Sparse Representation for Blind Image Quality Assessment

Lihuo He[1] Dacheng Tao[2] Xuelong Li[3] Xinbo Gao[1]

[1] School of Electronic Engineering, Xidian University, Xi'an 710071, Shaanxi, P. R. China

[2] Centre for Quantum Computation and Intelligent Systems, Faculty of Engineering and Information Technology, University of Technology, Sydney, Broadway NSW 2007, Australia

[3] Center for **OPT**ical **IM**agery **A**nalysis and **L**earning (**OPTIMAL**), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, P. R. China

lihuo.he@gmail.com dacheng.tao@uts.edu.au xuelong_li@opt.ac.cn xbgao@mail.xidian.edu.cn

Abstract

Blind image quality assessment (BIQA) is an important yet difficult task in image processing related applications. Existing algorithms for universal BIQA learn a mapping from features of an image to the corresponding subjective quality or divide the image into different distortions before mapping. Although these algorithms are promising, they face the following problems: 1) they require a large number of samples (pairs of distorted image and its subjective quality) to train a robust mapping; 2) they are sensitive to different datasets; and 3) they have to be re-trained when new training samples are available. In this paper, we introduce a simple yet effective algorithm based upon the sparse representation of natural scene statistics (NSS) feature. It consists of three key steps: extracting NSS features in the wavelet domain, representing features via sparse coding, and weighting differential mean opinion scores by the sparse coding coefficients to obtain the final visual quality values. Thorough experiments on standard databases show that the proposed algorithm outperforms representative BIQA algorithms and some full-reference metrics.

Blind/no-reference image quality assessment (BIQA) is of crucial importance in image processing and analysis in practice. Especially, it is hard to develop a universal or generalized BIQA algorithm to handle different types of distortions [11]. Technically, there are two major issues in developing a robust BIQA algorithm, which are features and mapping model. The features are required to reflect the visual quality, and the mapping model needs to bridge the gap between the image features and visual quality.

In recent years, two groups of universal BIQA algorithms have been developed: 1) BLINDS [7, 8] directly maps image features to subjective quality without distinguishing different types of distortions [4, 6]; and 2) BIQI [13, 14] and LBIQ [20] first determine the distortion type of a test image and then employ an associated distortion-specific BIQA metric to predict the quality of the given image. Both groups of BIQA algorithms construct a black-box mapping from the image features to the image quality, and perform well on the LIVE II database.

However, these algorithms face the following three problems. First, they require a large number of expensive samples (pairs of distorted image and its subjective quality) to train a robust BIQA metric, let alone it is impossible to obtain a minimally sufficient number of training samples in many practical applications. Secondly, they are sensitive to different datasets. Given a BIQA metric trained on a database, it usually performs poorly on another database. That is because the algorithm contains database-specific parameters, e.g., BIQI and LBIQ need to select different models for different distortion subsets. Finally, they need to retrain the metric when the new samples are available. It is time-consuming and impractical for real time utilizations.

To address the aforementioned problems, we propose a novel BIQA algorithm based on *sparse representation* (S-R). It assumes that the feature space and subjective quality space share an almost same intrinsic manifold. This means images of similar quality have similar features. In this paper, we use *natural scene statistics* (NSS) to represent images. The proposed new algorithm, or SRNSS for short, can be sketched in the following steps. First, a dictionary is constructed by combining a set of images with different types of distortions. The NSS features of a test image are then extracted and encoded in the dictionary via sparse representation. Finally, the coding coefficients are utilized to linearly weight the corresponding subject quality (*differential mean opinion score*, or DMOS for short) for predicting the image quality.

The NSS features which are used in BIQA reflect the self-similar and certainly less specific in representation of image content. Hence, the proposed metric is not sensitive to different datasets. The feature sparse representation can

