Blind Image Quality Assessment Based On a New Feature of Nature Scene Statistics

Li Song, Chen Chen, Yi Xu, Genjian Xue, Yi Zhou

[#]*Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University*
800 Dong Chuan Road, Shanghai, 200240 China
song_li@sjtu.edu.cn,445452902@qq.com,xuyi@sjtu.edu.cn
xgjsword@gmail.com,zy_21th@sjtu.edu.cn

Abstract— **A recently proposed model, known as blind/referenceless image spatial quality evaluator (BRISQUE), achieves the state-of-the-art performance in context of blind image quality assessment (IQA). This model used the predefined generalized Gaussian distribution (GGD) to describe the regularity of natural scene statistics, introducing fitting errors due to variations of image contents. In this paper, a more generalized model is proposed to better characterize the regularity of extensive image contents, which is learned from the concatenated histograms of mean subtracted contrast normalized (MSCN) coefficients and pairwise products of MSCN coefficients of neighbouring pixels. The new feature based on MSCN shows its capability of preserving intrinsic distribution of image statistics. Consequently support vector machine regression (SVR) can map it to more accurate image quality scores. Experimental results show that the proposed approach achieves a slight gain from BRISQUE, which indicates the crafted GGD modelling step in BRISQUE is not essential for final performance.**

Index Terms— **Image quality assessment, blind quality assessment, natural scenes statistics, concatenated histogram, support vector machine regression**

I. Introduction

Recent works on the natural scene statistics (NSS) show that the features derived from the NSS possess distinct patterns of different distortions. Based on this observation, several no-reference image quality assessment (NR-IQA) models have been developed and achieve state-of-the-art performance [1-6]. These approaches used the predefined generalized Gaussian distribution (GGD) to describe the regularity of natural scene statistics. As we know, the parametric models will introduce fitting errors due to variations of image contents.

In this paper, we propose a more generalized feature for distortion-generic NR-IQA approach, which directly uses the concatenated histograms of the spatial domain NSS as a feature to learn image quality scores. This non-parametric representation model shows its strong potential to characterize the regularity of natural scene statistics. Experiments show that the proposed method highly correlates with the human perception and outperforms the known best NR-IQA method. Furthermore, our work indicates that the crafted CGD/AGGD modelling step is not essential for final performance.

The rest of this paper is organized as follows. We review recent NSS based NR-IQA works first, and present the concatenated histograms feature extraction algorithm in Section II. Experimental results and comparisons are given in Section IV. Section V concludes the paper.

II. Previous Related Works

NSS based NR-IQA methods have drawn a lot of attention recently for their good performance and distortion-generic advantages. Among them, DIIVINE [1], BLIINDS-II [2], and BRISQUE [3] can be treated as state-of-the-art ones because of their excellent performances.

DIIVINE [1] deploys the statistics derived from an NSS wavelet coefficient model with two scales and six orientations. BLIINDS-II [2] measures image quality in the DCT domain, where a small number of features are extracted from a generalized NSS based model of local DCT coefficients.

Different from the above ones, BRISQUE [3] is a spatial domain NSS based NR-IQA model, which has achieved better performance than DIIVINE and BLIINDS-II while keeping low computational complexity. The underlying features are derived from the distribution of locally normalized luminance and products of locally normalized luminance in spatial domain. In specific, BRISQUE first estimates parameters of presumed generalized Gaussian distribution (GGD) or asymmetric generalized Gaussian distribution (AGGD) by fitting histogram of mean subtracted contrast normalized (MSCN) coefficients or pairwise products of MSCN coefficients from neighbouring pixels. Then the estimated model parameters are used as features for SVR which maps them to quality scores.

BRISQUE used parametric model to describe the regularity of natural scene statistics, introducing fitting errors due to variations of image contents. Our algorithm differs from BRISQUE method by directly using raw histograms of MSCN coefficients and pairwise products of MSCN coefficients as a discriminant feature. The proposed non-parametric representation of natural scene statistics can alleviate fitting error problem. Consequently, support vector machine regression (SVR) can map it to more accurate image quality scores.

III. CONCATENATED HISTOGRAMS OF LOCAL MCSN COEFFICIENTS

Similar to the workflow of BRISQUE, we first extract a statistical feature from an input image. Then such a feature is transformed into the quality scores using a regression algorithm.

