No-reference image quality assessment using statistical characterization in the shearlet domain

Article *in* Signal Processing Image Communication · August 2014

DOI: 10.1016/j.image.2014.05.007

CITATIONS
32

READS
326

4 authors, including:

Yuming Li
City University of Hong Kong
21 PUBLICATIONS **521** CITATIONS
SEE PROFILE

Lai Man Po
City University of Hong Kong
190 PUBLICATIONS **4,755** CITATIONS
SEE PROFILE

Xuyuan Xu
City University of Hong Kong
32 PUBLICATIONS **603** CITATIONS
SEE PROFILE

Signal Processing: *Image Communication* 29 (2014) 748–759

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

No-reference image quality assessment using statistical characterization in the shearlet domain

Yuming Li *, Lai-Man Po, Xuyuan Xu, Litong Feng

Department of Electronic Engineering, City University of Hong Kong, Hong Kong Special Administrative Region, China

ARTICLE INFO

Article history:
Received 4 January 2014
Received in revised form
26 April 2014
Accepted 28 May 2014
Available online 11 June 2014

Keywords:
No-reference image quality assessment
Shearlet transform
Sparse autoencoder
Softmax classification

ABSTRACT

Image and video quality measurements are crucial for many applications, such as acquisition, compression, transmission, enhancement, and reproduction. Nowadays, no-reference (NR) image quality assessment (IQA) methods have drawn extensive attention because it does not rely on any information of original images. However, most of the conventional NR-IQA methods are designed only for one or a set of predefined specific image distortion types, which are unlikely to generalize for evaluating image/video distorted with other types of distortions. In order to estimate a wide range of image distortions, in this paper, we present an efficient general-purpose NR-IQA algorithm which is based on a new multiscale directional transform (shearlet transform) with a strong ability to localize distributed discontinuities. This is mainly based on distorted natural image that leads to significant variation in the spread discontinuities in all directions. Thus, the statistical property of the distorted image is significantly different from that of natural images in fine scale shearlet coefficients, which are referred to as 'distorted parts'. However, some 'natural parts' are reserved in coarse scale shearlet coefficients. The algorithm relies on utilizing the natural parts to predict the natural behavior of distorted parts. The predicted parts act as 'reference' and the difference between the reference and distorted parts is used as an indicator to predict the image quality. In order to achieve this goal, we modify the general sparse autoencoder to serve as a predictor to get the predicted parts from natural parts. By translating the NR-IQA problem into classification problem, the predicted parts and distorted parts are utilized to form features and the differences between them are identified by softmax classifier. The resulting algorithm, which we name SHeArlet based No-reference Image quality Assessment (SHANIA), is tested on several database (LIVE, Multiply Distorted LIVE and TID2008) and shown to be suitable for many common distortions, consistent with subjective assessment and comparable to full-reference IQA methods and state-of-the-art general purpose NR-IQA algorithms.

1. Introduction

Visual quality measurement is vital in many image and video processing applications. The direct way of measuring quality is to seek opinions from human observers, but it is always not practical for real-time applications. Thus, automatic quality measurement methods that are consistent with human perception are preferable for assigning quality scores to images or videos. Over these years, the most popular approaches for image quality assessment (IQA) are based on the measurements of image similarity or fidelity (e.g. [1–7]). The dependency of reference images divides

* Corresponding author.
E-mail address: yumingli4-c@my.cityu.edu.hk (Y. Li).

http://dx.doi.org/10.1016/j.image.2014.05.007
0923-5965/

the objective IQA methods into three types: full-reference (FR), reduced-reference (RR) and no-reference (NR). In FR-IQA and RR-IQA methods, the whole reference images or partial information of the reference images are assumed to be available. However, the availability of the full or partial reference image's information may not be possible or very expensive in many practical applications. For example, in order to get the best quality of image and video data, real-time image and video acquisition system needs the feedback of the IQA method to adjust system parameters. Besides, in network digital video service, IQA method is also used to check the quality of current state of the video. Based on the feedback, the transmission system can distribute data flow resources in a reasonable way. In these cases, information about the original images or videos cannot be obtained. Because of these drawbacks, the NR-IQA (or blind IQA) method has recently received a great deal of attention.

Nowadays, most of the existing NR-IQA algorithms can be broadly classified as (1) distortion-specific approaches. This type of NR-IQA algorithms usually calibrate some specific distortions, such as JPG [8], JPG2K [9]. Since this kind of NR-IQA methods usually imply some prior information about distortions, it is very hard to generalize them to other new distortion types, (2) natural scene statistics (NSS) based approaches. NSS based approaches depend on the fact that natural scenes belong to a small set in the space of all possible signals and most distortions that are prevalent in image/video processing systems destroy the specific features of natural scenes. Recent works about this type of algorithms focused on developing advanced statistical models to describe the properties of natural images, and then the blind measurement of NR-IQA is achieved by measuring the variation in terms of NSS. For example, Lu et al. [10] improved the NSS model by contourlets. Moorthy et al. [11] extracted features of NSS in the wavelet domain. Saad et al. [12] applied a NSS model of discrete cosine transform coefficients. Mittal et al. [13] proposed a NSS model in the spatial domain, (3) training-based approaches. These approaches usually rely on a number of features extracted from images and a regression model is learned based on these features and labels to predict image quality. Recent works about this type of algorithms focused on using advanced machine learning methods to extract effective features to represent natural images and distorted images. For example, Li et al. [14] developed a NR-IQA algorithm using a general regression neural network. Ye et al. [15] presented a NR-IQA framework based on unsupervised feature learning framework.

In this paper, a new NR-IQA algorithm is proposed, which is called SHeArlet based No-reference Image quality Assessment (SHANIA). It is a combination of NSS-based and training-based approaches, and can estimate a wide range of image distortions. NSS model in shearlet domain is developed to identify natural images and distorted images. Training and learning methods are adopted through the entire framework to achieve this new NR-IQA method. The main idea of SHANIA is based on the finding that the statistical property of most natural images in shearlet domain is relatively constant. Nevertheless, distorted images usually contain more or less spread discontinuities in all directions. That is, real-world image distortions disturb the natural image statistical property and discriminate it from natural images to distorted images. Shearlets are apt at detecting these discontinuities. Thus, these variations in statistical property can be easily detected by shearlets and applied to describe image quality distortion. In particular, for natural images, the Mean of Subband Coefficient Amplitudes (MSCA) in different scales has special relationship in shearlet domain, that is, because of the dyadic transformation, the MSCA in coarse scale subbands are about the nth power of two times of fine scale subbands. Thus, when the MSCA are plotted versus the number of subbands in logarithmic coordinates ($\log_2$ of), it will have an approximate linear distribution in the same direction through different scales.

However, if a natural image is distorted by some common distortions, the linear relationship in coarse scales will retain, but it will be disturbed in fine scales. Motivated by this NSS model, the MSCA of coarse scales can be viewed as the 'natural parts' of a distorted image and the MSCA of fine scales is referred to as 'distorted parts'. We can make full use of the natural parts to predict the natural behavior of the MSCA in fine scales and use the 'predicted parts' as 'reference'. Finally, the difference between distorted parts and predicted parts can be measured. For natural images, the difference is small. For distorted images, the difference is large. Theoretically, this model can act as an indicator to blindly indicate image quality. In practice, our prediction strategy is achieved by modifying the general sparse autoencoder to create an accurate and robust predictor. Through the predictor, the predicted parts can be obtained from natural parts. Besides, by translating the NR-IQA problem into classification problem, the difference between predicted parts and distorted parts can be identified by softmax classifier. SHANIA does not incorporate any prior knowledge about distortions, making it suitable to many distortions.

The remainder of the paper is organized as follows. Section 2 introduces the detailed implementation and related techniques about SHANIA. In Section 3, experimental results and a thorough analysis of our results are given. Finally, conclusion and future works are presented in Section 4.

2. Methodology

The framework of the proposed approach is illustrated in Fig. 1. The major components in this framework include (1) MSCA extraction in shearlet domain, (2) predictor construction and training, (3) difference identification using softmax classifier, and (4) quality score calculation. More details will be described in the following sections.

