No-reference image quality assessment based on non-subsample shearlet transform and natural scene statistics*

WANG Guan-jun (王冠军)[1,2]**, **WU Zhi-yong** (吴志勇)[1], **YUN Hai-jiao** (云海姣)[1,2], and **CUI Ming** (崔明)[1,2]

1. Department of Photoelectric Measurement and Control, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China

2. University of Chinese Academy of Sciences, Beijing 100049, China

(Received 27 December 2015)

A novel no-reference (NR) image quality assessment (IQA) method is proposed for assessing image quality across multifarious distortion categories. The new method transforms distorted images into the shearlet domain using a non-subsample shearlet transform (NSST), and designs the image quality feature vector to describe images utilizing natural scenes statistical features: coefficient distribution, energy distribution and structural correlation (*SC*) across orientations and scales. The final image quality is achieved from distortion classification and regression models trained by a support vector machine (SVM). The experimental results on the LIVE2 IQA database indicate that the method can assess image quality effectively, and the extracted features are susceptive to the category and severity of distortion. Furthermore, our proposed method is database independent and has a higher correlation rate and lower root mean squared error (*RMSE*) with human perception than other high performance NR IQA methods.

Document code: A **Article ID:** 1673-1905(2016)02-0152-5

DOI 10.1007/s11801-016-5276-2

Objective no-reference (NR) image quality assessment (IQA) has drawn extensive attention for the major advantage of assessing image quality without any information of reference image. Nowadays, the NR IQA approaches can be classified into distortion-specific approaches[1-6] and general purpose approaches[7-12]. For distortion-specific IQA approaches, knowing the type of distortion in advance makes it easier to assess image quality, but also seriously limits the application. General purpose approaches extract statistical features to predict image quality across multifarious distortion categories. Natural scene statistics approaches are the representative for their excellent performance. However, the natural scene statistics methods generally pay more attention to the coefficients distribution and ignore the changes of energy and structure of coefficients across different orientations and scales. The performance of those methods remains to be improved.

We propose a new NR IQA method combined with non-subsample shearlet transform (NSST) and natural scene statistics features in this paper. Introducing NSST into IQA could effectively represent distorted image features.

For a distorted image, the high-frequency components are severely affected as distortion severity increases. To accentuate the changes in high-frequency components, the locally normalized luminance is performed on distorted images as follows.

$$\hat{I}(i,j)=\frac{I(i,j)-\mu(i,j)}{\sigma(i,j)+1}, \tag{1}$$

$$\mu(i,j)=\sum_{m=-M}^{M}\sum_{n=-N}^{N}w_{m,n}\times I_{m,n}(i,j), \tag{2}$$

$$\sigma(i,j)=\sqrt{\sum_{m=-M}^{M}\sum_{n=-N}^{N}w_{m,n}\times[I_{m,n}(i,j)-\mu(i,j)]^2}, \tag{3}$$

where $\hat{I}(i,j)$ is the locally normalized luminance, $i\in 1,2,\ldots H$ and $j\in 1,2,\ldots W$ are the pixel indices of image with the size of $H\times W$, μ and σ are the mean and local variance of the image, respectively, $w=\{w(m,n)|m=-M,\ldots M, n=-N,\ldots N\}$ is a two-dimensional Gaussian weighting function, and in this paper, $M=N=3$.

To extract image quality features, we transform the locally normalized images into shearlet domain using the NSST. The NSST is constructed by combining the non-sampled Laplacian pyramid transform (NSLP) with different combinations of shearing filters (SFs)[13]. NSLP

