Signal Processing: *Image Communication* 29 (2014) 856–863

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

No-reference image quality assessment based on spatial and spectral entropies

Lixiong Liu [a,*], Bao Liu [a], Hua Huang [a], Alan Conrad Bovik [b]

[a] *Beijing Laboratory of Intelligent Information Technology, School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China*

[b] *Laboratory for Image and Video Engineering, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712, USA*

ARTICLE INFO

Article history:
Received 4 October 2013
Received in revised form
13 June 2014
Accepted 15 June 2014
Available online 23 June 2014

Keywords:
Image quality assessment
No-reference
Spatial entropy
Spectral entropy
Support vector machine

ABSTRACT

We develop an efficient general-purpose no-reference (NR) image quality assessment (IQA) model that utilizes local spatial and spectral entropy features on distorted images. Using a 2-stage framework of distortion classification followed by quality assessment, we utilize a support vector machine (SVM) to train an image distortion and quality prediction engine. The resulting algorithm, dubbed Spatial–Spectral Entropy-based Quality (SSEQ) index, is capable of assessing the quality of a distorted image across multiple distortion categories. We explain the entropy features used and their relevance to perception and thoroughly evaluate the algorithm on the LIVE IQA database. We find that SSEQ matches well with human subjective opinions of image quality, and is statistically superior to the full-reference (FR) IQA algorithm SSIM and several top-performing NR IQA methods: BIQI, DIIVINE, and BLIINDS-II. SSEQ has a considerably low complexity. We also tested SSEQ on the TID2008 database to ascertain whether it has performance that is database independent.

1. Introduction

With the development of multimedia network technology and the widespread usage of intelligent mobile phones, digital images have become an increasingly important medium to obtain information and communicate with others. Because of the limits of access devices, storage media, processing technologies and transmission equipment, digital images are subject to distortions during acquisition, compression, transmission, processing and reproduction. Distortions seriously affect the ability of humans to extract and understand the information contained in images. Therefore, in order to ensure, control and enhance image quality, identifying and quantifying image distortions become particularly important. The development of objective image quality assessment (IQA) models has important practical significance [1,2].

Depending on whether there are available original reference images, objective IQA methods can be divided into three categories: full-reference (FR), reduced-reference (RR) and no-reference (NR). In most cases, an original reference image does not exist, leaving NR IQA as the only possible method that can be practically embedded into such application systems. At present, proposed NR methods can be divided into two categories: (1) Distortion-specific quality assessment methods quantify a specific distortion regardless of other factors and

* Corresponding author.
E-mail addresses: lxliu@bit.edu.cn (L. Liu), liubao@bit.edu.cn (B. Liu), huahuang@bit.edu.cn (H. Huang), bovik@ece.utexas.edu (A.C. Bovik).

http://dx.doi.org/10.1016/j.image.2014.06.006
0923-5965/

score a distorted image accordingly. Examples of such NR–IQA methods [3–6] compute different kinds of distortion effects to a specific distortion, such as JP2K in [3], noise in [4], blur in [5] and JPEG in [6]. Moreover, Sheikh et al. use Natural Scene Statistics (NSS) to assess JP2K distortion severity [7]. These methods are obviously limited by the fact that it is necessary to know the distortion types in advance, which makes their application field limited. (2) Thus, the general-purpose quality assessment method based on training and learning is an emerging research direction. Moorthy and Bovik proposed a two-step framework for NR IQA, which involves distortion classifying and distortion-specific quality assessment in sequence, making use of several NSS features to implement a simple blind IQA index named BIQI [8]. They later refined the idea, developing the DIIVINE index also based on the two-step framework. The DIIVINE index uses a series of NSS features in the wavelet domain to predict image quality, thereby achieving excellent performance [9]. Saad et al. proposed BLIINDS-II, which extracts NSS features in the block DCT domain using a fast single-stage framework [10]. Mittal et al. proposed the BRISQUE [11] index that achieves superior predictive performance. As its features are all extracted in the spatial domain, its time complexity is particularly low. And we also proposed such a training based CurveletQA method, which extracts a serious of NSS features in the curvelet domain and correlates well with human perception [12].

Information entropy is an effective measure of the amount of information in an image. Sporring studied the relationship between image entropy, the histogram and image moments and found that the entropy captures statistical information over scales [13]. Sheikh and Bovik utilized image entropy for FR IQA and achieved good results [14,15], while Hu et al. constructed an NR–IQA function based on the product of entropy and contrast [16]. Motivated by these findings, we develop an efficient general-purpose no-reference (NR) image quality assessment (IQA) model, which utilizes spatial and spectral entropy features from a distorted image. The proposed method achieves very good predictive performance with acceptable computation.

The rest of the paper is organized as follows. In Section 2 we describe the framework of the proposed method. In Section 3 we describe the features extracted from a distorted image. In Section 4 we introduced the two-stage framework for NR image quality assessment [8,9] which is utilized in our method. We evaluate the performance of the proposed method in Section 5 and conclude the paper in Section 6.