2.1. Shearlet transform

The new NR-IQA method proposed in this paper is based on the shearlet transform [16–22]. This multiscale transform is a multidimensional edition of the traditional wavelet transform [23–25], and is capable for addressing anisotropic and directional information at different scales. When the dimension $n=2$, the affine systems with

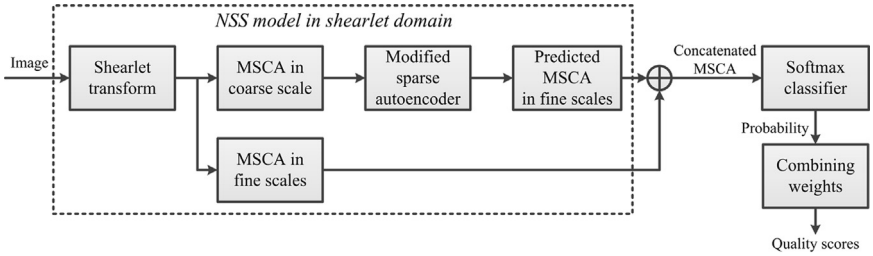

Fig. 1. Overview of the SHANIA framework.

composite dilations are the collections of the form

$$SH_\phi f(a,s,t) = <f,\ \phi_{a,s,t}>, a>0,\ \ s\in R, t\in R^2 \tag{1}$$

where the analyzing factor $\phi_{a,s,t}$ is called the shearlet coefficient, which is defined as

$$\phi_{a,s,t}(x) = |\det M_{a,s}|^{-(1/2)}\phi(M_{a,s}^{-1}x - t) \tag{2}$$

where $M_{a,s} = B_s A_a = \begin{pmatrix} a & \sqrt{a}s \\ 0 & \sqrt{a} \end{pmatrix}$, and $A_a = \begin{pmatrix} a & 0 \\ 0 & \sqrt{a} \end{pmatrix}$, $B_s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$. A_a is the anisotropic dilation matrix and B_S is the shear matrix. The analyzing functions associated to the shearlet transform are anisotropic and are defined at different scales, locations and orientations. Thus, shearlets have the ability to detect directional information and account for the geometry of multidimensional functions, which overcome the limitation of the wavelet transform.

Shearlets have a lot of very good mathematical properties [17]. For example, shearlet is well localized (which means they are compactly supported in the frequency domain and have fast decay in the spatial domain), highly directional sensitivity and optimally sparse. Let f be C^2 away from piecewise C^2 curves, and let $\hat{f}_N$ be the approximation to f using the N largest coefficients in the different transforms. The theoretical optimal approximation error decays as

$$\varepsilon_{Ideal} = \|f - \hat{f}_N\|_2^2 \le N^{-2} \tag{3}$$

For Fourier transform, the approximation error decays as

$$\varepsilon_{Fourier} = \|f - \hat{f}_N\|_2^2 \le N^{-(1/2)} \tag{4}$$

For wavelet transform, the approximation error is better than Fourier approximation, which is

$$\varepsilon_{Wavelet} = \|f - \hat{f}_N\|_2^2 \le CN^{-1} \tag{5}$$

For shearlet transform, the approximation is better than both Fourier and wavelet approximations, which can achieve

$$\varepsilon_{Shearlet} = \|f - \hat{f}_N\|_2^2 \le CN^{-2}(logN)^3 \tag{6}$$

Nowadays, only the curvelets [26,27] and contourlets [28] are claimed to satisfy similar sparsity property.

In summary, shearlets form a tight frame of well-localized waveforms, at various scales and directions, and are optimally sparse in representing images with edges. With these good properties, shearlets can provide more additional information about distorted images than the traditional wavelets and is suitable for NR-IQA work.

2.2. Shearlet coefficients statistical property of natural images and distorted images

Usually, natural images refer to images from the natural environment. These images form an extremely tiny subset of all possible scenes. Natural images are not random patterns, but show a number of consistent statistical properties. One of the properties that has received considerable attention by many authors is that natural image spectra follow a power law, which is defined as

$$S(f) \approx \frac{A_s(\theta)}{f^{2-\eta(\theta)}} \tag{7}$$

where $S(f)$ denotes the image spectra, $A_s(\theta)$ is called the amplitude scaling factor, $2-\eta(\theta)$ is the frequency exponent and η clusters around zero for natural images [29–33]. Thus, in the shearlet domain, this property can be expressed as

$$\mathbf{E}[|SH_\phi f(a,s,t)|] \approx a \tag{8}$$

where $SH_\phi f(a,s,t)$ is the shearlet transform of natural image and a is the scale parameter. Eq. (8) is the Mean of Subband Coefficient Amplitudes (MSCA) in different scales and different directions. For natural images, if Eq. (8) is calculated in logarithmic coordinates, MSCA in the same directions will exhibit a linear distribution through coarse scale to fine scale.

Similarly, in discrete shearlet domain, the MSCA of coarse scale are about the nth power of two times of fine scale. Thus, when MSCA are plotted versus the number of subbands in logarithmic coordinates ($\log_2$ of), it will also have an approximate linear distribution in the same directions. In addition, although the intercept of these lines differs from image to image, the slope of these lines shows slight change for different natural images. However, if the natural image is distorted by common distortions, the slope of these lines will deviate enormously in fine scales, but it is less affected in coarse scales. In a word, if a natural image is distorted by common distortions, MSCA in coarse scales are less affected than fine scales. Thus, MSCA in coarse scales are the most 'natural parts' of a distorted image, and MSCA in fine scales are the 'distorted parts'.

This property is well illustrated in Fig. 2, which is generated by all the 29 original images and their associated distorted versions in laboratory for image and video engineering (LIVE) database [34] (detailed information about LIVE will be given in Section 3). Distortions in LIVE include JPEG2K, JPEG, Gaussian blur (GB), fast fading (FF) and Gaussian white noise (GWN). In order to provide statically reasonable results, every original image and distorted image are randomly

sampled several times by the size of 256 × 256. Totally 12,000 sampled blocks are obtained and 2000 for each type. Shearlet transform with 4 scales and 6 directions for each scale is applied to each of the sampled blocks, and MSCA is calculated. The horizontal axis in Fig. 2 indicates the number of subbands and each scale is divided by the gray dashed line. The vertical axis represents the mean of MSCA of 2000 sampled blocks for each type in logarithmic coordinates.

It can be seen from Fig. 2 that common distortions disturb image statistics and make statistical property vary from that of natural images in shearlet domain. MSCA in fine scales (subbands from 7 to 24) are affected by distortions and become unnatural. However, MSCA in coarse scales (subbands from 1 to 6) are still very natural. Because of blurring, ringing and blocking artifact existed in JPEG2K, JPEG, GB and FF; fine scale coefficients decrease, and average energy of distorted images in fine scales becomes smaller than the original image, which is reflected by the decrease of MSCA in fine scales. On the contrary, for GWN, because much high frequency components are added, the average energy of distorted image becomes larger than the original image, which is reflected as the increase of MSCA in fine scales. Besides, MSCA in fine scales increase or decrease monotonously with the reduction in image quality.

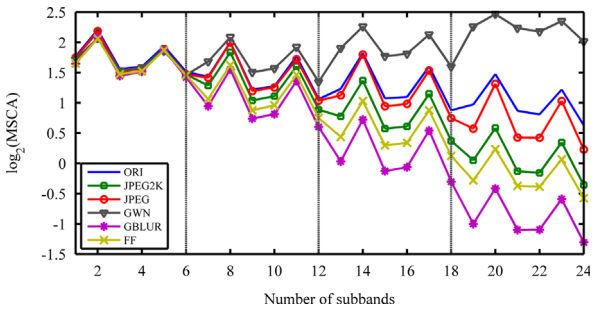

Fig. 2. Mean MSCA versus subband enumeration index for natural images and different distorted images in LIVE IQA database. ORI: original natural image, JPEG2K: JPEG2K compression, JPEG: JPEG compression, GWN: Gaussian white noise, GBLUR: Gaussian blur, FF: Rayleigh fast-fading channel simulation.

