUEE	$\uparrow\uparrow\uparrow\uparrow$	$\uparrow\uparrow\uparrow\uparrow$	$\uparrow\uparrow\uparrow\uparrow$	$\uparrow\uparrow\uparrow\uparrow$	$\downarrow\downarrow\uparrow$	$\downarrow\downarrow\uparrow$	$\downarrow\downarrow\uparrow$	$\downarrow\downarrow\uparrow$	$\downarrow\downarrow\uparrow$	0000	$\downarrow\downarrow\downarrow$
BJLC	$\uparrow\uparrow\uparrow\uparrow$	$\uparrow\uparrow\uparrow\uparrow$	$\uparrow\uparrow\uparrow\uparrow$	$\uparrow\uparrow\uparrow\uparrow$	$\uparrow\uparrow\uparrow\uparrow$	$\uparrow\uparrow\uparrow\uparrow$	$0\uparrow\uparrow\uparrow$	$\uparrow\uparrow\uparrow\uparrow$	$\uparrow\uparrow\uparrow\uparrow$	$\uparrow\uparrow\uparrow\uparrow$	0000

Table 4

SRCC comparison of BIQA methods on individual distortion type for LIVE and CSIQ.

BIQA model	LIVE (779)					CSIQ (866)					
	JP2K	JPEG	WN	GB	FF	WN	JPEG	JP2K	PGN	GB	CTD
NIQE [40]	0.9239	0.9423	0.9715	0.9395	0.8616	0.8366	0.8826	0.9267	0.3275	0.9062	0.2786
ILNIQE [28]	0.9002	0.9439	0.9791	0.9244	0.8443	0.8681	0.9043	0.9239	0.8835	0.8676	0.5447
DIIVINE [17]	0.9151	0.9142	0.9813	0.9377	0.8714	0.9404	0.8884	0.8706	0.8038	0.8563	0.7661
BLIINDS2 [22]	0.9324	0.9519	0.9435	0.9321	0.8821	0.9223	0.9519	0.9342	0.7818	0.9249	0.4150
CORNIA [41]	0.9329	0.9456	0.9791	0.9600	0.9023	0.9498	0.9066	0.9097	0.7717	0.9110	0.6477
BRISQUE [23]	0.9086	0.9639	0.9853	0.9493	0.8917	0.9604	0.9404	0.8838	0.8776	0.9104	0.5306
GMLOG [27]	0.9286	0.9711	0.9849	0.9368	0.9128	0.9435	0.9168	0.9137	0.8652	0.9123	0.7446
NFERM [42]	0.9409	0.9672	0.9853	0.9479	0.8278	0.9459	0.9346	0.9137	0.9008	0.9270	0.6522
NRSL [43]	0.9293	0.9580	0.9835	0.9515	0.8674	0.9509	0.9469	0.9099	0.8937	0.9160	0.8019
FRIQUEE [25]	0.9020	0.9553	0.9826	0.9586	0.8770	0.9453	0.9340	0.8910	0.9277	0.9413	0.8687
BJLC	0.9588	0.9734	0.9862	0.9542	0.9155	0.9622	0.9501	0.9474	0.9314	0.9301	0.8727

definite distortion category, we exclude CLIVE database in this experiment.

From Tables 4 and Table 5, we find that among the 35 distortion groups from the three databases, BJLC ranks the best 26 times. For other distortion groups, BJLC performs similarly with the best ones. Such superior performance indicates that BJLC is also a good distortion-specific BIQA method. All BIQA methods fail on three distortion types, i.e., #14 non eccentricity pattern noise, #15 local blockwise distortion of different intensity, #16 mean shift, with SRCC less than 0.6. This may be attributed to the characteristics of

these distortions. #14 and #16 are non-structural distortions; #15 is localized distortion.

3.6. Cross-database experiment

Cross-database experiment is performed to show the generalization capability of examined BIQA methods. These methods are first trained using the whole LIVE database and examined on CSIQ and TID2013 databases. The common four distortions shared by these databases are used for testing (i.e., JP2K, JPEG, WN, GB). From

Table 5
SRCC comparison of BIQA methods on individual distortion type for TID2013.