978-1-4673-1228-8/12/$31.00 ©2012 IEEE

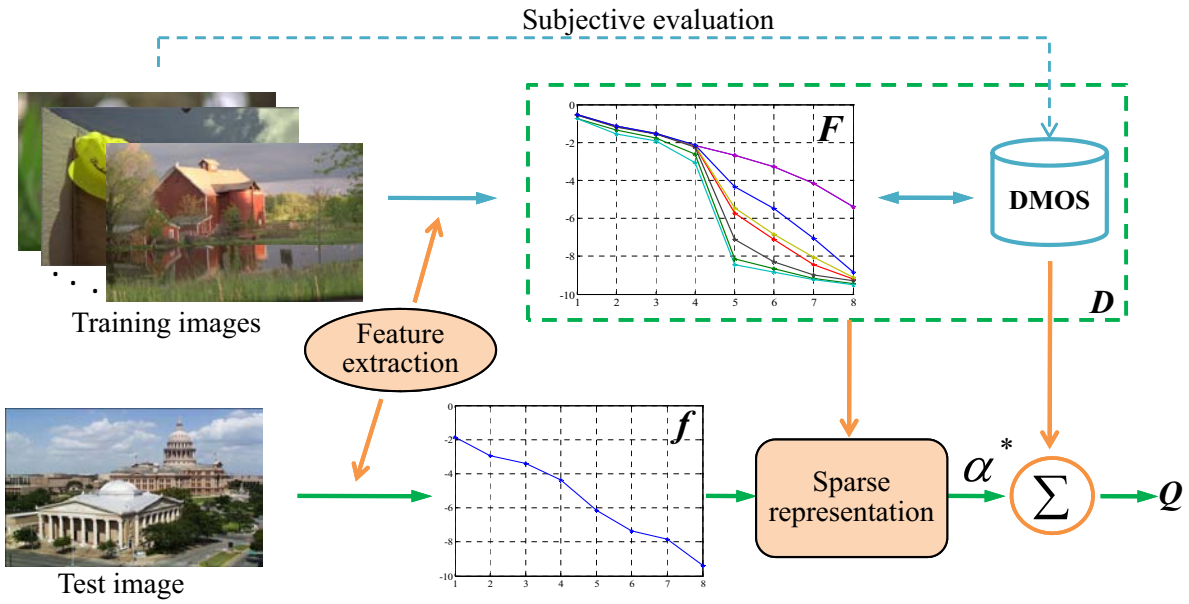


Figure 1. The proposed algorithm of the BIQA based on sparse representation in natural scene statistics.

effective avoid requirement of lots of samples. Furthermore, when the new samples are available, what we need to do is adding samples into dictionary and we do not retrain the whole metric.

The performance of the proposed BIQA metric is thoroughly validated on the LIVE II database [18] and other publicly available databases [10, 5, 1, 3]. The experimental results demonstrate that SRNSS for BIQA is consistent with the subjective quality, and outperforms some representative BIQA algorithms [7, 8, 13, 20, 16] as well as some classical full-reference image quality assessment metrics [23, 17].

The rest of the paper is organized as follows. Section 1 describes the proposed blind image quality assessment algorithm in detail. Experimental results of the proposed metric are given in Section 2. Section 3 concludes the paper.

1. Sparse Representation for BIQA

Figure 1 illustrates SRNSS for BIQA, which includes four modules, i.e., NSS features extraction, dictionary learning, sparse representation, and image quality quantification.

1.1. NSS features extraction

Natural scenes are defined by images and videos captured by high-quality devices operating in the visual spectrum [16]. They contain particular structures which can be described through natural scene statistics [24, 12, 19]. These NSS in the wavelet domain can be grouped into three levels, which are primary properties, secondary properties, and tertiary properties [12].

Primary properties reveal the significant statistical structure and can be further compacted in secondary properties which consist of non-Gaussianity and persistency. *General Gaussian distribution* (GGD) has been used to fit the non-Gaussianity distribution of wavelet coefficients. The persistency has been adopted to construct blind quality assessment for JPEG2000 compression images [16]. These features change significantly with different kinds of visual content. Therefore, they are not pertinent for developing a universal BIQA. Fortunately, the tertiary properties of the wavelet coefficients reflect the self-similar property of scenes. And the most important feature is the exponential decay across scales and it becomes stronger at fine scales. In particular, the exponential decay is certainly less specific in their representation of a particular image, and thus is suitable for representing the generalized behaviors of natural scenes.

An input image x is initially decomposed by the wavelet transform into subband wavelet coefficients C. A number of subbands are defined with the increasing of the wavelet subband frequency. Due to the similarity in the statistics of LH (low-high) and HL (high-low) subbands at the same scale, we calculate the features (magnitude, variance, and entropy) of the two subbands together, i.e., we do not distinguish LH and HL subbands at the same scale. By decomposing an image into 4 scales, we have 8 wavelet subbands in total. For each subband, we calculate the magnitude m_k to encode the generalized spectral behavior, the variance v_k to describe the fluctuations of the energy, and the entropy

e_k to represent the generalized information, according to

$$m_k = \frac{1}{N_k \times M_k} \sum_{j=1}^{N_k} \sum_{i=1}^{M_k} \log_2 |C_k(i,j)|, \quad (1)$$

$$v_k = \frac{1}{N_k \times M_k} \sum_{j=1}^{N_k} \sum_{i=1}^{M_k} \log_2 |C_k(i,j) - m_k|, \quad (2)$$

$$e_k = \sum_{j=1}^{N_k} \sum_{i=1}^{M_k} p[C_k(i,j)] \ln p[C_k(i,j)], \quad (3)$$

where $C_k(i,j)$ stands for the (i,j) coefficient of the k-th subband, M_k and N_k are the length and width of the k-th subband, respectively; $p[\cdot]$ is the probability density function of the subband.