2.3. Modified sparse autoencoder

The findings in Section 2.2 reveal that common distortions more or less affect the distribution of singularities. The singularities are mainly reflected in fine scale coefficients and coarse scale coefficients contain smooth components. Therefore, image quality can be estimated by determining the changes of singularities in fine scale coefficients. However, NR-IQA algorithms cannot reference any information about the original image, which brings about a great difficulty to measure the variation in fine scale coefficients. In order to find a reliable metrics, the least affected components in the distorted image should be considered as a reference. Undoubtedly, coarse scale coefficients are a good choice. According to this idea, we propose to modify the cost function of general sparse autoencoder to predict the MSCA in fine scales only through the coarse scale. The general sparse autoencoder is a symmetric system, which tries to learn an approximation to the identity function, such that the approximation output $\hat{x}$ is as similar as the input x. In this work, we modify the sparse autoencoder into an asymmetric system. The inputs are the MSCA in the coarse scale and the outputs are the predicted MSCA in the fine scales, which is shown in Fig. 3.

In order to get the weights $\mathbf{W}_1$, $\mathbf{W}_2$ and bias $\mathbf{b}_1$, $\mathbf{b}_2$ in this prediction system, suppose that we have a training set $\{(x^{(1)},y^{(1)}),\ \ldots\ ,(x^{(m)},y^{(m)})\}$from natural images, and there are m training examples in it. $x^{(i)}$ are the MSCA in coarse scale, $\hat{x}^{(i)}$ are the predicted MSCA in fine scales and $y^{(i)}$ are the actual MSCA in fine scales.

First, we define the forward propagation of this prediction system

$$\begin{cases} z^{(2)} = \mathrm{W}^{(1)}x + b^{(1)} \\ a^{(2)} = f(z^{(2)}) \\ z^{(3)} = \mathrm{W}^{(2)}x + b^{(2)} \\ \hat{x} = a^{(3)} = f(z^{(3)}) \end{cases} \tag{9}$$

where f is the sigmoid function, which is represented as the red circle in Fig. 3. The output range of sigmoid

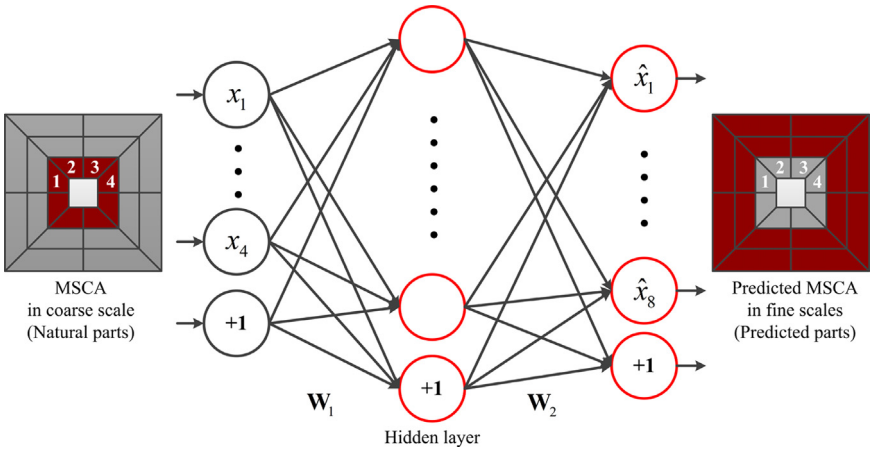

Fig. 3. Modified sparse autoencoder framework. In this illustration, a grayscale image is decomposed by shearlet transform, which has 3 scales and every scale has 4 directions. The inputs are MSCA only in the coarse scale (not the approximation component). The outputs are MSCA in the rest of fine scales. The red circle denotes sigmoid function.

function is from 0 to 1. The inputs and outputs need to be normalized before sending it into the prediction system. The normalization is achieved by

$$\begin{cases} x^{(i)} = \frac{x^{(i)}}{\max\{x^{(i)}, y^{(i)}\}} \\ y^{(i)} = \frac{y^{(i)}}{\max\{x^{(i)}, y^{(i)}\}} \end{cases} \tag{10}$$

where $\max\{x^{(i)}, y^{(i)}\}$ denote the maximum MSCA in all subbands.

Then, we define the overall cost function to be

$$J(W,b) = \left[\frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{2}\|\hat{x}^{(i)} - y^{(i)}\|^2\right)\right] + \frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}(W_{ji}^{(l)})^2 \tag{11}$$

where n_l is the number of layers, s_l is the number of neurons in layer l, λ is called the weight decay parameter, which controls the relative importance of the two terms.

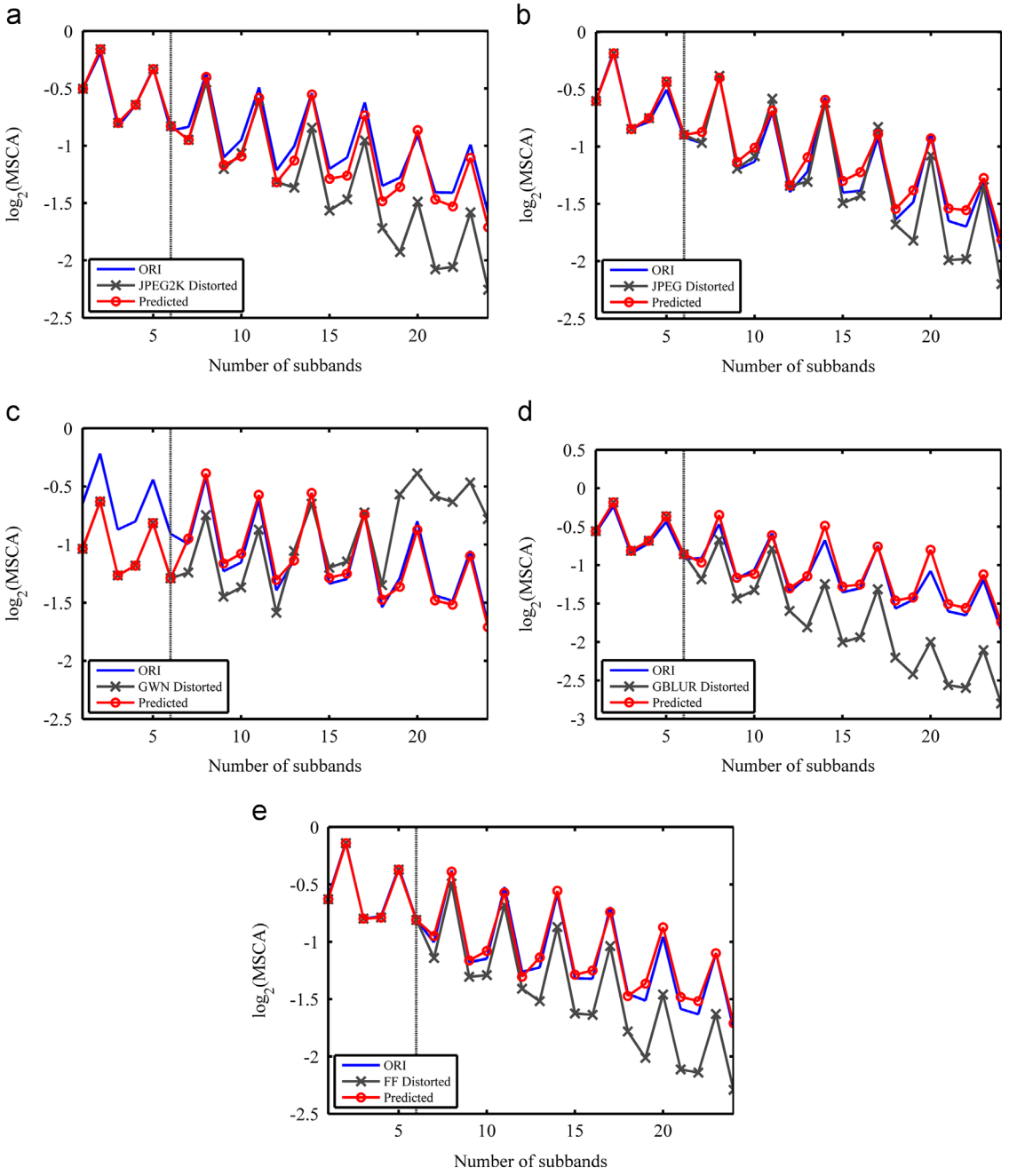

Fig. 4. Prediction results for five distortions in LIVE database using modified sparse autoencoder. The inputs and outputs are divided by the gray dashed line: (a) JPEG2000 compression, (b) JPEG compression, (c) Gaussian white noise, (d) Gaussian blur, and (e) Rayleigh fast-fading channel simulation.