BIQA model	TID2013 (2880)											
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
NIQE [40]	0.8492	0.6976	0.7155	0.8100	0.8613	0.8162	0.8746	0.8385	0.6577	0.8808	0.9077	0.0104
ILNIQE [28]	0.9034	0.8288	0.9377	0.6962	0.8843	0.8299	0.8867	0.8646	0.8019	0.8832	0.9209	0.3577
DIIVINE [17]	0.9063	0.8235	0.8896	0.7500	0.9271	0.9174	0.6376	0.9400	0.8192	0.8262	0.8923	0.6238
BLIINDS2 [22]	0.7415	0.5731	0.5696	0.5485	0.7537	0.7353	0.6185	0.8600	0.7419	0.7758	0.8850	0.6428
CORNIA [41]	0.9377	0.8907	0.8508	0.8484	0.9120	0.9200	0.9236	0.9246	0.8326	0.8916	0.9047	0.7055
BRISQUE [23]	0.9000	0.8244	0.8903	0.6226	0.9131	0.8512	0.8023	0.8969	0.7403	0.8695	0.9039	0.4938
GMLOG [27]	0.9454	0.8846	0.8887	0.7755	0.9386	0.8922	0.9022	0.9292	0.8700	0.9238	0.9277	0.5981
NFERM [42]	0.9314	0.8631	0.9162	0.7799	0.9377	0.8584	0.8615	0.9268	0.8331	0.9116	0.9431	0.5692
NRSL [43]	0.9127	0.8277	0.9138	0.7977	0.9452	0.9108	0.8212	0.9423	0.8541	0.8931	0.9292	0.6287
FRIQUEE [25]	0.9500	0.9085	0.9562	0.8415	0.9577	0.9285	0.9222	0.9446	0.8723	0.9127	0.9043	0.6219
BJLC	0.9585	0.8938	0.9227	0.8587	0.9609	0.9398	0.9448	0.9685	0.9252	0.9462	0.9531	0.7785
BIQA model	TID2013 (2880)											
	#13	#14	#15	#16	#17	#18	#19	#20	#21	#22	#23	#24
NIQE [40]	0.6004	-0.1762	-0.0465	-0.1335	0.0433	-0.2250	0.7503	0.1662	0.8431	0.8095	0.7131	0.8738
ILNIQE [28]	0.6600	-0.1772	-0.0569	0.2512	0.0623	-0.1187	0.7438	0.4285	0.8820	0.7977	0.8169	0.9054
DIIVINE [17]	0.8189	0.2528	-0.0015	0.0873	0.5895	0.0596	0.8660	0.6256	0.8125	0.8492	0.8145	0.8534
BLIINDS2 [22]	0.5654	0.2583	0.2362	0.2191	0.2696	-0.0762	0.7834	0.5645	0.6892	0.7454	0.7508	0.8854
CORNIA [41]	0.7905	0.4791	-0.1208	0.2046	0.5903	0.2912	0.9046	0.7665	0.9131	0.8742	0.8222	0.8965
BRISQUE [23]	0.7227	0.2960	0.4572	0.2262	0.3529	0.3131	0.8435	0.4296	0.7495	0.8738	0.7657	0.8789
GMLOG [27]	0.7962	0.1683	0.3898	0.3615	0.6581	0.3738	0.8423	0.6963	0.8485	0.9162	0.7638	0.9223
NFERM [42]	0.7262	0.1585	0.2462	0.2742	0.6633	0.2577	0.8546	0.5403	0.8292	0.8677	0.8484	0.9085
NRSL [43]	0.7885	0.1787	0.3542	0.1398	0.5896	0.1729	0.8737	0.7409	0.8543	0.9101	0.8616	0.9054
FRIQUEE [25]	0.8364	0.2306	0.3485	0.3956	0.7530	0.7677	0.9345	0.5677	0.9323	0.9069	0.8731	0.9077
BJLC	0.8726	0.4719	0.5467	0.4469	0.7173	0.4642	0.9292	0.7830	0.9377	0.9423	0.8959	0.9562

Table 6
SRCC comparison on cross-database validation.