2. Overview of the method

Natural photographic images are highly structured [17], in the sense that their pixels exhibit strong dependencies in space and frequency, and these dependencies carry important information about the visual scene. Localized image entropy features can capture the degree of local image structure and the entropy degree can denote the dependence level between pixels. We deploy two types of features: spatial and spectral entropies. The framework of the proposed IQA method is summarized in Fig. 1 and can be roughly divided into four steps:

1) The first step of the IQA "pipeline" is to preprocess incoming distorted images using a downsampling processing with a factor of 2, enabling simple multiscale analysis. In order to avoid aliasing, the bicubic interpolation method is applied during downsampling. In our implementation, we decompose each possibly distorted image into 3 scales: low, middle and high, yielding 3 scale responses. We found that increasing or decreasing the number of scales did not contribute to performance any more.
2) The second step of the pipeline is to partition the responses on which block-based entropy is computed. We partition each image into $M \times M$ blocks, henceforth referred to as local image patches, then compute spatial and frequency entropies within each block. The performance of the model has been found to be insensitive to the local patch size over a large number of experiments, hence was fixed at $M=8$. The process of computing spatial and spectral entropies on the patches is covered in more detail later.
3) The third step is feature pooling. At this stage, the 2 feature sets obtained during the second step – spatial and spectral entropies – are sorted in ascending order, thus yielding ordered sets $S=(se_1, se_2, \ldots, se_m)$ and $F=(fe_1, fe_2, \ldots, fe_m)$. Here se_i and fe_i are local spatial and spectral entropies respectively, and m is the number of blocks within each scale. The method of pooling used is percentile pooling [18]: 60% of the central

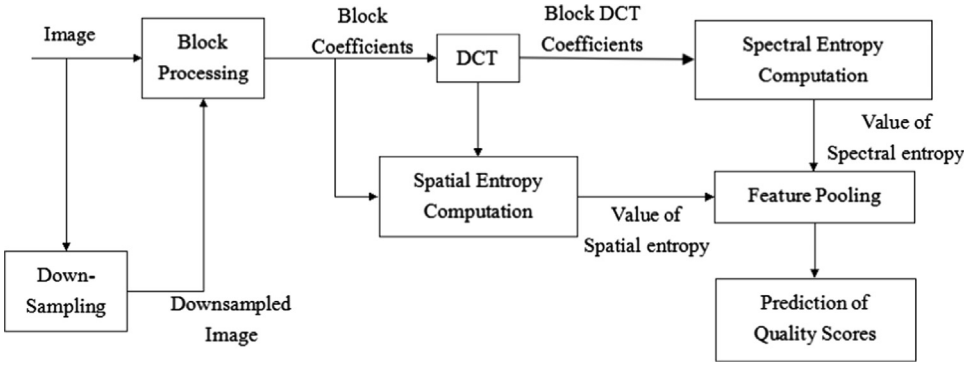

Fig. 1. High-level overview of the SSEQ framework.

elements are extracted from S yielding the set $Sc=(se_{\lfloor 0.2m\rfloor}, se_{\lfloor 0.2m\rfloor+1}, \ldots, se_{\lfloor 0.8m\rfloor})$, and 60% of the central elements are extracted from F yielding the set $Fc=(fe_{\lfloor 0.2m\rfloor}, fe_{\lfloor 0.2m\rfloor+1}, \ldots, fe_{\lfloor 0.8m\rfloor})$. The introduction of percentile pooling here is based on the hypothesis that distortions only occur in a subset of patches, leading to the variation of a part of entropy values. Our experiments show that percentile pooling improves our method's performance. The mean values of Sc and Fc and the skew values of S and F supply the final features from each scale:

$$f=(\text{mean}\,(Sc),\ \text{skew}\,(S),\ \text{mean}\,(Fc),\ \text{skew}\,(F)). \quad (1)$$

4) The final stage is prediction of the image quality score. The distorted image is decomposed into 3 scales from which 4 features are extracted. So the feature vector of an image has $3\times 4=12$ dimensions. Given the 12-dimensional feature vector, we utilize a two-stage framework for NR image quality assessment as in [8,9] to get the final image quality score.

3. Image entropy features

Image entropy indicates the amount of information contained within an image and when computed over multiscales reveals the statistical entropy of scale space [13]. The types and degrees of image distortions will generally affect the local entropy of images in diverse, yet predictable ways. Unlike local entropy, global image entropy captures global information of images but does not discriminate the spatial distribution of information. Thus images having the same global entropy may appear quite different.

Previous work has shown that there exists a close relationship between local image entropy and perceived image quality [15]. Likewise, we have found that entropy features are highly sensitive to the degrees and types of image distortions. Our method utilizes entropies computed from local image blocks, on both the block spatial scale responses and also on the block DCT coefficients. All of the entropies are computed locally. Spatial entropy is a function of the probability distribution of the local pixel values, while spectral entropy is a function of the probability distribution of the local DCT coefficient values. Compared with NSS based IQA methods, which utilizes the statistical characteristics at pixel level, our entropy based method analyzes the joint distribution of pixels within a local patch. Our method reflects the statistical characteristics of local regions rather than pixels, which will contribute to describe images' local structural information.