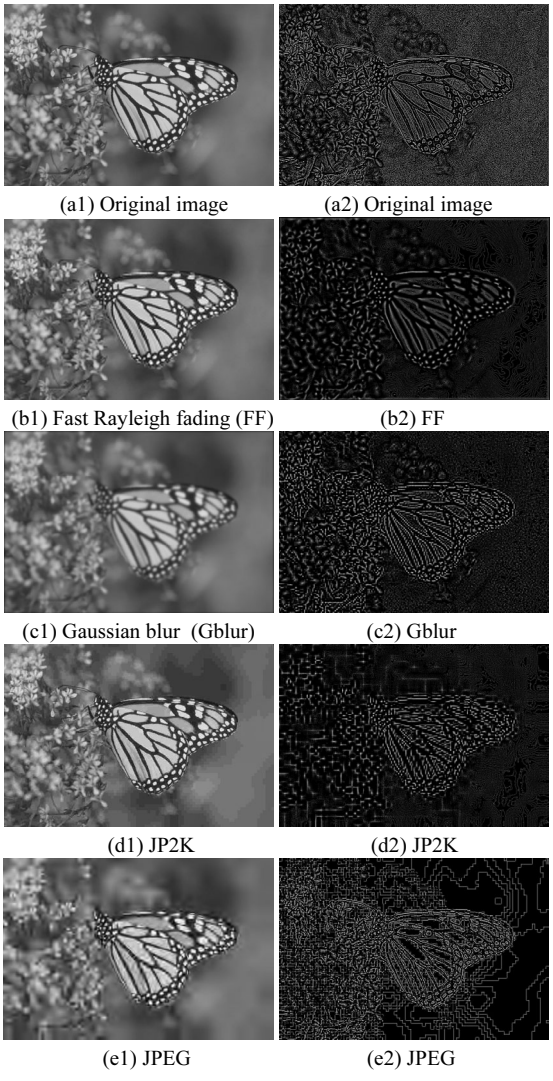
* This work has been supported by the National Natural Science Foundation of China (No.61405191), and the Jilin Province Science Foundation for Youths of China (No.20150520102JH).

** E-mail: wangguanjun198711@163.com

actualizes multi-scale decomposition to produce a high-frequency sub-image and a low-frequency sub-image at each level, and the NSLP is performed iteratively on the low-frequency component to capture the singular points of image. On the other hand, at each scale, 2^l directional sub-images are obtained through actualizing multi-directional decomposition with l stages using SFs on high-frequency sub-images. In this paper, the image is transformed into four scales and each scale owns eight orientation sub-images.

In order to demonstrate the effects of local normalization, Fig.1 gives an example of local normalization of a nature image. Furthermore, the shearlet coefficient distributions are plotted in Fig.2. To ensure a fair comparison, the magnitudes of coefficients and probabilities are normalized. From Fig.1 and Fig.2, it is easy to find that the local normalization of luminance will make the effects of distortions more obvious.

For NR IQA models, the reliability of the approach is as effective as the selection of features. Since different distortion categories significantly impact specific statistical regularities of natural scenes in different ways, we design an image quality feature vector utilizing



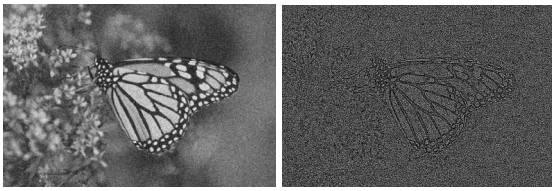
(a1) Original image (a2) Original image

(b1) Fast Rayleigh fading (FF) (b2) FF

(c1) Gaussian blur (Gblur) (c2) Gblur

(d1) JP2K (d2) JP2K

(e1) JPEG (e2) JPEG



(f1) White noise (WN) (f2) WN

Fig.1 The local normalization of luminance: (a1)—(f1) The non-locally normalized images; (a2)—(f2) The locally normalized luminance images

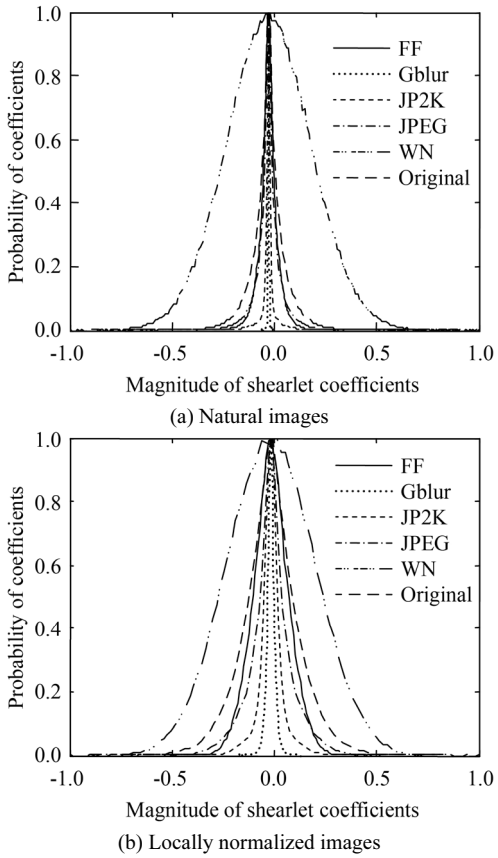


Fig.2 Distributions of shearlet coefficients

coefficient distribution, energy distribution and structural correlation (*SC*) across orientations and scales. Since the low-frequency components are less affected by distortions, we only extract features at fine scales, i.e., orientation index in the range of 17—32.