DB:DT	NIQE	ILNIQE	DIIVINE	BLIINDS2	CORNIA	BRISQUE	GMLOG	NFERM	NRSL	FRIQUEE	BJLC
CSIQ:JP2K	0.9062	0.9062	0.8582	0.8644	0.9125	0.8585	0.8859	0.8980	0.8788	0.8531	0.9222
CSIQ:JPEG	0.8820	0.8993	0.8707	0.9241	0.9076	0.8923	0.9153	0.9180	0.9080	0.9106	0.9423
CSIQ:WN	0.8098	0.8498	0.8763	0.8993	0.7503	0.8998	0.8963	0.9196	0.8556	0.8496	0.9327
CSIQ:GB	0.8948	0.8579	0.8797	0.8975	0.9172	0.8908	0.8907	0.8928	0.8901	0.8488	0.8995
CSIQ:ALL	0.8693	0.8800	0.8708	0.9009	0.8983	0.8879	0.8968	0.9077	0.8905	0.8754	0.9260
TID2013:JP2K	0.8980	0.9121	0.8653	0.9027	0.9033	0.9079	0.9297	0.9438	0.9503	0.9126	0.9281
TID2013:JPEG	0.8629	0.8672	0.8457	0.8575	0.8959	0.8896	0.8985	0.9148	0.9276	0.8363	0.9187
TID2013:WN	0.8163	0.8859	0.8683	0.7369	0.7364	0.8209	0.8990	0.9051	0.8460	0.8107	0.9032
TID2013:GB	0.8156	0.8349	0.8886	0.7864	0.9150	0.8724	0.9121	0.8820	0.8869	0.8657	0.8986
TID2013:ALL	0.8106	0.8768	0.8746	0.8452	0.8790	0.8776	0.9125	0.9142	0.9159	0.8712	0.9071

Table 7
PLS vs. SVR as regression models.

DB	METHOD	80%Train-20%Test		50%Train-50%Test		20%Train-80%Test	
		SRCC	Train/Test TIME(s)	SRCC	Train/Test TIME(s)	SRCC	Train/Test TIME(s)
LIVE	BJLC _{SVM}	0.9523	6.594/1.041	0.9452	2.850/1.516	0.9307	0.557/1.012
	BJLC	0.9561	0.094/0.007	0.9460	0.069/0.015	0.9310	0.034/0.025
CSIQ	BJLC _{SVM}	0.8769	5.731/1.247	0.8548	2.201/1.871	0.8236	0.377/1.216
	BJLC	0.8855	0.087/0.007	0.8555	0.057/0.018	0.8240	0.027/0.029
TID2013	BJLC _{SVM}	0.6309	86.842/13.663	0.6239	31.813/20.068	0.6027	5.053/13.322
	BJLC	0.7490	0.275/0.026	0.7062	0.181/0.060	0.6633	0.083/0.093
CLIVE	BJLC _{SVM}	0.6792	9.407/2.227	0.6751	3.760/3.365	0.6399	0.646/2.221
	BJLC	0.6996	0.103/0.008	0.6823	0.066 / 0.020	0.6404	0.035/0.036

Table 6, we can see that BJLC has good generalization ability to unseen distorted images. Among all these methods, CORNIA is most effective on GB; NFERM is most effective on WN; BJLC and NRSL are most effective on JP2K and JPEG artifacts.

3.7. Choosing PLS as the regression model

To validate the usage of PLS as the mapping module from quality-aware features to opinion score, we tested against SVM combined with the extracted features. We denote the method using SVM as BJLC_{SVM}, and it differs with the proposed BJLC method in only the regression module.

From Table 7, we can observe that PLS performs better when combined with our extracted features. Moreover, it is more efficient than SVR in both training and test stages.

3.8. Run-time comparison

To apply the BIQA model in real-world applications, the efficiency is an important factor. The run-time comparison of average feature extraction time on CSIQ database is tabulated in Table 8. Experiments are performed on a notebook with Intel Core i7-3520M CPU@2.9 GHz and 8 GB RAM. It shows that BJLC is a high

Table 8
Run-time comparison for feature extraction.

BIQA model	NIQE	ILNIQE	DIIVINE	BLIINDS2	CORNIA	BRISQUE	GMLOG	NFERM	NRSL	FRIQUEE	BJLC
Time (s)	0.227	9.840	15.519	61.393	2.449	0.079	0.063	54.025	0.101	28.025	3.023

performing BIQA method with moderate computational complexity.

4. Conclusion

Two natural image properties have been proved to be useful in perceptual quality evaluation, i.e., the non-Gaussian statistics and image correlations across space, scale and orientation. Joint NSS models that bear both of these two properties are overlooked in the BIQA research. In this work, we have developed a joint NSS model of directional log-contrast statistics to exploit these two properties. The difficulties in working with joint density can be largely alleviated by the utilization of Fisher vector for quality-aware feature extraction. Furthermore, we have developed a joint NSS based BIQA method that delivered better quality prediction performance beyond state-of-the-art competitors, and this has validated the effectiveness of applying joint NSS model in BIQA design. The proposed NSS model has captured the non-Gaussian, spatial and directional correlation properties. Nevertheless, it still lacks the modeling of inter-scale correlation, and more advanced joint NSS modes are expected to succeed in BIQA research in the future.