Our basic hypothesis is that the local entropy of undistorted images possesses certain statistical properties. The statistical properties are a result of the dependence between adjacent pixels. And the introduction of distortion will destroy this inherent dependence, yielding the change of local entropy. For example, additive white noise will cause larger local entropy values due to the introduction of much high-frequency information to undistorted images but blur distortion will reduce local entropy values owing to the loss of image details. Here we will describe the local entropy and discuss the relationship between the local entropy and image quality in detail.

3.1. Spatial entropy features (f_1-f_6)

The spatial entropy is

$$E_s=-\sum_x p(x)\log_2 p(x), \quad (2)$$

where x are the pixel values within a block, with empirical probability density $p(x)$, i.e. relative frequency here. To illustrate the behavior of the local spatial entropy values against different degrees and types of distortions, we conducted a series of validation experiments on an image. As shown in Fig. 2, different types of distortions (JP2K and JPEG compression, noise, blur and fastfading) exert systematically different influences on the spatial entropy values. The undistorted image (ori) has a spatial entropy histogram with mean about 4 and which is left-skewed. The "left-skewed" means that there are less data at the left of mean value than right, leading to a longer left tail than right. However, the introduction of distortion changes its mean and skew. For example, noise tends to sharply increase the mean, while blur and jp2k reduced the mean and skewed the histogram to the right. The jpeg histogram is quite different since many zeroes occur sharply decreasing the localized spatial entropy. Overall, spatial entropy features are strongly indicative of the type of distortion as we make clear going forward.

We utilize the mean (defined in Section 2) and skew as quality features that describe the histogram. We extracted these 2 features from each scale, yielding $2\times 3=6$ features.

3.2. Spectral entropy features (f_7-f_{12})

Since we believe that there exists a strong relationship between spectral entropy values and the degree and type of distortion [15], the block DCT coefficient matrix C is also computed on 8×8 blocks. Use of the DCT rather than the DFT reduces block edge energy in the transform

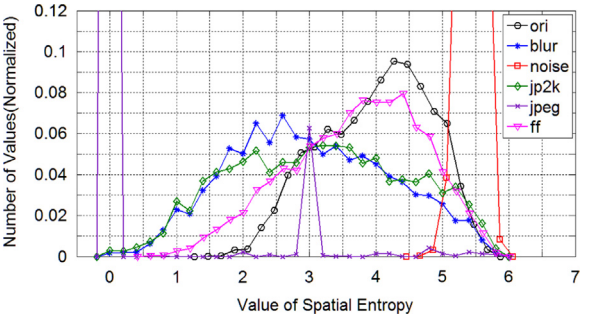

Fig. 2. Histograms of spatial entropy values for different types of distortion. The 6 curves correspond to an undistorted image and its distorted counterparts. Notice how each distortion affects the histograms of the spatial entropy values. "ori" (DMOS=0), "jp2k" (DMOS=65.3354), "jpeg" (DMOS=61.1951), "noise" (DMOS=67.5793), "blur" (DMOS=59.4296) and "ff" (DMOS=47.9970).

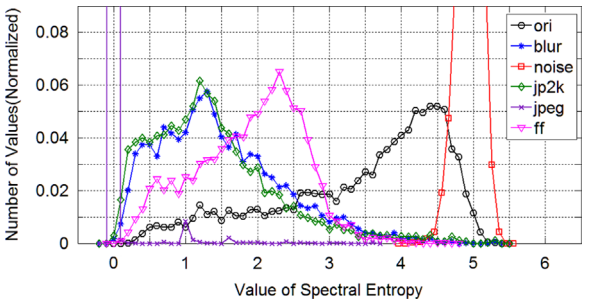

Fig. 3. Histograms of spectral entropy values for different types of distortion. The 6 curves correspond to an undistorted image and its distorted counterparts. The images here are the same with those in Fig. 2. Notice how each distortion affects the histograms of the spectral entropy values. "ori" (DMOS=0), "jp2k" (DMOS=65.3354), "jpeg" (DMOS=61.1951), "noise" (DMOS=67.5793), "blur" (DMOS=59.4296) and "ff" (DMOS=47.9970).