The distributions of shearlet coefficients are characterized by very large peaks, heavy tails and asymmetry. Therefore, we select an asymmetric generalized Gaussian distribution (AGGD) model to represent the coefficient distribution characteristics. The AGGD model is expressed as

$$f(x;\alpha,\beta_r,\beta_l,\mu)=\begin{cases}\dfrac{\alpha}{(\beta_l+\beta_r)\Gamma(1/\alpha)}\exp\left[-\left(\dfrac{x-\mu}{\beta_r}\right)^{\alpha}\right], & x\geq\mu\\ \dfrac{\alpha}{(\beta_l+\beta_r)\Gamma(1/\alpha)}\exp\left[-\left(\dfrac{-(x-\mu)}{\beta_r}\right)^{\alpha}\right], & x<\mu\end{cases}, \quad (4)$$

where α is the shape parameter, β_r and β_l represent the right and left scale parameters, respectively, μ is the mode of the AGGD model, and Γ is the gamma function.

In our method, the parameters (α, β_r, β_l, μ) at fine scales are used to design the features f_1—f_{64} and they will be estimated using the moment-matching based on the approach proposed in Ref.[14].

Since the cortical neurons of human are highly sensitive to energy in images, distortions will modify this energy distribution in unnatural ways[15]. To measure the energy distribution, we calculate the ratio of high frequency to low frequency components across different orientations and scales. The ratio R_k^l is calculated as

$$R_k^l = \frac{\sum\sum C_k^l(i,j)^2}{\sum\sum C_L(i,j)^2} , \tag{5}$$

where $C_k^l(i,j)$ is the coefficient of high-frequency sub-images with scale index k and orientation index l, and $C_L(i, j)$ is the low-frequency coefficient. The energy distribution at fine scales is used to design the features f_{65}—f_{80}.

The NSST is highly direction-sensitive and can produce a set of orientation sub-images. To describe the structural differences across different orientations, we calculate the *SC* of orientation sub-images. Since the entropy indicates the information amount of images, we choose the sub-image with the largest entropy at each scale as the cardinal orientation to calculate *SC*. The *SC* is computed as

$$SC = \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_1} , \tag{6}$$

where σ_x and σ_y are the standard variances of the cardinal and non-cardinal orientation sub-images, respectively, and σ_{xy} is the cross variance. C_1 is a constant to prevent the denominator to be zero, C_2 is also a constant, and C_2=2C_1. In our method, we extract the average of the *SC* at fine scales to construct the features f_{81}—f_{82}.

To clarify the relationship between these statistical features and human perception across the five distortion categories, we plot the Spearman's rank ordered correlation coefficient (*SROCC*) between these statistical features and the difference mean opinion score (*DMOS*) provided by LIVE2[16] database in Fig.3. From Fig.3, we observe that these statistical features have high correlation with human perception, so they are suitable to indicate image quality.

To express image quality intuitively, a regression model is needed to map feature vectors to image quality scores. However, it is not practical to build a general regression model for each distortion category. Thus, we implement a two-stage framework of distortion identification and distortion-specific quality assessment. First, we train a probabilistic classifier using an SVM. In this stage, the probability of distortion identification is estimated. Second, a regression model is built for each distortion category using support vector regression (SVR), and the distortion-specific quality score is mapped from the feature space. Finally, we calculate the image quality score as the sum of distortion-specific quality scores weighted by corresponding distortion probabilities.

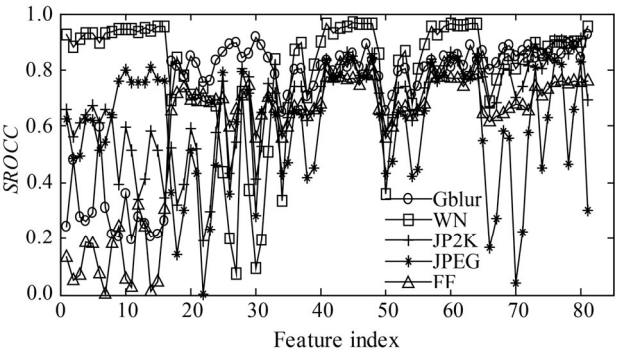


Fig.3 The *SROCC* between *DMOS* and statistical features on the LIVE2 database

In our method, the SVM and SVR are implemented utilizing the LIBSVM package[17] and the radial basis function kernel is selected for SVM and SVR.