The vertical and horizontal subbands with an identical mark in the same scale are combined through averaging after the above process. They can then be combined into a single vector

$$f = [m_1, m_2, ..., m_8, v_1, v_2, ..., v_8, e_1, e_2, ..., e_8]^T, \quad (4)$$

where f represents the exponential decay and generalizes the features of the natural images.

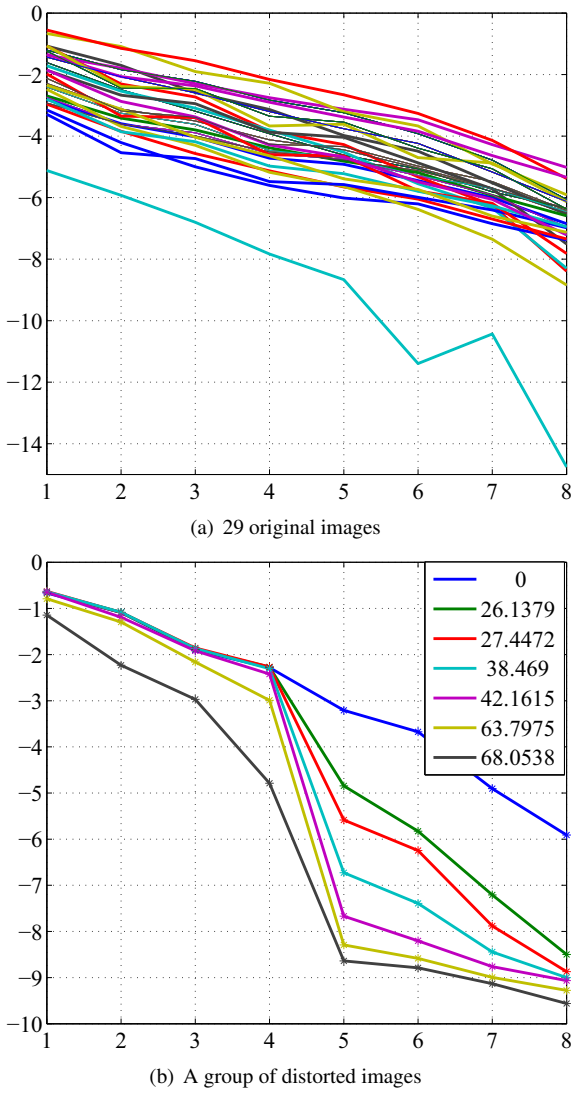
The distribution of magnitudes m_k vs. the sequence of subbands for the 29 original images in the LIVE II database on the log-log axes are shown in Figure 2(a). Since the magnitude spectra decay exponentially across scales, it is approximately a straight line on the log-log axes. And most of the reference images have similar exponential decay characteristics across scales. Figure 2(b) shows a group of distorted images with different qualities associated with the same reference image. It is obvious that one reference image and its corresponding distorted images have different departures. The increase of distortion degree brings out the sharper decrease and stronger persistence at fine scales. Furthermore, the distorted images with the same DMOS (differential mean opinion score) can be viewed roughly as having a similar tendency across scales; the quality of an input image can thus be estimated by a weighted average of the quality of images with similar tendency. In this paper, we solve this problem through applying the sparse representation to the extracted features.

1.2. Dictionary learning

In the proposed BIQA scheme, the dictionary learning is very simple by directly combining features and DMOS of the training images

$$\begin{bmatrix} \mathrm{D} \\ DMOS' \end{bmatrix} \doteq \begin{bmatrix} f_1 & f_2 & \cdots & f_n \\ DMOS'_1 & DMOS'_2 & & DMOS'_n \end{bmatrix}, \quad (5)$$

where the dictionary D is an $m \times n$ matrix ($m = 24$ and n is the number of training images). Vector $DMOS^{'}$ is the corresponding DMOS values of the training images.



(a) 29 original images

(b) A group of distorted images

Figure 2. Magnitudes of the wavelet coefficients decay exponentially across scales. (a) 29 original images, (b) A group of distorted images from one original image, showing different DMOS at top-right.

1.3. Sparse representation

Sparsity in human perception has been strongly supported by studies of the *human visual system* (HVS) [9]. And it is one of the most important features of HVS. In order to represent certain structures of a signal in a compact form, sparse representation can adaptively account for all or most of the information of a signal with the linear combination of a small number of elementary signals.

In this paper, sparse representation has been adopted to solve the sparsest representation problem which is similar to the cognitive behavior of object recognition and information representation. The sparse representation can consider

all possible supports and then adaptively choose the minimal number of training atoms. Thus, it can reduce the disturbance of images with large differences in visual quality in the proposed BIQA algorithm.

We find the most approximate and the sparsest representation of the feature f of a test image by using the dictionary $D = [f_1, f_2, ..., f_n]$ by solving

$$\alpha^* = \arg\min_{\alpha \in R^N} \|\alpha\|_0 \qquad s.t. \quad f = D\alpha, \tag{6}$$

where α^* is the sparse representation coefficient for f over the dictionary D and N is the number of atoms in D.