coefficients. We then normalize the DCT coefficients to produce a spectral probability map:

$$P(i,j)=\frac{C(i,j)^2}{\sum_i\sum_j C(i,j)^2}, \tag{3}$$

where $1\le i\le 8$, $1\le j\le 8$, and $i,j\neq 1$ (DC is excluded). Then define the local spectral entropy

$$E_f=-\sum_i\sum_j P(i,j)\log_2 P(i,j). \tag{4}$$

To explore the behavior of the spectral entropy features against different types of distortions, we conducted another visual validation experiment on the same images in Fig. 2. As shown in Fig. 3, different types of distortions (jp2k and jpeg compression, noise, blur and fastfading) exert systematically different influences on the spectral entropy values. Clearly, we may find that the undistorted image (ori) has a spectral entropy histogram that is typically left-skewed. However, the introduction of distortion will change its mean and skew. For example, noise tends to increase the mean sharply, while blur, jp2k and ff reduce the mean and skew the histogram to the right. The jpeg histogram is again heavily affected. The spectral entropy features are also strongly indicative of the type of distortions. Compared with the spatial entropy histogram, the spectral entropy histogram more clearly distinguishes the undistorted image from those affected by distortions. The spectral entropy defined here is an accurate descriptor of images' energy spectrum and emphasizes the main frequency and main orientations within a local patch. So it is able to distinguish noise and blur effect more clearly. Further, spectral entropy can capture texture variations more effectively, to which human perception is very sensitive.

As with the spatial entropy, we use the mean (defined in Section 2) and skew as descriptive quality features. We extract these 2 features from each scale, yielding $2\times 3=6$ features.

In both Figs. 2 and 3, we further find that the entropy distributions of "noise" and "jpeg" can be clearly distinguished from that of "ori". Among 5 kinds of distortion, the difference between "noise", "jpeg" and "ori" is most significant. The reason is that "noise" will introduce much high-frequency signal to the "ori", increasing both spatial and spectral entropy and "jpeg" will remove much high-frequency signal from the "ori", decreasing both spatial and spectral entropy. In addition, "jpeg" is based on DCT, which is employed in our method and can be attributed to discriminating this type distortion. Meanwhile, the entropy distributions of "blur" and "jp2k" also have different shapes from that of "ori" but "blur" and "jp2k" share almost the same entropy distribution. This observation is easy to understand that "jp2k" will introduce blur effect by removing much high-frequency information from "ori", while "ff" and "ori" have the most similar entropy distribution, particularly for spatial entropy distribution. Therefore, the similarity between "ff" and "ori" will result to relatively worse performance than others, which will be shown next.

Overall, our model extracts only 12 "quality-aware" features from the distorted image to be tested, as tabulated in Table 1. Due to extracting each kind of features from 3 scales (low, middle and high), each feature group includes 3 features.

4. Two-stage framework for blind image quality assessment

The NR IQA method proposed here uses the two-stage framework for NR image quality assessment [8,9] to map feature vectors to predicted quality scores. Thus, a probabilistic classifier is trained to compute the probability of occurrence of each distortion in an image, and then regression functions are trained on each distortion type against human scores. This yields two vectors: the distortion probability vector and the distortion-specific quality vector. A final predicted quality score is obtained by computing the dot product of these 2 vectors.

Machine learning has also been applied in the field of IQA for a long time, such as support vector regression (SVR) [19,20] and neural network [21,22]. As explained in detail later, we use a support vector machine (SVM) for classification and SVR for regression [23,24]. We used the LIBSVM package [25] to implement the SVM and SVR, both implemented using the radial basis function (RBF) kernel.

5. Experiments and results

We tested the performance of SSEQ on the LIVE IQA database [26], which contains 29 reference images distorted by the 5 distortion types: white noise (WN), JPEG and JP2K compression, Gaussian blur (Blur) and fast Rayleigh fading (FF), yielding 799 distorted images. Each distorted image is provided with a Difference Mean

Table 1
Features used for SSEQ.

Feature vector	Feature description
f_1-f_3	Means of spatial entropy values for 3 scales
f_4-f_6	Skews of spatial entropy values for 3 scales
f_7-f_9	Means of spectral entropy values for 3 scales
$f_{10}-f_{12}$	Skews of spectral entropy values for 3 scales

Opinion Score (DMOS) value, which is representative of the human subjective score of the image.

Since SSEQ is a training based model, the database needs to be partitioned into the training and test sets. The training set is used to train the classification and prediction models and the test set is used to evaluate the performance of IQA methods. To certify the validity of the evaluation results, the process of database partition was performed with the following rules: (1) The training and test sets were separated by content; (2) 80% of the reference images and corresponding distorted counterparts constituted the training set and the remaining 20% of the reference images and corresponding distorted counterparts constituted the test set; (3) The LIVE IQA database was repeated to be randomly partitioned for 1000 times and assessed the performance each time. Finally, the median performance indices across 1000 random trails were treated as the final evaluation results.

The performance indices of IQA algorithm contain the Pearson (Linear) Correlation Coefficient (LCC), the Spearman's Rank Ordered Correlation Coefficient (SROCC) and the Root Mean Squared Error (RMSE) between the predicted DMOS and the real DMOS provided by database. The better correlation with human perception means a value close to 0 for RMSE and a value close to 1 for LCC and SROCC. RMSE and LCC were calculated after passing the predicted DMOS through a logistic non-linearity shown in [27].