A. *Natural Scene Statistics After Normalization*

Ruderman's research [7] on spatial luminance statistics show that imposing a local non-linear operation on the log-contrast luminance map can correct local mean displacements and normalize the local variance. For natural images, such normalization pre-processing tends to achieve decorrelation effect and produce a normal Gaussian distribution. It is defined as:

$$I'(i,j) = \frac{I(i,j)-\mu(i,j)}{\sigma(i,j)+C}\,, \qquad (1)$$

Where $i \in \{1,2,\dots,M\}$, $j \in \{1,2,\dots,N\}$; M, N denotes the image height and width, respectively; C=1 is a constant to prevent the denominator from being zero; I refers to the original image and I' is the MSCN coefficients. $\mu(i,j)$ and $\sigma(i,j)$ are defined as:

$$\mu(i,j) = \sum_{k=-K}^{K}\sum_{l=-L}^{L} w_{k,l} I_{k,l}(i,j), \qquad (2)$$

$$\sigma(i,j) = \sqrt{\sum_{k=-K}^{K}\sum_{l=-L}^{L} w_{k,l}(I_{k,l}(i,j)-\mu(i,j))^2}, \qquad (3)$$

Where $w = \{w_{k,l} | k=-K,\dots,K, l=-L,\dots,L\}$ is a 2D circularly-symmetric Gaussian weighting function sampled out to three standard deviations and rescaled to unit volume. We adopt K=L=3 in this paper. The local mean value μ is computed within the block surrounding the image position(i,j), σ is the standard deviation of I within the block.

Similar to BRISQUE, we compute both the distribution of MSCN coefficients and the distribution of pairwise products of neighbouring MSCN coefficients along four orientations (horizontal (*H*), vertical (*V*), main-diagonal (*D1*), and secondary-diagonal (*D2*)), which are defined as:

$$H(i,j) = I'(i,j)I'(i,j+1) \qquad (4)$$

$$V(i,j) = I'(i,j)I'(i+1,j) \qquad (5)$$

$$D1(i,j) = I'(i,j)I'(i+1,j+1) \qquad (6)$$

$$D2(i,j) = I'(i,j)I'(i+1,j-1) \qquad (7)$$

B. *Concatenated Histograms of Local MCSN Coefficients*

Rather than using the parametric models of GGD and AGGD to describe natural scene statistics, we directly use a concatenation of *raw* histogram of local MSCN coefficients as the feature to discriminate a natural undistorted image against its various distorted counterparts. The parametric models of GGD/AGGD in BRISQUE only reflect shape and variance of a distribution, introducing fitting errors of original distributions. In fact, normalization operation in formula (1) can efficiently reduce high-order statistical dependencies of natural sensory signal [8]. Thus resulting signal should be less dependent upon specific distribution hypothesis. This observation motivates us to directly use original statistical information like histograms to construct a discriminant feature.

Specifically we propose a new image feature, *concatenated histograms of local MSCN coefficients* (dotted as ***CH-MCSN***) for quality assessment, which is established as follows:

(1) Compute five spatial histograms of I', *H*, *V*, *D1*, *D2*. The histograms range are set to be [-2, 2] for MSCN coefficients and [-1, 1] for pairwise products respectively in consideration of their dominant value range.

(2) Divide the range into 40 bins (further discussion can be found at Section 4.2). Each histogram represents a distinct *40*-dimension feature vector of the input image. We obtain histograms at multi-scales – the original image scale and a reduced resolution (low pass filtered and down sampled by factor of two) to get more complete description of image statistics.

(3) Construct a high dimension feature vector by concatenating all the extracted histogram vectors.

Fig.1 gives an illustration of computation of the proposed feature. Given this feature, a support vector machine regressor (SVR) is then applied to train a mapping model from feature space to image quality scores. In the next section, experimental results demonstrate that high correlation exists among ten feature vectors. As a result, we choose only a subset of possible combination of histograms in application with trivial performance loss.

IV. EXPERIMENTAL RESULTS

We use the LIVE IQA database [9] to test the performance of the proposed method, which consists of 29 reference images with 779 distorted images. Distortion categories include JPEG2000 (JP2K), JPEG, white noise (WN), Gaussian blur (Blur), and a Rayleigh fast-fading channel simulation (FF). Each of the distorted images has an associated difference mean opinion score (DMOS). We randomly divide the dataset into the training (80%) subsets and test (20%) subsets. For each split, a SVR is fit to the proposed features extracted from training data, and the predictive accuracy is assessed on the test data.

The Spearman's rank ordered correlation coefficient (SROCC) and Pearson linear correlation coefficient (LCC) between the predicted score and DMOS are used as the performance indices. Before providing a comparison between the proposed approach and BRISQUE, we investigate how the configuration of parameters in our algorithm affects performance.