Since the l_0-minimization (6) is an NP-hard problem, the approximate solution via l_1-minimization [25] is usually used. Then we have

$$\alpha^* = \arg\min_{\alpha \in R^N} \|\alpha\|_1 \qquad s.t. \quad f = D\alpha. \tag{7}$$

This problem can then be transformed into an unconstrained optimization problem

$$\alpha^* = \arg\min_{\alpha \in R^N} \lambda\|\alpha\|_1 + \|f - D\alpha\|_2, \tag{8}$$

where the parameter λ is a positive constant balance the fidelity term and the sparse regularization term. The least absolute shrinkage and selection operator [21] is used to solve the unconstrained convex optimization problems (8).

1.4. Image quality quantification

Through the previous step, the sparse solution vector α^* of f over the dictionary is obtained. According to the assumption that *if images have identical quality values their features have similar distributions*, the quality of the test image can be quantified by the qualities of the training images, i.e.,

$$Q = \frac{\sum_{i=1}^{N} \alpha_i^* DMOS_i'}{\sum_{i=1}^{N} \alpha_i^*}, \tag{9}$$

where $DMOS_i'$ is the differential mean opinion scores of the i-th image in the dictionary. The estimation Q, ranging from 0 to 100, is the final quality score of the test image, and a lower value of Q implies a higher quality of the test image.

2. Experiments

In order to validate the effectiveness of the proposed S-RNSS for BIQA, we conduct four groups of experiments: the consistency experiment, the rationality experiment, the sensitivity experiment, and the extensibility experiment.

In experiments, LIVE II [18], TID [10], CSIQ [5], IVC [1] and MICT [3] are used as the standard databases. Additionally, the model parameters of the proposed SRNSS are trained on the LIVE II database.

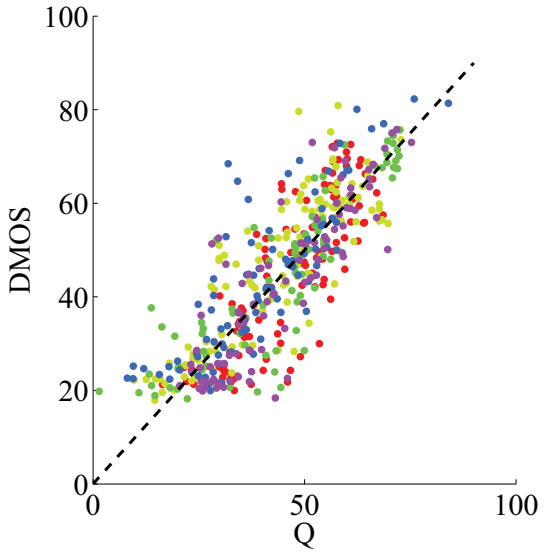


Figure 3. Predicted Q *vs.* subjective DMOS. 15 original images and their associated distorted images as the training set, the rest as the test set. The original images are removed.

VQEG provides the comparison criterion among the metrics [22]. Two important evaluation criteria are then used to compare the performance. **Criterion 1** is the *Pearson linear correlation coefficient* (LCC) between the subjective quality DMOS and the corresponding objective quality, which provides an evaluation of the prediction accuracy. **Criterion 2** is the *Spearman rank-order correlation coefficient* (SROCC) which estimates the test for agreement between the rank orders of DMOS and model predictions. It is considered as a measure of the prediction monotonicity. In addition, the *root mean square error* (RMSE) and the *mean absolute error* (MAE) of the fitting procedure after the nonlinear mapping are calculated to quantify the direct errors.

2.1. Consistency experiment

In this experiment, the LIVE II database is utilized as the benchmark database. First, part of original images and their associated distorted images are randomly selected for BIQA model training, with the rest for test. We run our algorithms 100 times in this way to verify that our algorithms are robust to the image content. The performance evaluation metrics of LCC, SROCC, RMSE and MAE are the average of the 100 random experimental results. The full-reference metrics: PSNR, SSIM [23], VIF [15], IFC [17], and the blind metrics: BIQI [7], DIIVINE [8], BLINDS [13], BLINDS-II [14], LBIQ [20], NSS [16] are compared with the proposed SRNSS. The evaluation results for all the compared metrics are given in Table 1.