5.1. Correlation of individual feature vectors with human perception

In order to study the predictive power of each group of features considered here, we developed limited implementations of our model whereby only specific feature groups were deployed, then conducted a performance evaluation of each feature group on each distortion category, and across all distortion categories. Tables 2, 3 and 4 tabulate the median experimental results of these feature subsets over 1000 iterations in terms of median SROCC, LCC and RMSE, respectively. The feature groups f_1-f_6 and f_7-f_{12} are representatives of spatial and spectral entropy features, respectively.

From Tables 2–4, we observed that the spectral entropy features contributed more to performance than the spatial entropy features. However, the spatial entropy features are collectively a valuable, performance-enhancing complement to the spectral entropy features. Clearly, we have similar observations for these features within Figs. 2 and 3. The spectral entropy distributions of distortions are more different than spatial, especially clearly for “jp2k”, “blur” and “ff”. The spatial entropy distributions of “jp2k”, “blur” and “ff” are nearly confused with that of “ori”, while their spectral entropy distributions are further away from that of “ori”. And hence this confusion for spatial entropy makes spatial features perform worse than spectral ones. More importantly, the main reason is the greater ability of spectral entropy, including describing images' energy spectrum, emphasizing the main frequency and orientations and capturing texture variations.

Further, we also observed that our method performs much better for “jpeg” and “noise” than other distortions but that for “ff” is very poor, as can be shown from Tables 2–4. This is mainly because the spatial and spectral entropy distributions of “jpeg” and “noise” are farthest away from that of “ori” and that of “ff” is most similar to that of “ori”. So “jpeg” and “noise” will be easy to compute while “ff” is much harder.

Table 2
Median SROCC across 1000 train-test trials.

	JP2K	JPEG	Noise	Blur	FF	All
f_1-f_6	0.5874	0.8105	0.9485	0.7855	0.5966	0.6873
f_7-f_{12}	0.9390	0.9420	0.9509	0.9465	0.8917	0.9199
f_1-f_{12}	0.9420	0.9510	0.9784	0.9483	0.9035	0.9348

Table 3
Median LCC across 1000 train-test trials.

	JP2K	JPEG	Noise	Blur	FF	All
f_1-f_6	0.5858	0.8209	0.9434	0.7956	0.6579	0.6706
f_7-f_{12}	0.9424	0.9623	0.9516	0.9525	0.9145	0.9199
f_1-f_{12}	0.9464	0.9702	0.9806	0.9607	0.9198	0.9383

Table 4
Median RMSE across 1000 train-test trials.

	JP2K	JPEG	Noise	Blur	FF	All
f_1-f_6	19.6867	13.7856	7.2912	13.2175	16.6872	17.1644
f_7-f_{12}	8.1624	6.5453	6.7611	6.5939	8.8435	9.0211
f_1-f_{12}	7.8285	5.8467	4.3211	6.0027	8.5418	8.0039

5.2. Comparison with other IQA approaches

We conducted performance comparisons between SSEQ and three FR approaches (PSNR [28], SSIM [17] and VIF [15]) and other four NR IQA models (BIQI [8], DIIVINE [9], BLIINDS-II [10] and BRISQUE [11]). The results were shown in Tables 5–8. To make a fair comparison, we perform a similar random 20% test set selection for 1000 times to get median performance indices of the FR algorithms, since the FR algorithms do not need training. And we only test the FR approaches on the distorted images (not including the reference images of the LIVE IQA database). For the NR approaches, the same random 80–20% train test trails are conducted and the median performance was treated as the overall performance indices. We also calculated the standard deviations (STD) of the performance indices to gauge the algorithm stability in Table 8. Higher LCC and SROCC with the lower STD and RMSE mean excellent quality prediction performance.

5.3. Statistical significance testing

Although the preceding results suggest that there are meaningful differences in algorithm performance, we conducted statistical tests to determine significance. Fig. 4 shows a box plot of the SROCC distribution for some popular approaches across 1000 train-test trials. We applied the *t*-test between correlation scores generated

Table 5
Median SROCC across 1000 train-test trials on the LIVE IQA database.

	JP2K	JPEG	Noise	Blur	FF	All
PSNR	0.8990	0.8484	0.9835	0.8076	0.8986	0.8293
SSIM	0.9510	0.9173	0.9697	0.9513	0.9555	0.8996
VIF	0.9515	0.9104	0.9844	0.9722	0.9631	0.9521
BIQI	0.8551	0.7767	0.9764	0.9258	0.7695	0.7599
DIIVINE	0.9352	0.8921	0.9828	0.9551	0.9096	0.9174
BLIINDS-II	0.9462	0.9350	0.9634	0.9336	0.8992	0.9331
BRISQUE	0.9457	0.9250	0.9892	0.9511	0.9028	0.9468
SSEQ	0.9420	0.9510	0.9784	0.9483	0.9035	0.9348

Table 6
Median LCC across 1000 train-test trials on the LIVE IQA database.