If we further incorporate sparsity parameter ρ, the overall cost function is modified as

$$J_{sparse}(W,b)=J(W,b)+\beta\sum_{j=1}^{s_2}\mathrm{KL}(\rho||\hat{\rho}_j) \tag{12}$$

where ρ controls the weight of the sparsity penalty term. KL is the Kullback–Leibler divergence defined as

$$\mathrm{KL}(\rho||\hat{\rho}_j)=\rho\log\frac{\rho}{\hat{\rho}_j}+(1-\rho)log\frac{1-\rho}{1-\hat{\rho}_j} \tag{13}$$

and

$$\hat{\rho}_j=\frac{1}{m}\sum_{i=1}^{m}[a_j^{(2)}(x^{(i)})] \tag{14}$$

Finally, weights $\mathbf{W}_1$, $\mathbf{W}_2$ and bias $\mathbf{b}_1$, $\mathbf{b}_2$ are obtained by backpropagation (BP) algorithm. Detailed implementation about BP algorithm is available in [35].

Fig. 4(a)–(e) shows the prediction results for five distortions in LIVE database. These figures are generated by randomly selecting 15 original images to train the parameters (weights $\mathbf{W}_1$, $\mathbf{W}_2$ and bias $\mathbf{b}_1$, $\mathbf{b}_2$) of the predictor. The rest 14 original images and their distorted versions are randomly sampled several times by the size of 256×256. There are no overlap between the training images and testing images. 1000 sampled blocks are obtained for each distortion type. The inputs and outputs are divided by the gray dashed line. The vertical axis of Distorted and Predicted represent the mean of actual MSCA and predicted MSCA of 1000 sampled blocks for each type in logarithmic coordinates. The vertical axis of ORI represents the mean of MSCA of sampled blocks of original images in logarithmic coordinates. It can be seen that the predicted MSCA are very close to the MSCA of original images for all these five distortions and is different from MSCA of distorted images. This result shows that the modified sparse autoencoder is competent for the prediction work.

2.4. Classification and quality evaluation

In Section 2.3 we apply the modified sparse autoencoder as a predictor to get the predicted parts from the natural parts. Besides, the predicted parts are very similar to the distorted parts for natural images, and are different from the distorted parts for distorted images. In this section, we translate the NR-IQA into classification problem and use Softmax classifier to identify the image quality through predicted parts and distorted parts. In order to create labeled data to train the Softmax classifier, we first classify training images into 7 different classes based on their Mean Opinion Score (MOS). The class labels are created based on Table 1.

Now, we have a training set $\{(x^{(1)},y^{(1)}),\ \ldots\ ,(x^{(m)},y^{(m)})\}$ of m labeled data, where the input features are $x^{(i)}$, which is a concatenated MSCA vector defined as {Distorted parts, Predicted parts}. $y^{(i)}\in\{1,2,\ldots,7\}$ is the label of a training image.

Given an image feature $x^{(i)}$, we want to use Softmax classifier to predict the probability that $p(y^{(i)}=k|x^{(i)})$ for each value of $k=1,\ldots,7$. Thus the output of a Softmax classifier in this problem is a 7 dimensional probability vector, which is defined as

$$P_\theta(x^{(i)})=\begin{bmatrix}p(y^{(i)}=1|x^{(i)};\theta)\\p(y^{(i)}=2|x^{(i)};\theta)\\\vdots\\p(y^{(i)}=7|x^{(i)};\theta)\end{bmatrix}=\frac{1}{\sum_{j=1}^{7}e^{\theta_j^Tx^{(i)}}}\begin{bmatrix}e^{\theta_1^Tx^{(i)}}\\e^{\theta_2^Tx^{(i)}}\\\vdots\\e^{\theta_7^Tx^{(i)}}\end{bmatrix} \tag{15}$$

where $\theta=\begin{bmatrix}-\theta_1^T-\\-\theta_2^T-\\\vdots\\-\theta_7^T-\end{bmatrix}$ are the parameters of Softmax classifier, which can be obtained by training dataset. Since our work is not a pure classification problem, a mapping is needed to convert the hypothesis $P_\theta(x^{(i)})$ into quality scores. This mapping can be done by multiplying a combining weights $\omega\in R^7$, which is a row vector. Thus, the final quality score of an image is calculated by

$$Q_i=\omega\cdot P_\theta^T(x^{(i)}) \tag{16}$$

where the combining weights ω are learned by calculating the least square solution of this over determined equation upon the training set. The overall implementation steps and related training parameters of SHANIA are summarized in Table 2.

3. Experiments and related analysis

In order to effectively calibrate, train, test and compare the proposed NR-IQA algorithm, the following three IQA databases were used:

(1) LIVE IQA database. This IQA database contains 29 high-resolution 24-bits/pixel RGB original images distorted

Table 1
The relationship between image MOS and its label.

MOS	<25	25–34	35–44	45–54	55–64	65–74	>75
Class	1	2	3	4	5	6	7

Table 2
Summary of implementation steps and training parameters of SHANIA.

Steps	Description	Train parameter
1	MSCA extraction in shearlet domain	Weights $\mathbf{W}_1$, $\mathbf{W}_2$ and bias $\mathbf{b}_1$, $\mathbf{b}_2$
2	MSCA prediction using modified sparse autoencoder	
3	Concatenating distorted parts and predicted parts to form features	
4	Hypothesis probability calculation using Softmax	Softmax parameter θ
5	Quality score calculation	Combining weights ω

using five types of distortions at different distortion levels. These original images are distorted using the following distortion types: JPEG2000, JPEG, white Gaussian noise in the RGB components, Gaussian blur in the RGB components, and bit errors in JPEG2000 bit stream when transmitted over a simulated fast-fading Rayleigh channel. Besides, Mean Opinion Score (MOS) and the standard deviation between subjective scores were computed for each image. MOS for LIVE is in the range 0–100. Higher MOS indicates higher image quality.

(2) LIVE Multiply Distorted IQA database [36]. This IQA database extends one type of distorted images to two types of multiple distorted images. A subjective study was conducted on 15 natural images and their distorted versions. This study was conducted in two parts to obtain human judgments on images corrupted under two multiple distortion scenarios: (a) image storage where images are first blurred and then compressed by a JPEG encoder. (b) Camera image acquisition process where images are first blurred due to narrow depth of field or other defocus and then corrupted by white Gaussian noise to simulate sensor noise. Differential Mean Opinion Score (DMOS) associated with distorted images is given, which is in the range 0–100. Different from MOS, lower DMOS indicates higher image quality.

(3) TID2008 database [37,38]. This IQA database contains 25 reference images and 1700 distorted images. These 1700 distorted images are obtained using 17 types of distortions for each reference image and every distortion has 4 levels. Mean Opinion Scores (MOS) for this database were computed for each image, which is in the range 0–9. Higher value indicates higher visual quality of the image.

3.1. Classification accuracy

In this section, we first used the LIVE IQA database to test the performance of Softmax classification accuracy. The experiments are conducted in the following way. Firstly, 15 natural images in LIVE Multiply Distorted IQA database are converted into grayscale images and randomly sampled several times by the size of 256×256. 2000 sampled blocks are obtained to determine the training parameters of the predictor. Secondly, the original LIVE database is split into a training subset and a testing subset randomly without overlapping. The training subset consists of 15 natural images and their corresponding distorted versions. The testing subset is other 14 images and their distorted versions. The training parameters of Softmax and combining weights are obtained using training subset and the classification accuracy is measured on the testing subset. So as to eliminate performance bias, this random train-test procedure is repeated 1000 times and the following performance report is the median of 1000 trials.