Table 1 shows the performance of the metric on the five subsets: JPEG2000, JPEG, WN, Gblur, FF and the entire

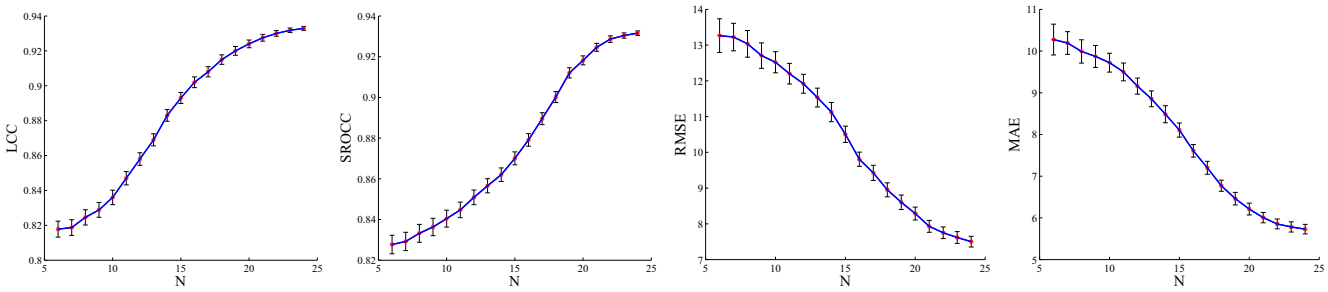


Figure 4. Performances *vs.* number of original images. The images and their distorted images are selected randomly as the training set.

LIVE II database. Although VIF performs best among all metrics, but it is a full reference IQA metric that needs the reference image to evaluate the distorted images. For BIQA metrics, we divide them into two groups by the number of selected original images used in BIQA model training - 15 original images named group-I and 23 original images named group-II. The proposed SRNSS_1 which belongs to group-I obtains the best performance among the benchmarks: NSS, BIQI, LBIQ and BLINDS. This is due to the effectiveness of the tertiary NSS features of the natural images captured by the statistical model and the rationality of sparse representation. The statistical features reflect the visual quality and the sparse representation can then find the best representation for the features of the test image.

Group-II includes the proposed SRNSS_2, DIIVINE and BLINEDS-II which are the improved versions of BIQI and BLINEDS, in which the performance of the proposed SRNSS_2 is better than DIIVINE for JPEG2000, JPEG, F-F and the entire database. BLINDS-II demonstrates better performance than SRNSS_2 on an individual sub-database, but SRNSS_2 shows the best performance over the entire database. That is mainly because of the following two reasons. First, the features used in SRNSS_2 are insufficient for representing the image quality. The generalized parametric model of the extracted DCT coefficients is used in BLINDS-II and 88 features extracted from NSS are utilized in DIIVINE. However, only the tertiary features of NSS, which has 8 features, are utilized in the SRNSS. Secondly, the test set of sub-databases only has 30~48 images and the strategy of random content-separated divisions has a tremendous influence on the performance shown in Figure 4. It influences all the metrics, including the training process.

Figure 3 shows the scatter plots of DMOS versus the predicted score of the proposed algorithm (trained on 15 original images and their distorted images). It demonstrates that the proposed algorithm is stable across different distortions and demonstrates good performance across the entire LIVE II database.

Figure 4 presents the relationship between the number of selected original images and the quality prediction performance. The number ranges from 6 to 24. For each, we run our algorithms 100 times and obtain the average and the 95% confidence interval. It demonstrates that the strategy of the random content-separated divisions has a tremendous influence on the performance.

2.2. Rationality experiment

In this subsection, the rationality experiment [2] is conducted on four types of distortions: JPEG2000, JPEG (the compression rate is R), blurring (the window size is W) and Gaussian noise with zero-mean (the variance is V). The prediction trends of the proposed algorithm for image "*Window*" with different types of distortions are presented in Figure 5. In order to facilitate observation, part (flower) of images is shown in figure. It is found that the prediction tends to rise when the degree of the distortion is increasing, because the proposed algorithm has the same meaning of the quality with DMOS which has a higher value with lower visual perception quality. Hence, it is consistent with the decrease in visual quality of the distorted images, which demonstrates the rationality of the proposed method.

The most typical artifacts or distortions are ringing and blurring in JPEG2000 because many small wavelet coefficients are set to zero when quantifying. The distortions and blurring will lose the high frequency information when the images are compressed. While the JPEG compression brings about blurring within blocks and blocking artifacts across block boundaries. According to the experimental observation and theoretical derivation in [12], NSS captures the property and the statistical features reflect distortions or artifacts. As illustrated in Figure 5, the tendency is not obvious when the noise reaches a certain degree, which is consistent with subjective perception. When the noise in an image is too severe, even the HVS finds it difficult to quantify the quality of the images, so the result in Figure 5(d) is tolerable and is consistent with the human perception.