	JP2K	JPEG	Noise	Blur	FF	All
PSNR	0.8837	0.8515	0.9817	0.8006	0.8939	0.8081
SSIM	0.9601	0.9485	0.9861	0.9537	0.9616	0.9100
VIF	0.9664	0.9478	0.9924	0.9774	0.9698	0.9520
BIQI	0.8414	0.7603	0.9732	0.9118	0.7342	0.7422
DIIVINE	0.9409	0.9097	0.9744	0.9393	0.9128	0.9116
BLIINDS-II	0.9493	0.9505	0.9614	0.9375	0.9079	0.9241
BRISQUE	0.9472	0.9330	0.9883	0.9463	0.9142	0.9365
SSEQ	0.9464	0.9702	0.9806	0.9607	0.9198	0.9383

Table 7
Median RMSE across 1000 train-test trials on the LIVE IQA database.

	JP2K	JPEG	Noise	Blur	FF	All
PSNR	7.5641	8.3269	3.0741	9.4289	7.3990	9.4973
SSIM	4.5389	5.0771	2.6584	4.6823	4.4855	6.6355
VIF	4.1943	5.0856	1.9608	3.3315	3.9624	4.9180
BIQI	13.7871	17.0133	5.3804	9.6562	15.5515	15.9547
DIIVINE	8.5703	10.6070	5.2137	8.0663	9.6520	9.9347
BLIINDS-II	8.1730	7.7658	6.5009	8.0696	9.7141	9.0473
BRISQUE	8.3625	9.3782	3.5294	7.5636	9.4359	8.3295
SSEQ	7.8285	5.8467	4.3211	6.0027	8.5418	8.0039

Table 8
Standard deviation of SROCC, LCC and RMSE across 1000 train-test trials on the LIVE IQA database.

	SROCC STD	LCC STD	RMSE STD
PSNR	0.0249	0.0566	2.5850
SSIM	0.0096	0.0144	0.4735
VIF	0.0067	0.0072	0.4323
BIQI	0.0654	0.0663	1.5994
DIIVINE	0.0273	0.0286	1.2705
BLIINDS-II	0.0235	0.0247	1.1653
BRISQUE	0.0117	0.0140	0.9455
SSEQ	0.0173	0.0197	1.1650

by the algorithms across the 1000 train-test trials. Table 9 tabulates the results of the *t*-test. The null hypothesis is that the mean correlation value of the row is equal to the mean correlation value of the column at the 95% confidence level. The alternative hypothesis is that the mean correlation value of the row is greater (or less) than the mean correlation value of the column. Table 9 indicates which row is statistically superior ('1'), statistically equivalent ('0') or statistically inferior ('−1') to which column.

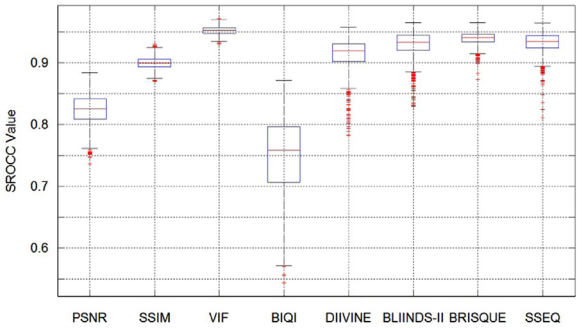

Fig. 4. Box plot of SROCC distributions of the compared IQA methods over 1000 trials on the LIVE IQA database.

From the above experimental results, SSEQ was found to be statistically superior to the FR approaches PSNR and SSIM and the NR approaches BIQI, DIIVINE and BLIINDS-II across all distortion categories. It was statistically close to the top-performing NR approach BRISQUE and statistically inferior to the top-performing FR approach VIF.

5.4. *Classification performance analysis*

We also calculated the classification accuracies of SSEQ, which is based on the 2-stage framework and tabulated the median classification accuracies of 1000 train-test trails, for each distortion category as well as across all distortion categories. The results can be shown in Table 10.

Further, to visualize what distortion categories may be confused with each other, a confusion matrix was plotted in Fig. 5, in which the value represents the probability of the distortion category on the vertical axis may be confused with that on the horizontal axis. The values in the confusion matrix are the mean classification accuracies across 1000 train-test trails. Obviously, Noise and JPEG are hard to be confused with other distortion categories while the other 3 distortion categories are more easily confused with one another. Since FF distortion is a combination of JP2K compression followed by packet-loss errors, and therefore shares distortion characteristics with JP2K compression distortion, it is obvious that FF distortion is more easily confused with JP2K compression distortion. We can also get expiation from Figs. 2 and 3, the entropy distributions of "noise" and "jpeg" are very specific and "jp2k", "blur" and "ff" have quite similar entropy distributions, leading to more confusion between them.

In the classification stage of the 2-stage framework, a weighted probability estimate is applied, rather than the "hard classification". However, classification confusion or the actual accuracy of the classifier does not necessarily adversely affect the performance of the IQA algorithm [9], since the distortions sometimes present similar artifacts.