In this experiment, Fourier based shearlet transform is applied and the grayscale image is decomposed into 4 scales (exclude approximation component) and every scale has 6 directions. The number of total subbands is 24. For the predictor, the number of neurons in hidden layer is 64. Weight decay parameter λ is $3e-6$. Sparsity parameter ρ is 0.1 and weight of sparsity penalty term β is 5. For Softmax classifier, weight decay parameter λ is also $3e-6$. Combining weights ω is learned by calculating the least square solution of overdetermined equation upon the training subset. In the following experiments, we also use the same parameters for predictor and classifier.

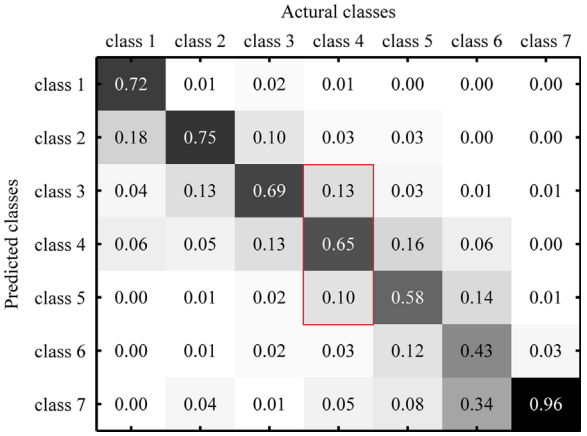

Fig. 5. Mean confusion matrix for Softmax classifier across 1000 trials for all the images in testing subset. The red box indicates the enlarged acceptance range.

Fig. 5 plots the mean confusion matrix for all the images in testing subset. The sum of each row in this matrix is 1 and the actual values represent the mean confusion percentage across the 1000 train-test trials. It can be seen from the confusion matrix that the general prediction accuracy is acceptable. However, as mentioned in last sensation, our work is not a pure classification problem. We just want to utilize the output probability of the classifier and it is also a reasonable output if the misclassification only occurs between the neighbors of the actual classes. For example, if the actual class is 4, we can consider it is also a right output if the prediction is 3 or 5, because the finally quality score is a weighted average of combining weights. Thus, we also report the classification accuracy after enlarging the acceptance range.

Fig. 6 is the box plots of classification accuracy for each of the 1000 trials and Table 3 gives median classification accuracy across 1000 trials before and after enlarging the acceptance range. We can see from the results that after enlarging the acceptance range, the classification accuracy is improved drastically, which means most misclassifications appear only between the neighbors of the actual classes.

3.2. Variation with scale and direction

Since our proposed method relies on the extracted features from shearlet domain, we report quality score prediction results for features extracted using shearlet transform with different scales and different directions. To evaluate the prediction results, two measurement criterions of linear correlation coefficient (LCC) and Spearman's rank correlation

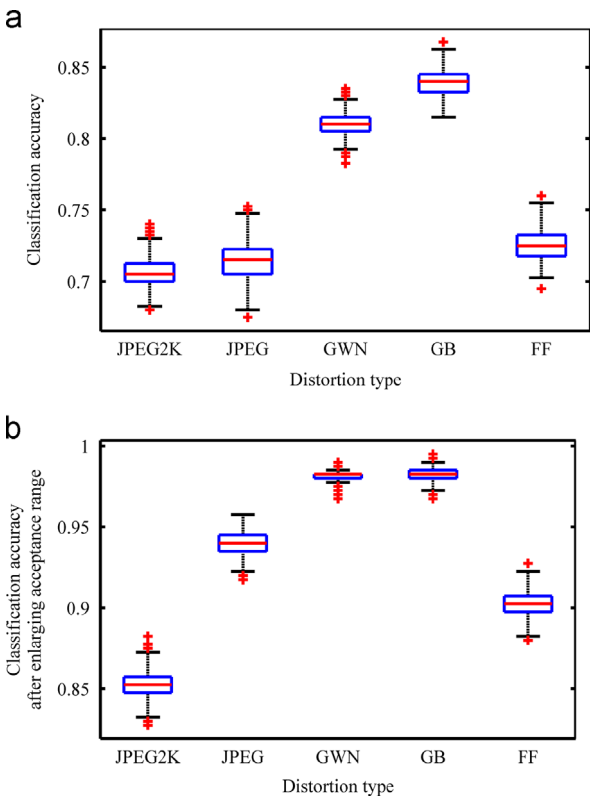

Fig. 6. Box plot of classification accuracy over 1000 trials on LIVE IQA database: (a) actual classification accuracy and (b) classification accuracy after enlarging acceptance range.

Table 3
Median classification accuracy across 1000 train-test trials.

	JPEG2K	JPEG	GWN	GBlur	FF
Classification accuracy (%)	70.5	71.5	81.8	84.8	72.5
Enlarged acceptance range (%)	85.3	94.8	98.3	98.7	91.5

coefficient (SROCC) for all the images in LIVE database are computed. LCC is used to reflect the prediction accuracy of the objective evaluation method and SROCC is used to reflect the monotonicity of the objective evaluation method. The experiment procedure is the same as described in Section 3.1. The comparison of median LCC and SROCC for different scales and directions is shown in Fig. 7. It can be seen from the results that both LCC and SROCC are improved with the increasing number of scale level and directions.

3.3. Influence of color information

Previous NR-IQA works usually convert color images into grayscale images and image quality is also measured using grayscale images. However, some distortion information is hidden in color components especially for multiple distorted images. Thus, in this section, we use color images to train and test our proposed method and compare the prediction results with those obtained through grayscale images.

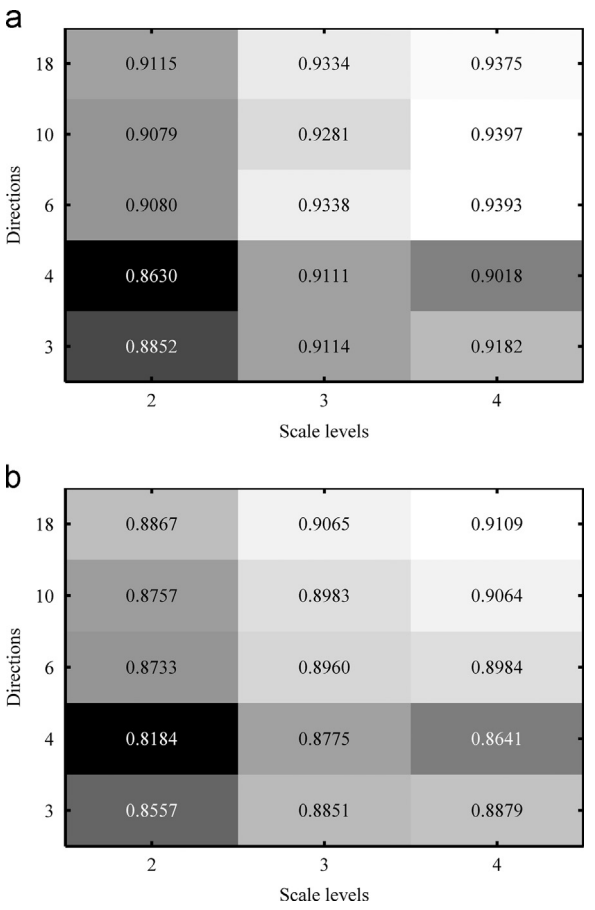

Fig. 7. Median LCC and SROCC correlations for 1000 iterations of randomly chosen train-test sets. LCC and SROCC are computed using all images in testing set: (a) median LCC for different scales and directions and (b) median SROCC for different scales and directions.