2.3. Sensitivity experiment

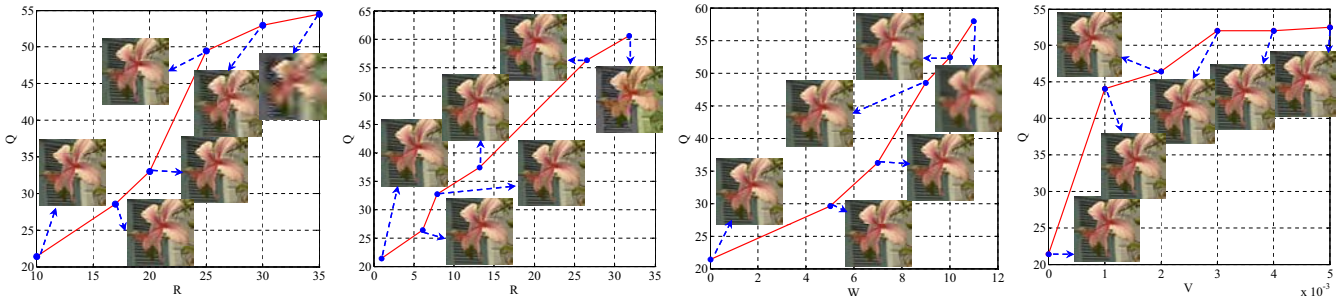
Sensitivity experiment tests whether a BIQA metric has reasonable prediction when the images have the same or similar PSNR (or other benchmark metrics) values. Figure

Distortion		JPEG2000				JPEG			
Metric	Type	LCC	SROCC	RMSE	MAE	LCC	SROCC	RMSE	MAE
PSNR	FR	0.8962	0.8898	7.1865	5.5283	0.8596	0.8409	8.1700	6.3797
SSIM [23]	FR	0.9367	0.9317	5.6706	4.4332	0.9283	0.9028	5.9468	4.4846
IFC [15]	FR	0.9027	0.8920	6.9720	5.4616	0.9047	0.8661	6.8129	4.7920
VIF [17]	FR	0.9615	0.9527	4.4493	3.4450	0.9430	0.9131	5.3212	3.8070
NSS [16]	NR	0.9210	0.9081	9.5060	8.3307	0.3661	0.1798	22.5284	18.8351
BIQI [7]	NR	0.8086	0.7995	14.8427	——	0.9011	0.8914	13.7552	——
LBIQ [20]	NR	——	0.9000	——	——	——	0.9200	——	——
BLINDS [13]	NR	——	0.9219	——	——	——	0.8391	——	——
DIIVINE [8]	NR	0.9220	0.9130	9.6600	——	0.9210	0.9100	12.2500	——
BLINDS-II [14]	NR	0.9630	0.9506	——	——	0.9793	0.9419	——	——
SRNSS$_1$	NR	**0.8859**	**0.8626**	**10.8831**	**7.9653**	**0.8901**	**0.8713**	**10.9128**	**7.8649**
SRNSS$_2$	NR	**0.9359**	**0.9283**	**7.8916**	**6.0152**	**0.9391**	**0.9306**	**7.9481**	**5.9681**
Distortion		WN				Gblur			
Metric	Type	LCC	SROCC	RMSE	MAE	LCC	SROCC	RMSE	MAE
PSNR	FR	0.9858	0.9853	2.6797	2.1639	0.7834	0.7816	9.7723	7.7425
SSIM [23]	FR	0.9695	0.9629	3.9163	3.2566	0.8740	0.8942	7.6391	5.7595
IFC [15]	FR	0.9581	0.9383	4.5738	3.8162	0.9608	0.9590	4.3604	3.4103
VIF [17]	FR	0.9839	0.9857	2.8514	2.3039	0.9744	0.9731	3.5334	2.8182
NSS [16]	NR	0.8217	0.8774	12.5284	10.4882	0.7007	0.7366	15.5178	10.8844
BIQI [7]	NR	0.9538	0.9510	8.4094	——	0.8293	0.8463	10.2347	——
LBIQ [20]	NR	——	0.9700	——	——	——	0.8800	——	——
BLINDS [13]	NR	——	0.9735	——	——	——	0.9569	——	——
DIIVINE [8]	NR	0.9880	0.9840	4.3100	——	0.9230	0.9210	7.0700	——
BLINDS-II [14]	NR	0.9854	0.9783	——	——	0.9481	0.9435	——	——
SRNSS$_1$	NR	**0.8802**	**0.8605**	**10.2694**	**8.0151**	**0.8654**	**0.8601**	**10.8634**	**8.3124**
SRNSS$_2$	NR	**0.9404**	**0.9382**	**7.9705**	**6.1058**	**0.9356**	**0.9327**	**7.5907**	**6.0861**
Distortion		FF				Entire database			
Metric	Type	LCC	SROCC	RMSE	MAE	LCC	SROCC	RMSE	MAE
PSNR	FR	0.8895	0.8903	7.5158	5.800	0.8240	0.8197	9.1236	7.3249
SSIM [23]	FR	0.9428	0.9411	5.4846	4.2968	0.8634	0.8510	8.1262	6.2752
IFC [15]	FR	0.9614	0.9630	4.5280	3.6196	0.9106	0.9128	6.6564	5.1822
VIF [17]	FR	0.9618	0.9649	4.5022	3.5469	0.9501	0.9526	5.0241	3.8866
NSS [16]	NR	0.7224	0.7383	15.2775	10.7881	0.4946	0.3333	20.0911	15.8479
BIQI [7]	NR	0.7328	0.7067	19.2911	——	0.8205	0.8195	15.6223	——
LBIQ [20]	NR	——	0.7800	——	——	——	0.8900	——	——
BLINDS [13]	NR	——	0.7503	——	——	——	0.7996	——	——
DIIVINE [8]	NR	0.8880	0.8630	12.9300	——	0.9170	0.9160	10.9000	——
BLINDS-II [14]	NR	0.9436	0.9268	——	——	0.9232	0.9202	——	——
SRNSS$_1$	NR	**0.8728**	**0.8651**	**10.3291**	**7.6245**	**0.8862**	**0.8761**	**10.7287**	**8.1056**
SRNSS$_2$	NR	**0.9473**	**0.9406**	**7.1571**	**5.2684**	**0.9318**	**0.9304**	**7.6183**	**5.8729**