In this paper, a series of experiments are conducted on LIVE2 IQA database to test the performance of our proposed method. The database is divided into training and testing sets randomly. Note that the two sets are non-overlapped. 80% of the original images and corresponding distorted images construct the training set which is used to train the distortion classification and regression models. The testing set consisting of the remained images is used to evaluate the performance of our IQA method. The splitting is randomly repeated for 1 000 times to avoid that those experiments are regulated by particular training-testing sets. The median performance indices across the 1 000 iterations are chosen as the final results. *SROCC*, Pearson linear correlation coefficient (*LCC*) and root mean squared error (*RMSE*) between predicted image quality scores and *DMOS* scores are calculated for evaluating the performance of IQA method.

We compare the performance of our proposed method with that of two FR methods (i.e., PSNR and SSIM[18]) and six NR IQA approaches (i.e., BIQI, DIIVINE, BLIINDS-II, BRISQUE, SHANIA and CurveletQA). The FR methods are only tested on the testing set for the reason that they need not training. Tabs.1—3 show the median experimental results and the superior results are listed in bold.

From Tabs.1—3, we can observe that our proposed method is correlated highly with human subjective opinions, and it performs better than the FR IQA methods as well as other general purpose IQA methods. Remarkably, for each distortion type as well as all the distortion types, our proposed NR method could be used to assess image quality effectively.

Tab.1 Median *SROCC* across 1 000 iterations

Model	JP2K	JPEG	WN	Gblur	FF	ALL
PSNR	0.864	0.879	0.941	0.783	0.878	0.869
SSIM	0.932	0.946	0.957	0.903	0.938	0.915
BIQI	0.842	0.792	0.954	0.915	0.775	0.798
DIIVINE	0.919	0.907	0.976	0.932	0.879	0.917
BLIINDS-II	0.931	0.928	0.947	0.904	0.883	0.917
BRISQUE	0.924	**0.967**	0.973	0.945	0.892	0.934
SHANIA	0.913	0.932	0.971	0.958	**0.941**	0.937
CurveletQA	0.937	0.909	0.979	0.961	0.896	0.930
Proposed	**0.948**	0.931	**0.986**	**0.973**	0.914	**0.945**

Tab.2 Median *LCC* across 1 000 iterations

Model	JP2K	JPEG	WN	Gblur	FF	ALL
PSNR	0.879	0.842	0.958	0.774	0.882	0.867
SSIM	0.937	0.941	0.974	0.907	**0.942**	0.904
BIQI	0.836	0.742	0.974	0.903	0.726	0.754
DIIVINE	0.921	0.925	0.981	0.924	0.899	0.915
BLIINDS-II	0.923	0.938	0.962	0.894	0.887	0.912
BRISQUE	0.928	**0.952**	0.983	0.951	0.903	0.940
SHANIA	0.876	0.897	0.958	0.964	0.907	0.911
CurveletQA	**0.946**	0.914	0.974	0.958	0.907	0.930
Proposed	0.942	0.924	**0.985**	**0.964**	0.923	**0.949**

Tab.3 Median *RMSE* across 1 000 iterations

Model	JP2K	JPEG	WN	Gblur	FF	ALL
PSNR	7.842	8.436	5.271	9.473	7.426	8.271
SSIM	**5.695**	**6.115**	4.729	4.768	**4.475**	**5.742**
BIQI	13.814	17.083	5.415	9.892	15.482	15.915
DIIVINE	8.570	10.217	5.914	8.594	9.671	10.138
BLIINDS-II	8.158	7.825	6.458	8.413	9.483	9.153
BRISQUE	8.058	9.453	3.502	7.871	9.842	9.043
SHANIA	8.914	10.110	6.518	7.612	10.043	8.716
CurveletQA	7.585	9.078	3.718	5.472	8.628	8.476
Proposed	6.846	8.713	**3.471**	**3.783**	7.320	7.657

In order to demonstrate the database independence of our approach, we train the method on LIVE IQA database and test it on TID2008 database[19]. The *SROCC* results are listed in Tab.4. The results show that our proposed method is more database independent than the high performance NR IQA methods of CurveletQA, SHANIA and BRISQUE.