In this experiment, MSCA are extracted from RGB channels separately. The input of predictor is a concatenated natural part (NP), which is defined as {NPR, NPG, NPB}. The input feature for classifier is also a concatenated vector containing distorted part (DP) and predicted part (PP), which are denoted as {DPR, DPG, DPB, PPR, PPG, PPB}. Experiments are conducted using images in LIVE database and their grayscale versions. Similarly, the training subset consists of 15 natural images and their corresponding distorted versions. The testing subset is other 14 images and their distorted versions. This random train-test procedure is also repeated 1000 times and the median LCC and SROCC of 1000 trials are reported. Besides, image is decomposed into 4 scales and every scale has 6 directions. The comparison of prediction results for grayscale images and color images is shown in Table 4. We found that some improvements are achieved by considering color components.

3.4. Comparison with full-reference and no-reference IQA approaches

In this section, we used the LIVE IQA database and LIVE Multiply Distorted IQA database to test the performance of

Table 4
Median LCC and SROCC correlations for 1000 iterations using grayscale images and color images.

Measure criterion		Distortion type					
		JPEG2K	**JPEG**	**GWN**	**GBLUR**	**FF**	**ALL**
LCC	Grayscale	0.8976	0.9257	0.9698	0.9704	0.9134	0.9393
	Color	0.9018	0.9329	0.9717	0.9759	0.9318	0.9404
SROCC	Grayscale	0.8539	0.8674	0.9527	0.9604	0.8837	0.8984
	Color	0.8526	0.8830	0.9560	0.9630	0.9071	0.9021

Table 5
Median LCC and SROCC correlations for 1000 iterations of experiments on the LIVE IQA database. Italicized algorithms are NR-IQA algorithms.

Measure criterion		Distortion type					
		JPEG2K	JPEG	GWN	GBLUR	FF	ALL
LCC	PSRN	0.8692	0.8402	0.9412	0.8318	0.8569	0.8588
	SSIM	0.9048	0.8303	0.9321	0.8823	0.8538	0.8823
	MS-SSIM	0.8807	0.8062	0.8841	0.8298	0.8004	0.7939
	BIQI	0.8086	0.9011	0.9538	0.8293	0.7328	0.8205
	DIIVINE	0.9220	0.9210	0.9880	0.9230	0.8680	0.9170
	BLIINDS-II	0.9386	0.9426	0.9635	0.8994	0.8790	0.9164
	BRISQUE	0.9229	0.9734	0.9851	0.9506	0.9030	0.9424
	SHANIA	0.9135	0.9380	0.9731	0.9790	0.9413	0.9412
SROCC	PSRN	0.8483	0.8314	0.8964	0.8393	0.8354	0.8461
	SSIM	0.9363	0.9730	0.9638	0.9758	0.9832	0.9719
	MS-SSIM	0.9460	0.9742	0.9660	0.9770	0.9795	0.9795
	BIQI	0.7995	0.8914	0.9510	0.8463	0.7067	0.8195
	DIIVINE	0.9319	0.9483	0.9821	0.9210	0.8714	0.9116
	BLIINDS-II	0.9323	0.9331	0.9463	0.8912	0.8519	0.9124
	BRISQUE	0.9139	0.9647	0.9786	0.9511	0.8768	0.9395
	SHANIA	0.8611	0.8918	0.9582	0.9674	0.9169	0.9033

SHANIA. For comparison purpose, we also tabulated the performance of three FR-IQA and four state-of-art NR-IQA algorithms. Three FR-IQA methods include peak-signal-to-noise ratio (PSNR), structural similarity index (SSIM) [39] and multi-scale structural similarity index (MS-SSIM) [40]. Four general purpose NR-IQA methods include BIQI [11], DIIVINE [41], BLIINDS-II [12] and BRISQUE [13]. When testing on LIVE database, we divide the database into two randomly chosen subsets and there is no overlap between train and test sets. The training subsets contain 80% original images and their distorted versions and the testing subsets contain 20%. This train-test procedure is repeated 1000 times, and the median of the performance across these 1000 iterations is reported. For SHANIA, we train the predictor using 2000 randomly sampled image blocks (256×256) from 15 natural images in LIVE Multiply Distorted IQA database and the original images in this database are different from LIVE database. When testing on LIVE Multiply Distorted IQA database, the training subset consists of 8 natural images and their corresponding distorted versions. The testing subset is other 7 images and their distorted versions. We train the predictor using 2000 randomly sampled image blocks (256×256) from 29 natural images in LIVE IQA database. We adopted 4 scales and 6 directions shearlet transform and tested on color images.

Table 5 reports the median LCC and SROCC results tested on LIVE IQA database. It can be seen that SHANIA achieves comparable testing results and approaches the performance of the reliable FR-IQA methods and state-of-art general purpose NR-IQA methods. For some distortion types (GBLUR and FF), it also slightly outperforms these NR-IQA methods. Table 6 reports the median LCC and SROCC results tested on LIVE Multiply Distorted IQA database. Since the difficulty to measure the quality of multiple distorted images is larger than single distorted images, the performance for both FR-IQA methods and

Table 6
Median LCC and SROCC correlations for 1000 iterations of experiments on the LIVE Multiply Distorted IQA database. Italicized algorithm is NR-IQA algorithms.

Measure criterion		Distortion type	
		Part 1 (GBLUR+JPEG)	**Part 2 (GBLUR+GWN)**
LCC	PSNR	0.7679	0.8083
	SSIM	0.8376	0.8566
	MS-SSIM	0.8361	0.8698
	SHANIA	0.7629	0.7073
SROCC	PSNR	0.7183	0.7551
	SSIM	0.8745	0.9028
	MS-SSIM	0.8743	0.8887
	SHANIA	0.8014	0.7528

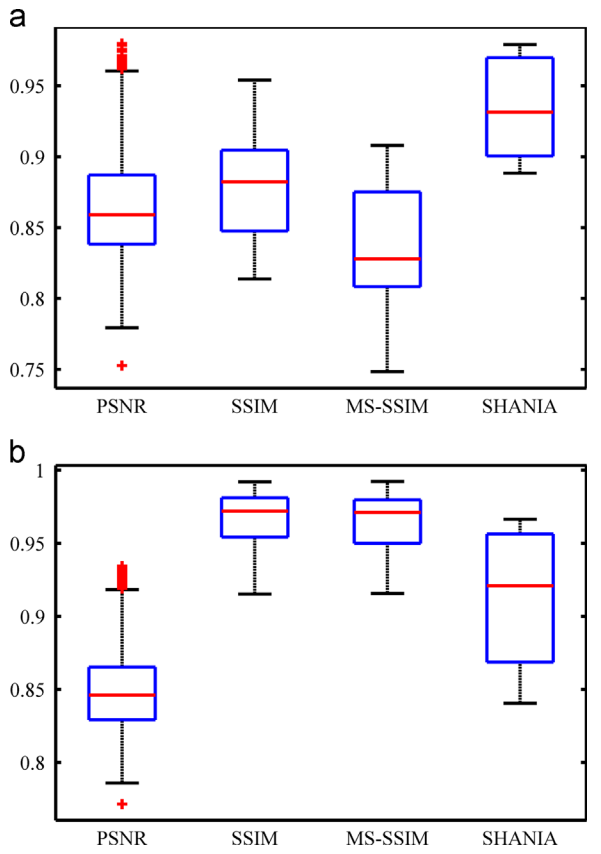

Fig. 8. Box plot of LCC and SROCC distributions of algorithms across 1000 trials of experiments on the LIVE IQA database. Box plot of (a) LCC distribution and (b) SROCC distribution.

SHANIA decreases. However, SHANIA can still achieve acceptable testing results.

Fig. 8(a) and (b) shows the box plot of LCC and SROCC distributions over 1000 experimental trials respectively. It is clear that the LCC distribution of SHANIA outperforms these three FR-IQA methods. The SROCC distribution of SHANIA does not outperform SSIM and MS-SSIM, but it still significantly outperforms PSNR. Fig. 9(a)–(f) shows the scatter plots of predicted MOS versus subjective MOS for each of the distortion sets as well as for the entire LIVE IQA database. We can see the distributions cluster along the diagonal line.