References

- [1] Z. Wang, A. Bovik, Reduced- and no-reference image quality assessment, *IEEE Signal Process. Mag.* 28 (6) (2011) 29–40, doi:10.1109/MSP.2011.942471.
- [2] W. Lin, C.-C. J. Kuo, Perceptual visual quality metrics: a survey, *J. Vis. Commun. Image Represent.* 22 (4) (2011) 297–312.
- [3] D. Zoran, Y. Weiss, Scale invariance and noise in natural images, in: Proceedings of the 2009 IEEE 12th International Conference on Computer Vision, 2009, pp. 2209–2216, doi:10.1109/ICCV.2009.5459476.
- [4] L. Li, Y. Zhou, W. Lin, J. Wu, X. Zhang, B. Chen, No-reference quality assessment of deblocked images, *Neurocomputing* 177 (2016) 572–584. <https://doi.org/10.1016/j.neucom.2015.11.063>.
- [5] L. Li, W. Xia, W. Lin, Y. Fang, S. Wang, No-reference and robust image sharpness evaluation based on multiscale spatial and spectral features, *IEEE Trans. Multimed.* 19 (5) (2017) 1030–1040.
- [6] L. Li, W. Lin, X. Wang, G. Yang, K. Bahrami, A.C. Kot, No-reference image blur assessment based on discrete orthogonal moments, *IEEE Trans. Cybern.* 46 (1) (2016a) 39–50, doi:10.1109/TCYB.2015.2392129.
- [7] Q. Li, W. Lin, Y. Fang, No-reference quality assessment for multiply-distorted images in gradient domain, *IEEE Signal Process. Lett.* 23 (4) (2016b) 541–545, doi:10.1109/LSP.2016.2537321.
- [8] T. Dai, K. Gu, L. Niu, Y. Bing Zhang, W. Lu, S.-T. Xia, Referenceless quality metric of multiply-distorted images based on structural degradation, *Neurocomputing* (2018), doi:10.1016/j.neucom.2018.02.050.
- [9] C. Zhang, J. Pan, S. Chen, T. Wang, D. Sun, No reference image quality assessment using sparse feature representation in two dimensions spatial correlation, *Neurocomputing* 173, Part 2 (2016) 462–470, doi:10.1016/j.neucom.2015.01.105.
- [10] Q. Wang, J. Chu, L. Xu, Q. Chen, A new blind image quality framework based on natural color statistic, *Neurocomputing* 173, Part 3 (2016) 1798–1810, doi:10.1016/j.neucom.2015.09.057.
- [11] L. Xu, J. Li, W. Lin, Y. Zhang, L. Ma, Y. Fang, Y. Yan, Multi-task rank learning for image quality assessment, *IEEE Trans. Circuits Syst. Video Technol.* 27 (9) (2017) 1833–1843.
- [12] Q. Li, W. Lin, Y. Fang, BSD: blind image quality assessment based on structural degradation, *Neurocomputing* 236 (2017) 93–103.
- [13] Q. Wu, H. Li, Z. Wang, F. Meng, B. Luo, W. Li, K.N. Ngan, Blind image quality assessment based on rank-order regularized regression, *IEEE Trans. Multimed.* 19 (11) (2017a) 2490–2504, doi:10.1109/TMM.2017.2700206.
- [14] J. Wu, J. Zeng, Y. Liu, G. Shi, W. Lin, Hierarchical feature degradation based blind image quality assessment, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017b, pp. 510–517.
- [15] H.R. Sheikh, A.C. Bovik, L. Cormack, No-reference quality assessment using natural scene statistics: JPEG2000, *IEEE Trans. Image Process.* 14 (11) (2005) 1918–1927.
- [16] A.K. Moorthy, A.C. Bovik, A two-step framework for constructing blind image quality indices, *IEEE Signal Process. Lett.* 17 (5) (2010) 513–516.
- [17] A.K. Moorthy, A.C. Bovik, Blind image quality assessment: from natural scene statistics to perceptual quality, *IEEE Trans. Image Process.* 20 (12) (2011) 3350–3364.
- [18] L. He, D. Tao, X. Li, X. Gao, Sparse representation for blind image quality assessment, in: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, pp. 1146–1153, doi:10.1109/CVPR.2012.6247795.
- [19] X. Gao, F. Gao, D. Tao, X. Li, Universal blind image quality assessment metrics via natural scene statistics and multiple kernel learning, *IEEE Trans. Neural Netw. Learn. Syst.* 24 (12) (2013) 2013–2026, doi:10.1109/TNNLS.2013.2271356.
- [20] Y. Zhang, A.K. Moorthy, D.M. Chandler, A.C. Bovik, C-DIIVINE: no-reference image quality assessment based on local magnitude and phase statistics of natural scenes, *Signal Process.: Image Commun.* 29 (7) (2014) 725–747, doi:10.1016/j.image.2014.05.004.