5.5. *Time complexity analysis*

The amount of time consumed on feature extraction in IQA models is usually much higher than that computed for distortion classification and prediction with SVM. The latter mainly relies upon the size of the feature group. SSEQ only uses a 12-dimensional feature vector, so we only

Table 9
Results of the t-test performed between SROCC values.

	PSNR	SSIM	VIF	BIQI	DIIVINE	BLIINDS-II	BRISQUE	SSEQ
PSNR	0	−1	−1	1	−1	−1	−1	−1
SSIM	1	0	−1	1	−1	−1	−1	−1
VIF	1	1	0	1	1	1	1	1
BIQI	−1	−1	−1	0	−1	−1	−1	−1
DIIVINE	1	1	−1	1	0	−1	−1	−1
BLIINDS-II	1	1	−1	1	1	0	−1	−1
BRISQUE	1	1	−1	1	1	1	0	1
SSEQ	1	1	−1	1	1	1	−1	0

Table 10
Median classification accuracy across 1000 train-test trials.

	JP2K	JPEG	Noise	Blur	FF	All
Class. Acc (%)	61.7647	89.5644	100	70	46.6667	73.2919

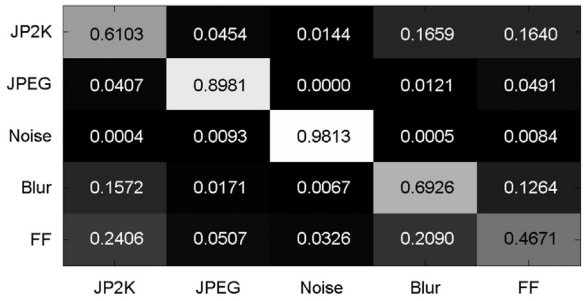

Fig. 5. Mean confusion matrix for stage 1 distortion classifier across 1000 train-test trials.

considered the time consumption of feature extraction here. In the time complexity experiments we used the image bikes.bmp with resolution 512×768, using a Dell desktop computer with quad-core CPU, 3.4 GHz and 4 GB RAM.

The process of feature extraction in SSEQ can be divided into spatial and spectral entropies features extraction. We list the percentage of time spent by each sub-step in Table 11.

We compared the time complexity of SSEQ with three good performance NR IQA approaches (DIIVINE, BLIINDS-II and BRISQUE). The unoptimized MATLAB code of each approach was run on the image bikes.bmp and the runtime was treated as the time complexity indices. The results are shown in Table 12. We observe that SSEQ is superior to DIIVINE and BLIINDS-II while significantly inferior to BRISQUE.

5.6. Database independence

In order to determine the degree of database independence of the proposed method, we tested SSEQ, BRISQUE and the 2 FR approaches for comparison on a portion of the TID2008 image database [29]. The TID2008 database consists of 25 reference images and 1700 distorted images over 17 distortion categories. Of these 25 reference images only 24 are natural images, so we tested it only on the 24 natural reference images over the 4 distortion categories in common with the LIVE database: JP2K, JPEG, Gaussian noise, and Gaussian blur.

We trained SSEQ on the entire LIVE database, and then tested it on the selected portion the TID2008 database. We report SROCC results in Table 13. Compared with PSNR, SSIM, VIF and BRISQUE, the correlation with subjective perception of SSEQ remained consistently competitive.

Table 11
Perceptage of time consumed by each stage of SSEQ.

	Percentage of time
Spatial entropy features extraction	0.55
Spectral entropy features extraction	0.45

Table 12
Comparison of time complexity (time for feature extraction) of three NR IQA algorithms.

	Time (s)
DIIVINE	25.40
BLIINDS-II	76.12
BRISQUE	0.142
SSEQ	1.824

Table 13
SROCC results obtained by training on the LIVE IQA database and testing on the TID2008 database.

	JP2K	JPEG	Noise	Blur	All
PSNR	0.8248	0.8753	0.9177	0.9335	0.8703
SSIM	0.9603	0.9354	0.8168	0.9598	0.9016
VIF	0.9697	0.9307	0.9136	0.9576	0.9403
BRISQUE	0.9037	0.9102	0.8227	0.8742	0.8977
SSEQ	0.8460	0.8661	0.8012	0.8354	0.8501

6. Conclusion

We proposed an efficient general-purpose no-reference (NR) image quality assessment (IQA) model called Spatial–Spectral Entropy-based Quality (SSEQ) index, which utilizes downsampled responses as inputs, then extracts a 12-dimensional local entropy feature vector from the inputs and learns to predict image quality scores from these features. We found that SSEQ delivers quality prediction performance that is competitive with top-performing FR

and NR IQA models. Currently, we are working towards creating effective NR video quality prediction models using space-time entropy features.

Acknowledgements

This work is supported in part by the National Natural Science Foundation of China under Grant 61133008 and in part by Specialized Fund for Joint Building Program of Beijing Municipal Education Commission.