3.5. Database independence

In this section, we test whether the performance of SHNIA is influenced by the IQA database on which it is tested. In order to demonstrate SHANIA is not limited to the testing database, we trained SHANIA on the entire LIVE IQA database and tested a portion of the entire TID2008 image database. The testing subset includes JPEG2000, JPEG, additive Gaussian white noise in luminance component (GWNL), additive Gaussian noise in color components (GWNC) and Gaussian blur. Since TID2008 contain 24 natural images and one artificially synthesized image, the distorted versions of the synthesized image are not included in the testing subset. Table 7 reports the testing results of SHANIA and three FR-IQA methods and Fig. 10 shows the scatter plot of predicted MOS versus subjective MOS on the testing subset. Although the performance dropped because distorted images are generated in different ways between LIVE and TID2008, SHANIA still performs well and is consistent with human evaluations.

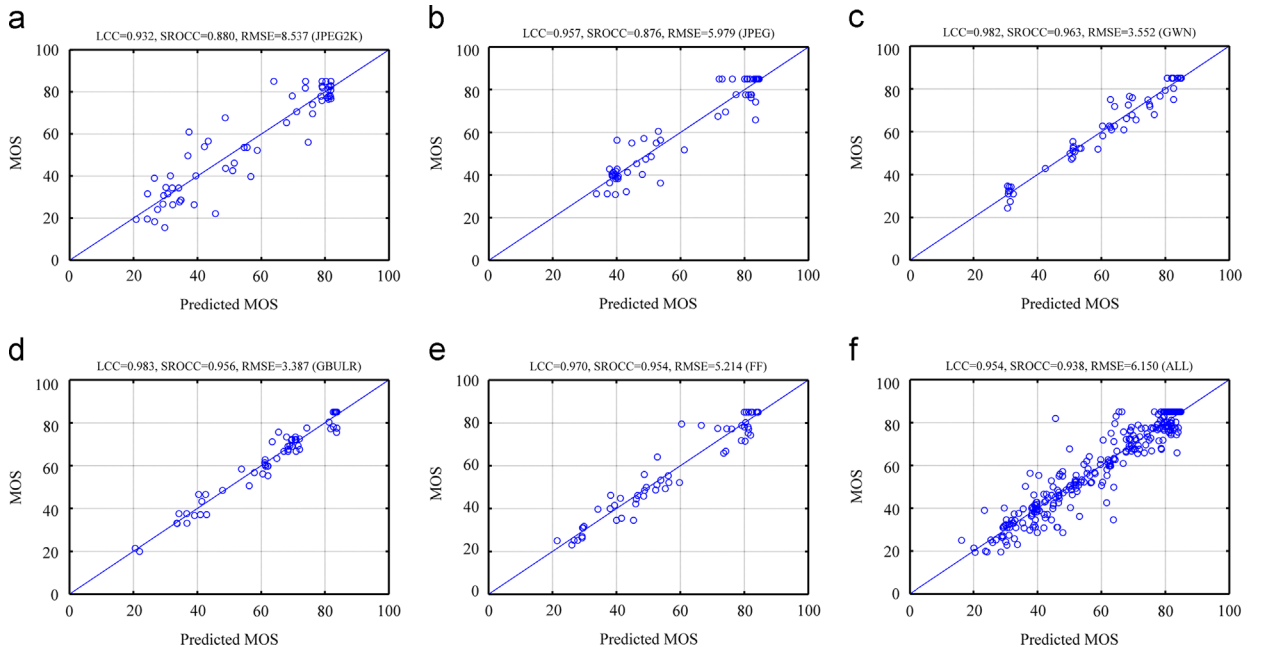

Fig. 9. Scatter plot of predicted MOS versus subjective MOS on LIVE IQA database. Root Mean Square Error (RMSE) is also calculated: (a) JPEG2000 database subset, (b) JPEG database subset, (c) Gaussian white noise database subset, (d) Gaussian blur database subset, (e) fast-fading database subset, and (f) entire LIVE IQA database.

Table 7
Median LCC and SROCC correlations for 1000 iterations of experiments on the TID2008 database. Italicized algorithm is NR-IQA algorithms.

Measure criterion		Distortion type					
		JPEG2K	JPEG	GWNL	GWNC	GB	ALL
LCC	PSRN	0.8814	0.8681	0.9416	0.9218	0.9271	0.8434
	SSIM	0.9472	0.9469	0.7576	0.7783	0.8906	0.9044
	MS-SSIM	0.9361	0.9353	0.7546	0.7817	0.8721	0.9040
	SHANIA	0.8588	0.9201	0.7083	0.7062	0.9038	0.8952
SROCC	PSRN	0.8248	0.8753	0.9177	0.8962	0.9335	0.9075
	SSIM	0.9604	0.9354	0.8168	0.7997	0.9598	0.9113
	MS-SSIM	0.9361	0.9371	0.8184	0.8042	0.8721	0.9087
	SHANIA	0.8383	0.8708	0.6841	0.6681	0.9049	0.8353

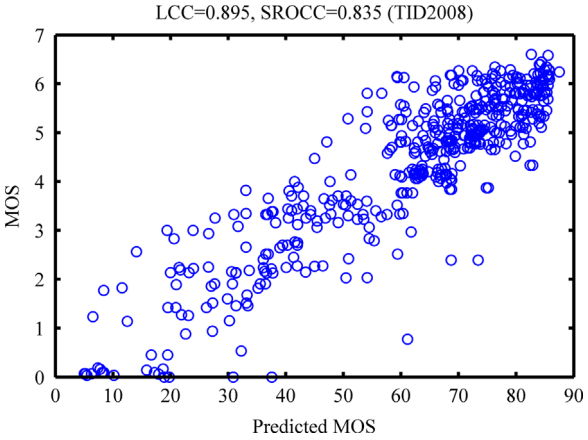

Fig. 10. Scatter plot of predicted MOS versus subjective MOS on TID2008 database.

4. Conclusion

In this paper, we proposed a general purpose NR-IQA algorithm SHANIA, which is based on the statistical characterization in the shearlet domain. We have demonstrated the natural scene statistics model in shearlet domain and proposed to use the most natural parts of an image to predict the natural tendency of other vulnerable parts, and compared the predicted parts with the actual distorted parts. Usually, the predicted parts are very close to the actual distorted parts for natural images, but they are very different for distorted images. Since the general idea about this algorithm is to quantify the naturalness of an image, no prior information about distortions is incorporated in this algorithm. SHANIA is tested on LIVE database, LIVE Multiply distorted database and TID2008 database. It has shown good performance on these databases and is comparative to FR-IQA methods and state-of-art general purpose NR-IQA algorithms.

Future work will involve improving SHANIA by considering the relationship between different subbands, such as using multivariate Laplace distribution [42,43] to model the relationship between them. Besides, more advanced machine learning methods, such as deep learning [44], can also be considered to improve NSS model. Furthermore, this method can also be extended to video quality assessment using 2D or 3D shearlet transform [45].

Acknowledgment

The work described in this paper was substantially supported by a grant from the City University of Hong Kong, Kowloon, Hong Kong with Project number 7004058.