- [21] M.A. Saad, A.C. Bovik, C. Charrier, DCT statistics model-based blind image quality assessment, in: Proceedings of the 18th IEEE International Conference on Image Processing, 2011, pp. 3093–3096, doi:10.1109/ICIP.2011.6116319.
- [22] M.A. 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.
- [23] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, *IEEE Trans. Image Process.* 21 (12) (2012) 4695–4708.
- [24] 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.
- [25] D. Ghadiyaram, A.C. Bovik, Perceptual quality prediction on authentically distorted images using a bag of features approach, *J. Vis.* 17 (1) (2017). 32–32
- [26] D. Lee, K.N. Plataniotis, Toward a no-reference image quality assessment using statistics of perceptual color descriptors, *IEEE Trans. Image Process.* 25 (8) (2016) 3875–3889, doi:10.1109/TIP.2016.2579308.
- [27] W. Xue, X. Mou, L. Zhang, A.C. Bovik, X. Feng, Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features, *IEEE Trans. Image Process.* 23 (11) (2014) 4850–4862.
- [28] L. Zhang, L. Zhang, A. Bovik, A feature-enriched completely blind image quality evaluator, *IEEE Trans. Image Process.* 24 (8) (2015) 2579–2591, doi:10.1109/TIP.2015.2426416.
- [29] H.R. Sheikh, Z. Wang, L. Cormack, A.C. Bovik, LIVE image quality assessment database release 2, 2005. <http://live.ece.utexas.edu/research/quality>
- [30] E.C. Larson, D. Chandler, Categorical image quality (CSIQ) database, Online, <http://vision.okstate.edu/csiq/> (2010).
- [31] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, C.-C. J. Kuo, Image database TID2013: Peculiarities, results and perspectives, *Signal Processing: Image Commun.* 30 (2015) 57–77, doi:10.1016/j.image.2014.10.009.
- [32] D. Ghadiyaram, A.C. Bovik, Massive online crowdsourced study of subjective and objective picture quality, *IEEE Trans. Image Process.* 25 (1) (2016) 372–387, doi:10.1109/TIP.2015.2500021.
- [33] T. Ojala, M. Pietikäinen, T. Mäenpää, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (7) (2002) 971–987.
- [34] F. Perronnin, C. Dance, Fisher kernels on visual vocabularies for image categorization, in: Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007, pp. 1–8, doi:10.1109/CVPR.2007.383266.
- [35] F. Perronnin, J. Sánchez, T. Mensink, Improving the fisher kernel for large-scale image classification, in: Proceedings of the 2010 11th European Conference on Computer Vision (ECCV), Berlin, Heidelberg, 2010, pp. 143–156.
- [36] Q. Wang, F. Chen, W. Xu, M.H. Yang, Object tracking via partial least squares analysis, *IEEE Trans. Image Process.* 21 (10) (2012) 4454–4465, doi:10.1109/TIP.2012.2205700.
- [37] M.A. Haj, J. Gonzalez, L.S. Davis, On partial least squares in head pose estimation: how to simultaneously deal with misalignment, in: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2602–2609, doi:10.1109/CVPR.2012.6247979.
- [38] R. Rosipal, N. Krämer, Overview and recent advances in partial least squares, in: Proceedings of the Subspace, Latent Structure and Feature Selection: Statistical and Optimization Perspectives Workshop, 2006, pp. 34–51. Berlin, Heidelberg
- [39] Video Quality Experts Group. Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment, Phase II (FR_TV2), 2003. ftp://ftp.its.bldrdoc.gov/dist/ituidvq/Boulder_VQEG_jan_04/VQEG_PhaseII_FRTV_Final_Report_SG9060E.doc
- [40] A. Mittal, R. Soundararajan, A. Bovik, Making a completely blind image quality analyzer, *IEEE Signal Process. Lett.* 20 (3) (2013) 209–212, doi:10.1109/LSP.2012.2227726.
- [41] P. Ye, J. Kumar, L. Kang, D. Doermann, Unsupervised feature learning framework for no-reference image quality assessment, in: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2012, pp. 1098–1105.