A. *Performance with Different Combination of Histograms of Local MCSN Coefficients*

To investigate prediction performance with different combination of histograms, we randomly select a pair of train-test dataset. A logistic non-linearity in [9] is applied to the results before computing the LCC. The results are listed in Table I

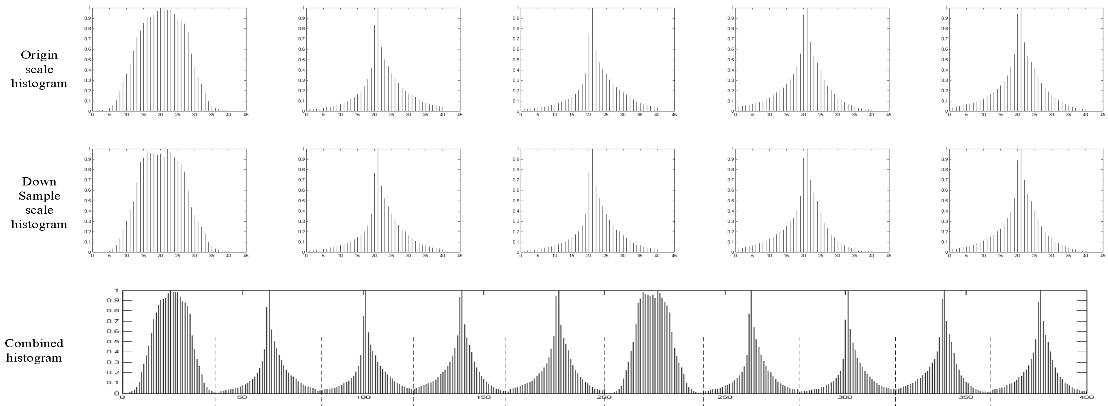

Figure. 1. The illustration of concatenated histograms of local MSCN coefficients

TABLE I
SROCC/LCC WITH DIFFERENT COMBINATION OF HISTOGRAMS OF LOCAL MSCN COEFFICIENTS

Scale numbers	Categories	SROCC/LCC
2	**I', H, V, D1, D2**	**0.977/0.966**
2	I'	0.958/0.954
2	H, V, D1, D2	0.972/0.959
2	H, V	0.971/0.959
2	D1, D2	0.967/0.956
2	H	0.959/0.953
2	V	0.962/0.956
2	D1	0.954/0.949
2	D2	0.953/0.947
1	I', H, V, D1, D2	0.973/0.966
1	I'	0.899/0.898
1	**H, V, D1, D2**	**0.973/0.966**
1	H, V	0.958/0.958
1	D1, D2	0.943/0.941
1	H	0.925/0.927
1	V	0.932/0.930
1	D1	0.924/0.922
1	D2	0.929/0.930

We provide the performance comparison of 18 cases under one or two scales with different combination of histograms of I', H, V, D1 and D2. We observe that high correlation exists among different categories and scales. Although best performance comes from full combination of histograms at two scales, there are non-significant performance differences between the top two in Table 1. As a result, we choose to extract single scale histogram of pairwise MSCN coefficients products along 4 orientations (H, V, D1 and D2). Since dimension of each histogram is 40, the final feature used in following experiments is a 160-dimension vector.

B. Performance Under Varied Histogram Dimension

Another key factor is histogram dimension or the number of bins in each histogram. Feature's dimension is dependent on this parameter. Fig.2 shows that performance approaches convergence when bins number rises to 40. Based on this observation, we choose 40 as the histogram size for each distribution.

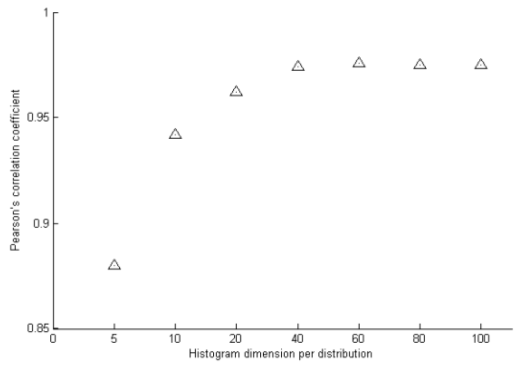

Fig. 2 LCC with different number of histogram bins

C. Performance Under Varied Training-set Size

At the beginning of section IV, we divide the LIVE dataset into the training subsets (80%) and test (20%) subsets for SVR as suggested by BRISQUE [3]. To validate this scheme under new feature space, we increase the percentage of training set from 10% to 90%. Small percentage means more challenges as the learned model sees less data. As shown in Fig. 3, we find that the performance keeps stable with the training set size decreasing from 90% to 20%, which indicates that the proposed approach has good generalization ability.