References

[1] ZhouWang, Alan C. Bovik, Hamid R. Sheikh, Eero P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[2] Bernd Girod, What's wrong with mean-squared error? In: Andrew B. Watson (Ed.), Digital images and human vision table of contents, MIT Press, Cambridge, MA, USA, 1993, 207–220 (ISBN: 0-262-23171-9).
[3] Ahmet M. Eskicioglu, Paul S. Fisher, Image quality measures and their performance, IEEE Trans. Commun. 43 (12) (1995) 2959–2965.
[4] Zhou Wang, Alan C. Bovik, A universal image quality index, IEEE Signal Process. Lett. 9 (3) (2002) 81–84.
[5] Thrasyvoulos N. Pappas, Robert J. Safranek, J. Chen, Perceptual criteria for image quality evaluationHandbook of Image and Video Processing, 669–684.
[6] Kalpana Seshadrinathan, Thrasyvoulos N. Pappas, Robert J. Safranek, Junqing Chen, Wang Zhou, Hamid R. Sheikh, Alan C. Bovik, Image quality assessment, The Essential Guide to Image Processing, , 2009.
[7] Zhou Wang, Alan C. Bovik, Ligang Lu, Why is image quality assessment so difficult? In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 4, 2002, pp. IV–3313.
[8] Zhou Wang, Hamid R. Sheikh, Alan C. Bovik, No-reference perceptual quality assessment of JPEG compressed images, in: Proceedings of the International Conference on Image Processing, vol. 1, 2002, pp. I–477.
[9] Hamid R. Sheikh, Alan C. Bovik, Lawrence Cormack, No-reference quality assessment using natural scene statistics: JPEG2000, IEEE Trans. Image Process. 14 (11) (2005) 1918–1927.
[10] Lu WenKai ZengDacheng Tao, Yuan Yuan, Xinbo Gao, No-reference image quality assessment in contourlet domain, Neurocomputing 73 (4) (2010) 784–794.
[11] Moorthy, Anush Krishna, Alan Conrad Bovik, A two-step framework for constructing blind image quality indices, IEEE Signal Process. Lett. 17 (5) (2010) 513–516.
[12] Michele A. Saad, Alan C. Bovik, Christophe Charrier, Blind image quality assessment: A natural scene statistics approach in the DCT domain, IEEE Trans. Image Process. 21 (8) (2012) 3339–3352.
[13] Anish Mittal, Anush Krishna Moorthy, Alan Conrad Bovik, No-reference image quality assessment in the spatial domain, IEEE Trans. Image Process. 21 (12) (2012) 4695–4708.
[14] Chaofeng Li, Alan Conrad Bovik, Xiaojun Wu, Blind image quality assessment using a general regression neural network, IEEE Trans. Neural Netw. 22 (5) (2011) 793–799.
[15] Peng Ye, Jayant Kumar, Le Kang, David Doermann, Unsupervised feature learning framework for no-reference image quality assessment, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, pp. 1098–1105.

[16] Yi Sheng, Demetrio Labate, Glenn R. Easley, Hamid Krim, A shearlet approach to edge analysis and detection, IEEE Trans. Image Process. 18 (5) (2009) 929–941.
[17] Glenn Easley, Demetrio Labate, Wang-Q. Lim, Sparse directional image representations using the discrete shearlet transform, Appl. Comput. Harmonic Anal. 25 (1) (2008) 25–46.
[18] Gitta Kutyniok, Wang-Q. Lim. Image separation using wavelets and shearlets. arXiv preprint arXiv: 1101.0553, 2011.
[19] Gitta Kutyniok, Wang-Q. Lim, Xiaosheng Zhuang, Digital shearlet transforms, In: Gitta Kutyniok, Demetrio Labate (Eds.), Shearlets, Birkhäuser, Boston, 2012, 239–282 (ISBN: 978-0-8176-8315-3).
[20] Gitta Kutyniok, Morteza Shahram, Xiaoshesng Zhuang, Shearlab: a rational design of a digital parabolic scaling algorithm, arXiv preprint arXiv:1106.1319, 2011.
[21] David L. Donoho, Gitta Kutyniok, Morteza Shahram, Xiaoshesng Zhuang, A rational design of a digital shearlet transform, RN (RN+1) 8, vol. 2, 2011, p. 1.
[22] Gitta Kutyniok, Morteza Shahram, David L. Donoho, Development of a digital shearlet transform based on pseudo-polar FFT, in: Proceedings of the SPIE Optical Engineering+Applications, International Society for Optics and Photonics, 2009, p. 74460B.
[23] Ingrid DaubechiesTen Lectures on Wavelets, vol. 61, Society for Industrial and Applied Mathematics, Philadelphia, 1992.
[24] Stéphane Mallat, A Wavelet Tour of Signal Processing, Access Online via Elsevier, 1999.
[25] Sidney C. Burrus, Ramesh A. Gopinath, Haitao Guo, Jan E. Odegard, Ivan W. Selesnick, Introduction to Wavelets and Wavelet Transforms: A Primer, vol. 23, Prentice Hall, Upper Saddle River, 1998.
[26] Jean-Luc Starck, Emmanuel J. Candès, David L. Donoho, The curvelet transform for image denoising, IEEE Trans. Image Process. 11 (6) (2002) 670–684.
[27] Emmanuel Candes, Laurent Demanet, David Donoho, Lexing Ying, Fast discrete curvelet transforms, Multiscale Model. Simul. 5 (3) (2006) 861–899.
[28] Minh N. Do, Martin Vetterli, The contourlet transform: an efficient directional multiresolution image representation, IEEE Trans. Image Process. 14 (12) (2005) 2091–2106.
[29] R.P. Millane, S. Alzaidi, W.H. Hsiao. Scaling and power spectra of natural images, in: Proceedings of the Image and Vision Computing, New Zealand, 2003, pp. 148–153.
[30] David 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.
[31] D.J. Tolhurst, Y. Tadmor, Tang Chao, Amplitude spectra of natural images, Ophthalmic Physiol. Opt. 12 (2) (1992) 229–232.
[32] Daniel L. Ruderman, Origins of scaling in natural images Electronic Imaging: Science & Technology, International Society for Optics and Photonics, Proc. SPIE 2657, Human Vision and Electronic Imaging, Vol. 120, April 22, 1996 10.1117/12.238707.
[33] David J. Field, Nuala Brady, Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes, Vision Res. 37 (23) (1997) 3367–3383.
[34] Hamid R. Sheikh, Zhou Wang, Alan C. Bovik, L.K. Cormack, Image and video quality assessment research at LIVE, 2003. ⟨http://live.ece.utexas.edu/research/quality/⟩.
[35] Andrew Ng, Jiquan Ngiam, Chuan Yu Foo, Yifan Mai, Caroline Suen, Unsupervised feature learning and deep learning. ⟨http://ufldl.stanford.edu/wiki/index.php/UFLDL_Tutorial⟩.
[36] Dinesh Jayaraman, Anish Mittal, Anush K. Moorthy, Alan C. Bovik, Objective quality assessment of multiply distorted images, in: Proceedings of the IEEE Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), 2012 Conference Record, 2012, pp. 1693–1697.
[37] NikolayPonomarenko, Vladimir Lukin, Alexander Zelensky, Karen Egiazarian, M. Carli, F. Battisti, TID2008 – a database for evaluation of full-reference visual quality assessment metrics, Adv. Mod. Radioelectron. 10 (4) (2009) 30–45.
[38] Nikolay Ponomarenko, Federica Battisti, Karen Egiazarian, Jaakko Astola, Vladimir Lukin, Metrics performance comparison for color image database, in: Proceedings of the Fourth International Workshop on Video Processing and Quality Metrics for Consumer Electronics, vol. 27. 2009.
[39] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, Eero P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[40] Zhou Wang, Eero P. Simoncelli, Alan C. Bovik, Multiscale structural similarity for image quality assessment, in: Proceedings of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers, Conference Record, vol. 2, 2003, pp. 1398–1402.
[41] Moorthy, Anush Krishna, Alan Conrad Bovik, Blind image quality assessment: from natural scene statistics to perceptual quality, IEEE Trans. Image Process. 20 (12) (2011) 3350–3364.
[42] Eltoft, Torbjørn, Taesu Kim, Te-Won Lee, On the multivariate Laplace distribution, IEEE Signal Process. Lett. 13 (5) (2006) 300–303.
[43] Samuel Kotz, Tomaz J. Kozubowski, Krzysztof Podgórski, Asymmetric Multivariate Laplace distribution, The Laplace Distribution and GeneralizationsBirkhäuser, Boston, 2001, 239–272.
[44] Matthew D. Zeiler, Rob Fergus, Visualizing and understanding convolutional neural networks. arXiv preprint arXiv:1311.2901, 2013.
[45] PooranNegi, Demetrio Labate, 3-D discrete shearlet transform and video processing, IEEE Trans. Image Process. 21 (6) (2012) 2944–2954.