- [42] K. Gu, G. Zhai, X. Yang, W. Zhang, Using free energy principle for blind image quality assessment, *IEEE Trans. Multimed.* 17 (1) (2015) 50–63, doi:[10.1109/TMM.2014.2373812](https://doi.org/10.1109/TMM.2014.2373812).
- [43] Q. Li, W. Lin, J. Xu, Y. Fang, Blind image quality assessment using statistical structural and luminance features, *IEEE Trans. Multimed.* 18 (12) (2016) 2457–2469, doi:[10.1109/TMM.2016.2601028](https://doi.org/10.1109/TMM.2016.2601028).



Qiaohong Li received the B.E. degree and M.E. degree in School of Information and Communication Engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 2009 and 2012, and the Ph.D. degree from Nanyang Technological University, Singapore, in 2017. She is currently a Research Fellow with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. Her research interests include multimedia quality assessment, visual perceptual modelling, and machine learning.



Weisi Lin (M'92-SM'98-F'16) received the Ph.D. degree from King's College London, U.K. He served as the Lab Head of visual processing, Institute for Infocomm Research, Singapore. He is currently an Associate Professor with the School of Computer Engineering. His technical expertise includes perceptual modeling and evaluation of multimedia signals, image processing, and video compression, in which he has published 160 journal papers and 230 conference papers, filed seven patents, authored two books, edited three books, and written nine book chapters. He is a Chartered Engineer, a fellow of IET, and an Honorary Fellow of the Singapore Institute of Engineering Technologists. He has been a Technical

Program Chair of IEEE ICME 2013, PCM 2012, QoMEX 2014, and VCIP 2017. He chaired the IEEE MMTC Special Interest Group on QoE (2012–2014). He has been a keynote/invited/panelist/tutorial speaker in over 20 international conferences, as well as a Distinguished Lecturer of the Asia-Pacific Signal and Information Processing Association from 2012 to 2013 and the IEEE Circuits and Systems Society from 2016 to 2017. He served as a guest editor for seven special issues in different scholarly journals. He was an Associate Editor (AE) of the *IEEE TRANSACTIONS ON MULTIMEDIA* and the *IEEE SIGNAL PROCESSING LETTERS*. He is an AE of the *IEEE TRANSACTIONS ON IMAGE PROCESSING*, the *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY*, and the *Journal of Visual Communication and Image Representation*.



Ke Gu received the B.S. and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009 and 2015, respectively. He has reviewed over 50 journal papers each year. His research interests include quality assessment, contrast enhancement, visual saliency detection, and air quality prediction. He received the Best Paper Award at the IEEE International Conference on Multimedia and Expo in 2016, and received the excellent Ph.D. thesis award from the Chinese Institute of Electronics in 2016. He is the leading special session organizer in VCIP2016 and ICIP2017. He is currently an Associated Editor of the *IEEE ACCESS*, and is the Reviewer of the *IEEE T-NNLS*, *T-IP*, *T-MM*, *TCYB*, *T-CSVT*, *T-IE*, *T-BC*, *J-STSP*, *SPL*, *Access*, *Information Sciences*, *Neurocomputing*, *SPIC*, *JVCI*, *DSP*, *MTAP*, *ELL*, and so on.



Yabin Zhang received the B.E. degree in Electronic Information Engineering in the Honors School, Harbin Institute of Technology and the Ph.D. degree from the School of Computer Science and Engineering, Nanyang Technological University, Singapore in 2013 and 2018, respectively. He is currently a senior researcher in Media Lab, Tencent, Shenzhen. His research interests include video coding, image/video processing, image quality assessment and computer vision.



Yuming Fang received the B.E. degree from Sichuan University, Chengdu, China, the M.S. degree from the Beijing University of Technology, Beijing, China, and the Ph.D. degree from Nanyang Technological University, Singapore. He is currently a Professor with the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, China. He has authored and co-authored over 90 academic papers in international journals and conferences in the areas of multimedia processing. His research interests include visual attention modeling, visual quality assessment, image retargeting, computer vision, 3-D image/video processing, and so on. He serves as an Associate Editor for the *IEEE ACCESS*, and he is on the Editorial Board of *Signal Processing: Image Communication*.