Table 1. The performance of image quality assessment metrics on the LIVE II database.

6 shows four images: (a) the original "*Window*" image, (b) the contrast stretching image, (c) the mean shift image, and (d) the JPEG compressed image. Table 2 reports the quality of PSNR, SSIM, IFC, VIF, NSS, BIQI, BLINDS and the proposed SRNSS for images shown in Figure 6.

From Figure 6 and Table 2, we find that the quality of the contrast stretching image is better than that of the mean shift image. However, the prediction of fidelity or similarity based metrics (SSIM, IFC and VIF) and BIQA metrics (NSS, BIQI and BLINDS) are inconsistent with human visual perception. In fact, the contrast stretching image has a higher visual quality than the original image, and this pro-



(a) JPEG2000 compression (b) JPEG compression (c) Blurring (d) Gaussian noise contamination

Figure 5. Quality trend of *Window* with different types of distortion using the proposed algorithm.

Metric	(b)	(c)	(d)
PSNR	30.53	30.45	30.61
SSIM [23]	0.9379	0.9939	0.7186
IFC [15]	12.28	72.00	1.14
VIF [17]	0.9998	0.9979	0.1830
NSS [16]	78.23	78.28	75.13
BIQI [7]	29.94	26.70	66.04
BLINDS [13]	28.56	27.34	69.25
SRNSS	**27.61**	**32.52**	**70.91**

Table 2. The quality values of different metrics for images in Figure 6.

Database	TID [10]	CSIQ [5]	IVC [1]	MICT [3]
PSNR	0.5643	0.8772	0.7192	0.6355
SSIM [23]	0.6387	0.8060	0.7924	0.7979
IFC [15]	0.5692	0.7482	0.8978	0.8387
VIF [17]	0.7496	0.9193	0.8966	0.9086
NSS [16]	0.2027	0.5667	0.4266	0.4541
BIQI [7]	0.4192	0.6601	0.5346	0.6853
BLINDS [13]	0.5086	0.7529	0.7013	0.7924
SRNSS	**0.7327**	**0.8157**	**0.7943**	**0.8094**

Table 3. The LCC of different metrics on other publicly available databases.

cessing can be considered to be an enhanced method. The proposed algorithm is trained on the LIVE II database that contains high-resolution color images, so the parameter setting of the proposed metric has a higher quality basis and suggests that the "*Window*" image (a) is not the best quality image. For the contrast stretching and mean shift distortions, similar regular patterns of tertiary properties exist. Hence, we can predict the two types of distorted image through the dictionary constructed by the LIVE II database. Results consistent with human visual perception can be obtained, sensitivity can be verified and the generalization of the proposed algorithm can be demonstrated.

2.4. Extensibility experiment

In order to verify the effectiveness of the generalization, the expansibility experiment is conducted in this subsection. The model parameters are trained on the LIVE II database. Table 3 presents LCC of PSNR, SSIM, IFC, VIF, NSS, BIQI, BLINDS and the proposed SRNSS on other publicly available databases. The results show that the proposed SRNSS performs better than other BIQA (NSS, BIQI and BLINDS) on other databases: TID [10], CSIQ [5], IVC [1] and MICT [3]. This good performance is mainly attributable to two factors. First, the LIVE II database includes almost all kinds of image distortions. For example, the TID database includes additive noise, blurring, JPEG and JPEG2000, which are included in the LIVE II database. Secondly, the distortions that are not included in the LIVE II database also follow the tertiary features of NSS. Therefore, the experimental results suggest that the proposed SRNSS trained on the LIVE II database can be applied to different databases.