References

[1] W. Lin, C.-C.J. Kuo, Perceptual visual quality metrics: a survey, J. Vis. Commun. Image Represent. 22 (4) (2011) 297–312.
[2] X. Gao, W. Lu, D. Tao, X. Li, Image quality assessment and human visual system, in: Proceedings of SPIE 7744, Visual Communications and Image Processing 2010, August 2010, 7744, pp. 77440Z-1–77440Z-10.
[3] Z.M.P. Sazzad, Y. Kawayoke, Y. Horita, No-reference image quality assessment for jpeg2000 based on spatial features, Signal Process. Image Commun. 23 (4) (2008) 257–268.
[4] X. Zhu, P. Milanfar, A no-reference sharpness metric sensitive to blur and noise, International Workshop on Quality of Multimedia Experience, San Diego, CA, July 2009, pp. 64–69.
[5] M.J. Chen, A.C. Bovik, No-reference image blur assessment using multiscale gradient, EURASIP J. Image Video Process. 1 (2011) 1–11.
[6] Z. Wang, A.C. Bovik, B.L. Evans, Blind measurement of blocking artifacts in images, in: Proceedings of the IEEE International Conference on Image Processing, September 2000, pp. 981–984.
[7] H.R. Sheikh, A.C. Bovik, L.K. Cormack, No-reference quality assessment using natural scene statistics: JPEG2000, IEEE Trans. Image Process. 14 (11) (2005) 1918–1927.
[8] 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.
[9] 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.
[10] M. Saad, A.C. Bovik, C. Charrier, Blind image quality assessment: a natural scene statistics approach in the DCT domain, IEEE Trans. Image Process. 21 (8) (2012) 3339–3352.
[11] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, IEEE Trans. Image Process. 1 (1) (2012) 4695–4708.
[12] L. Liu, H. Dong, H. Huang, A.C. Bovik, No-reference image quality assessment in curvelet domain, Signal Process. Image Commun. 29 (4) (2014) 494–505.
[13] J. Sponring, The entropy of scale-space, in: Proceedings of the 13th International Conference on Pattern Recognition, vol. 1, August 1996, pp. 900–904.
[14] H.R. Sheikh, A.C. Bovik, G. De Veciana, An information fidelity criterion for image quality assessment using natural scene statistics, IEEE Trans. Image Process. 14 (12) (2005) 2117–2128.
[15] H.R. Sheikh, A.C. Bovik, Image information and visual quality, IEEE Trans. Image Process. 15 (2) (2006) 430–444.
[16] Q. Hu, Z.X. Xie, Z.F. Wang, Y.H. Liu, Constructing NR-IQA function based on product of information entropy and contrast, in: International Symposium on Information Science and Engineering, vol. 2, December 2008, pp. 548–550.
[17] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[18] A.K. Moorthy, A.C. Bovik, Visual importance pooling for image quality assessment, IEEE J. Select. Top. Signal Process. 3 (2) (2009) 193–201.
[19] M. Narwaria, W. Lin, Objective image quality assessment based on support vector regression, IEEE Trans. Neural Netw. 21 (3) (2010) 515–519.
[20] M. Narwaria, W. Lin, SVD-based quality metric for image and video using machine learning, IEEE Trans. Syst. Man Cybern. Part B (Cybern.) 42 (2) (2012) 347–364.
[21] P. Carrai, I. Heynderickx, P. Gastaldo, R. Zunino, Image quality assessment by using neural networks, in: Proceedings of IEEE International Symposium on Circuits and Systems, vol. 5, 2002, pp. V-253–V-256.
[22] A. Bouzerdoum, A. Havstad, A. Beghdadi, Image quality assessment using a neural network approach, in: Proceedings of the 4th IEEE International Symposium on Signal Processing and Information Technology, 2004, pp. 330–333.
[23] B. Schölkopf, A.J. Smola, R.C. Williamson, P.L. Bartlett, New support vector algorithms, Neural Comput. 12 (5) (2000) 1207–1245.
[24] C.J.C. Burges, A tutorial on support vector machines for pattern recognition, Data Min. Knowl. Discov. 2 (2) (1998) 121–167.
[25] C.C. Chang, C.J. Lin. LIBSVM: a library for support vector machines. Available from: ⟨http://www.csie.ntu.edu.tw/∼cjlin/libsvm/⟩, 2001.
[26] H.R. Sheikh, Z. Wang, L. Cormack, A.C. Bovik. LIVE image quality assessment database release 2. Available from: ⟨http://live.ece.utexas.edu/research/quality⟩, 2006.
[27] H.R. Sheikh, M.F. Sabir, A.C. Bovik, A statistical evaluation of recent full reference image quality assessment algorithms, IEEE Trans. Image Process. 15 (11) (2006) 3440–3451.
[28] Z. Wang, A.C. Bovik, Mean squared error: love it or leave it? A new look at signal fidelity measures, IEEE Signal Process. Mag. 26 (1) (2009) 98–117.
[29] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, F. Battisti, TID 2008 a database for evaluation of full-reference visual quality assessment metrics, Adv. Mod. Radioelectron. 10 (2009) 30–45.