Tab.4 *SROCC* obtained by training on the LIVE2 database and testing on the TID2008 database

Model	JP2K	JPEG	WN	Gblur	ALL
BRISQUE	0.768	**0.872**	0.813	0.863	0.852
CurveletQA	0.569	0.864	0.848	0.852	0.867
SHANIA	0.843	0.852	0.803	0.833	0.849
Proposed	**0.866**	0.852	**0.896**	**0.878**	**0.874**

Furthermore, we analyze the capability of identifying distortion types. The median classification accuracies for all distortion types across 1 000 iterations are shown in Tab.5. We note that our proposed method could effectively identify distortion category, especially for WN distortion. Fig.4 shows the confusion matrix for each distortion type, where the horizontal and vertical axes represent the true distortion type and predicted distortion type, respectively. The numerical values represent the mean confused probabilities across 1 000 iterations. However, different distortion categories sometimes present similar effects, so the actual classification accuracy will be impacted by the performance of IQA algorithm inevitably[9].

Tab.5 Median accuracy across 1 000 iterations

Model	Accuracy	Model	Accuracy
JP2K	73.36%	Gblur	78.15%
JPEG	76.12%	FF	57.18%
WN	99.48%	ALL	76.29%

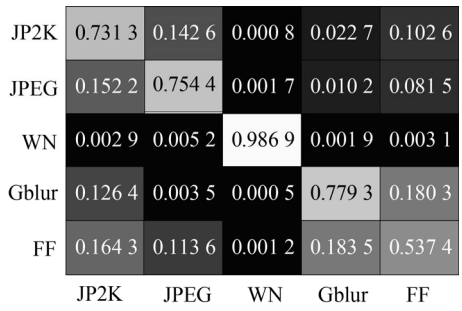


Fig.4 Mean confusion matrix across 1 000 iterations

In this paper, we present a novel NR IQA method to assess image quality across multifarious distortion categories. The method introduces locally normalized luminance and NSST into image quality assessment and extracts a series of natural scenes statistical features for training classification and regression models. We demonstrate that the method can effectively assess image quality and performs better than the most popular FR IQA method of SSIM and other high performance NR IQA methods, including BIQI, DIIVINE, BLIINDS-II, BRISQUE, SHANIA and CurveletQA. In addition, we develop this work from a set of statistical features of nature scenes rather than specific distortion features, making our method easily extended beyond the set of distortions considered in this paper.

References

[1] H. R. Sheikh, A. C. Bovik and L. Cormack, IEEE Transactions on Image Processing **14**, 1918 (2005).

[2] T. Brandao and M. P. Queluz, Signal Processing **88**, 822 (2008).

[3] A. Mittal, G. S. Muralidhar and A. C. Bovik, IEEE Signal Processing Letters **19**, 75 (2012).
[4] Q. B. Sang, D. L. Liang, X. J. Wu and C. F. Li, Journal of Optoelectronics·Laser **25**, 595 (2014). (in Chinese)
[5] L. Y. Zhou and Z. B. Zhang, Optik **125**, 5677 (2014).
[6] N. D. Narvekar and L. J. Karam, IEEE Transactions on Image Processing **20**, 2678 (2011).
[7] Y. Li, L. M. Po, X. Xu and L. Feng, Signal Processing Image Communication **29**, 748 (2014).
[8] A. K. Moorthy and A. C. Bovik, IEEE Signal Processing Letters **17**, 513 (2010).
[9] A. K. Moorthy and A. C. Bovik, IEEE Transactions on Image Processing **20**, 3350 (2011).
[10] M. A. Saad, A. C. Bovik and C. Charrier, IEEE Transactions on Image Processing **21**, 3339 (2012).
[11] A. Mittal, A. K. Moorthy and A. C. Bovik, IEEE Transactions on Image Processing **21**, 4695 (2012).
[12] Y. Zhang and D. M. Chandler, Journal of Electronic Imaging **22**, 043025 (2013).
[13] H. Y. Yang, X. Y. Wang, P. P. Niu and Y. C. Liu, Neural Networks **57**, 152 (2014).
[14] N. E. Lasmar, Y. Stitou and Y. Berthoumieu, Multiscale Skewed Heavy Tailed Model for Texture Analysis, Proceedings of the 16th IEEE International Conference on Image Processing, 2281 (2009).
[15] L. X. Liu, H. P. Dong, H. Huang and A. C. Bovik, Signal Processing Image Communication **29**, 494 (2014).
[16] H. R. Sheikh, Z. Wang, L. Cormack and A. C. Bovik, LIVE Image Quality Assessment Database, http://live.ece.utexas.edu/research/ quality.
[17] C. C. Chang and C. J. Lin, Acm. Transactions on Intelligent Systems & Technology **2**, 389 (2011).
[18] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, IEEE Transactions on Image Processing **13**, 600 (2004).
[19] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli and F. Battisti, Advances of Modern Radioelectronics **10**, 30 (2009).