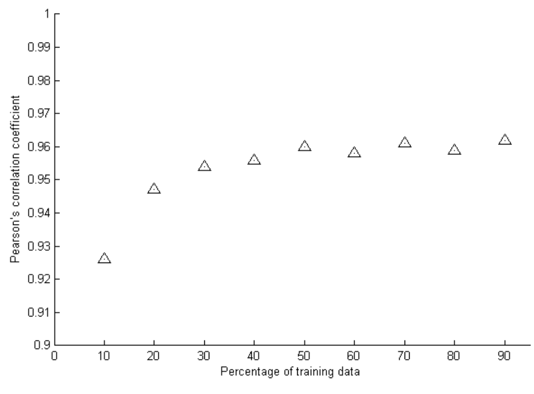

Fig. 3 LCC with different training set size

D. Overall Performance

We test the overall performance on LIVE IQA and part of Tid2008 database [10] (JPEG2000, JPEG, white noise and Gaussian blur). To make a fair comparison, this random train-test procedure is repeated 1000 times and the median performance is reported here. The final results are listed in Table II.

TABLE II
OVERALL PERFORMANCE (LCC, SROCC) OF PROPOSED APPROACH AND OTHER IMAGE QUALITY INDICES ON LIVE AND TID2008 DATABASE

Database	Index	PSNR	SSIM	BRISQUE	Proposed
LIVE IQA	LCC	0.803	0.892	0.945	**0.961**
	SROCC	0.812	0.890	0.949	**0.953**
TID2008	LCC	0.837	0.892	0.947	**0.954**
	SROCC	0.871	0.903	0.941	**0.949**

TABLE III
DETAILED PERFORMANCE (LCC, SROCC) OF THE PROPOSED METHOD AND BRISQUE ON LIVE DATABASE

metric	Index	JP2K	JPEG	WN	Blur	FF	ALL
BRISQUE	LCC	0.930	0.965	**0.974**	0.940	0.912	0.945
	SROCC	0.918	0.958	**0.971**	0.952	0.887	0.949
Proposed	LCC	**0.943**	0.967	0.964	**0.961**	0.910	**0.961**
	SROCC	**0.934**	0.951	0.957	**0.970**	0.877	**0.953**

Besides the benchmark BRISQUE - the best open NR-IQA algorithm (source code are from authors homepage: *http://live.ece.utexas.edu/research/quality/BRISQUE_release.zip*) to our knowledge, we also listed the performance of PSNR and SSIM, two well-known full reference IQA metrics for comparison. From the table it can be seen that the proposed method performs slightly better in correlation with human visual perception. Therefore, it indicates that the concatenated histogram feature has better generalization capability since more details of natural scene statistics are preserved.

Further detailed information on LIVE database for each category of distortions can be found in Table III, wherein our approach achieves superior performance for JP2K and BLUR distortions, and comparable performance for JPEG and FF distortions, but inferior performance for WN distortion. A small drop for white noise may come from imperfect model parameter configuration.

V. CONCLUSIONS

Rather than using parametric model features, a new feature based on a concatenated histogram of local MSCN coefficients is proposed for blind image quality assessment. It can be mapped to a more accurate image quality scores via a SVR regression prediction model.

ACKNOWLEDGMENT

This work was supported by 863 project (2012AA011703), NSFC (61221001, 61201384), the 111 Project (B07022) and the Shanghai Key Laboratory of Digital Media Processing and Transmissions.

REFERENCES

[1] A. K. Moorthy and A. C. Bovik, “Blind image quality assessment: From natural scene statistics to perceptual quality,” *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3350-3364, 2011.

[2] M. Saad, A. C. Bovik, and C. Charrier, “Blind image quality assessment: A natural scene statistics approach in the DCT domain,” *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3339-3352, 2012.

[3] A. Mittal, A. K. Moorthy and A. C. Bovik, “No-reference image quality assessment in the spatial domain,” *IEEE Transactions on Image Processing*, vol. 21, no.12, pp.4695-4708, 2012.

[4] H. Tang, N. Joshi, and A. Kapoor, “Learning a blind measure of perceptual image quality,” In *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*(*CVPR*), pp. 305-312, 2011.

[5] C. Chen, L. Song, X. Wang and M. Guo, No-reference video quality assessment on mobile devices, in *Proceeding of IEEE International Symposium on Broadband Multimedia Systems and Broadcasting* (*BMSB*), pp. 1-6, 2013.

[6] P. Ye, J. Kumar, L. Kang, and D. Doermann, “Real-time no-reference image quality assessment”, In *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*(*CVPR*), pp. 987-994, 2013.

[7] D. L. Ruderman, “The statistics of natural images,” *Network computation in neural systems*, vol. 5, no. 4, pp. 517-548, 1994.

[8] Siwei Lyu, “Dependency reduction with divisive normalization: Justification and effectiveness”, *Neural Computation*, vol.23, no. 11, pp.2942-2973, 2011.

[9] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, “A statistical evaluation of recent full reference image quality assessment algorithms,” *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440-3451, 2006.

[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 Radio electronics*, vol. 10, no. 10, pp. 30-45, 2009.