3. Conclusions

This paper presents a universal BIQA metric, i.e., SRNSS, which uses the sparse representation of the tertiary natural scene statistics to blindly evaluate image quality. Comprehensive experimental results demonstrate that 1) SRNSS can blindly yet effectively evaluate the quality of images with different kinds of distortions; 2) SRNSS performs consistently with the subjective perception and is stable across various kinds of distortions in terms of LCC, SROCC, RMSE and MAE criteria; 3) SRNSS outperforms conventional image quality assessment algorithms on publicly available databases; and 4) SRNSS, trained on the LIVE II database, has a good expansibility (or generalization) on other publicly available databases comparing with the representative BIQA algorithms under the same setting.

(a) Reference image (b) Contrast stretching (c) Mean shift (d) JPEG compression

Figure 6. *Window* image with the same PSNR but different perceived quality.

4. Acknowledgement

This work is supported by the National Basic Research Program of China (973 Program) (Grant No. 2012CB316400), the National Natural Science Foundation of China (Grant Nos. 61125204, 61172146, 61125106, 91120302, 61001203, 60832005, 61072093), the Fundamental Research Funds for the Central Universities, the Ph.D. Programs Foundation of Ministry of Education of China (Grant No. 20090203110002), and the Australian ARC discovery project (ARC DP-120103730).

References

[1] P. L. Callet and F. Autrusseau. Subjective quality assessment-ivc database, 2006. Available Online: http://www.irccyn.ec-nantes.fr/ivcdb/.

[2] X. Gao, W. Lu, D. Tao, and X. Li. Image quality assessment based on multiscale geometric analysis. *IEEE TIP*, 18(7):1409–1423, 2009.

[3] Y. Horita, K. Shibata, Y. Kawayoke, and Z. M. P. Sazzad. Mict image quality evaluation database, 2000. Available Online: http://mict.eng.u-toyama.ac.jp/mictdb.html.

[4] Y. Ke, X. Tang, and F. Jing. The design of high-level features for photo quality assessment. *In CVPR*, 2006.

[5] E. C. Larson and D. M. Chandler. Categorical image quality (csiq) database, 2009. Available Online: http://vision.okstate.edu/csiq.

[6] Y. Luo and X. Tang. Photo and video quality evaluation: focusing on the subject. *In ECCV*, 2008.

[7] A. K. Moorthy and A. C. Bovik. A two-step framework for constructing blind image quality indices. *IEEE SPL*, 17(6):513–516, 2010.

[8] A. K. Moorthy and A. C. Bovik. Blind image quality assessment: from scene statistics to perceptual quality. *IEEE TIP*, 2011. to appear.

[9] B. Olshausen and D. Field. Sparse coding with an overcomplete basis set: a strategy employed by v1? *Vision Research*, 37:3311–3325, 1997.

[10] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti. Tid2008 - a database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radioelectronics*, 10:30–45, 2009.

[11] G. Ramanarayanan, J. Ferwerda, B. Walter, and K. Bala. Visual equivalence: Towards a new standard for image fidelity. *In SIGGRAPH*, 2007.

[12] J. K. Romberg, H. Choi, and R. G. Baraniuk. Bayesian tree-structured image modeling using wavelets-domain hidden markov models. *IEEE TIP*, 10(7):1056–1068, 2001.

[13] M. A. Saad, A. C. Bovik, and C. Charrier. A dct statistics-based blind image quality index. *IEEE SPL*, 17(6):583–586, 2011.

[14] M. A. Saad, A. C. Bovik, and C. Charrier. Model-based blind image quality assessment using natural dct statistics. *IEEE TIP*, 2011. perprint.

[15] H. R. Sheikh and A. C. Bovik. Image information and visual quality. *IEEE TIP*, 15(2):430–444, 2006.

[16] H. R. Sheikh, A. C. Bovik, and L. Cormack. No-reference quality assessment using natural scene statistics: Jpeg2000. *IEEE TIP*, 14(11):1918–1927, 2005.

[17] H. R. Sheikh, A. C. Bovik, and G. de Veciana. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE TIP*, 14(12):2117–2128, 2005.

[18] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik. Live image quality assessment database release 2, 2003. Available Online: http://live.ece.utexas.edu/research/quality.

[19] E. P. Simoncelli and B. A. Olshausen. Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24:1193–216, 2001.

[20] H. Tang, N. Joshi, and A. Kapoor. Learning a blind measure of perceptual image quality. *In CVPR*, 2011.

[21] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Royal. Statist. Soc B.*, 58(1):267–288, 1996.

[22] VQEG. Validation of reduced-reference and no-reference objective models for standard definition television, phase i, 2009. Available Online: http://www.vqeg.org/.

[23] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, 13(4):600–612, 2004.

[24] Y. Weiss and W. T. Freeman. What makes a good model of natural images? *In CVPR*, 2007.

[25] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE T-PAMI*, 31(2):1–18, 2009.