Journal of Network Intelligence ©2019 ISSN 2414-8105 (Online)
Taiwan Ubiquitous Information Volume **4**, Number **1**, February **2019**

Perceptual Evaluation of Scanning Electron Microscopy Imaging

Qiaoyue Li, Zhaolin Lu*, Leida Li

School of Information and Control Engineering
China University of Mining and Technology
Xuzhou, 221116, China
*Corresponding author: zhaolinlu@cumt.edu.cn

Received January, 2018; Revised November, 2018

Abstract. *In recent years, scanning electron microscopy (SEM) has aroused widespread concern. It broadens the human vision, and provides a way to obtain the microstructure information. Image sharpness is the most important technical indicator of SEM in the imaging process. It usually takes a lot of time and effort to obtain a clear image by repeatedly adjusting imaging parameters and settings. Therefore, it is necessary to develop an effective method to evaluate the quality of SEM imaging. Image database plays an important role in the research of image quality evaluation. In order to explore how human eyes perceive different micro blurred images, it is necessary to have a SEM database. In this paper, we create a database composed of 650 SEM images in the beginning, and then conduct user study. Finally, 12 blurred image quality metrics and 8 general-purpose no-reference image quality metrics are tested by using the 150 blurred images in the microscopic image database. The present study shows that the existing quality metrics are limited in predicting the quality of SEM images.*
Keywords: Scanning electron microscopy, Image quality assessment, Database, Blurred image

1. **Introduction.** Scanning electron microscopy (SEM) is a widely used electronic instrument, which uses the electrical system to enlarge the observed object and the microstructure of the display, and to study the characteristics of the object from the microscopic morphology. At present, it has been widely used in biology, pathology, cell histology, and many other fields. The SEM image is slightly different from the macroscopic image, mainly in: the scene in the SEM image is usually simple, the relationship between the points in the image, the relationship between the line and the surface is far from the complexity of the macroscopic image. The SEM image is affected by light, the contrast of the image is low, the feature recognition is difficult, and it is easy to produce a large number of unrecognized pixels. The stain attached to the microscopic object is enlarged by the microscope and becomes very obvious in the microscope.

Blur is one of the most common distortion types in SEM imaging. It usually takes a lot of time and effort to obtain a clear image by repeatedly adjusting imaging parameters and settings. Surprisingly, in the literature of image quality assessment (IQA), little work has been devoted to the evaluation of SEM images. The existing methods for objective image quality assessment based on the information of the reference image can be classified into Full Reference (FR) IQA [1], Reduced Reference (RR) IQA [2] and No Reference (NR) IQA [3, 4, 5, 6]. In practice, researchers often resort to common IQA measures, such as peak signal to noise ratio (PSNR) and structural similarity (SSIM) index to evaluate

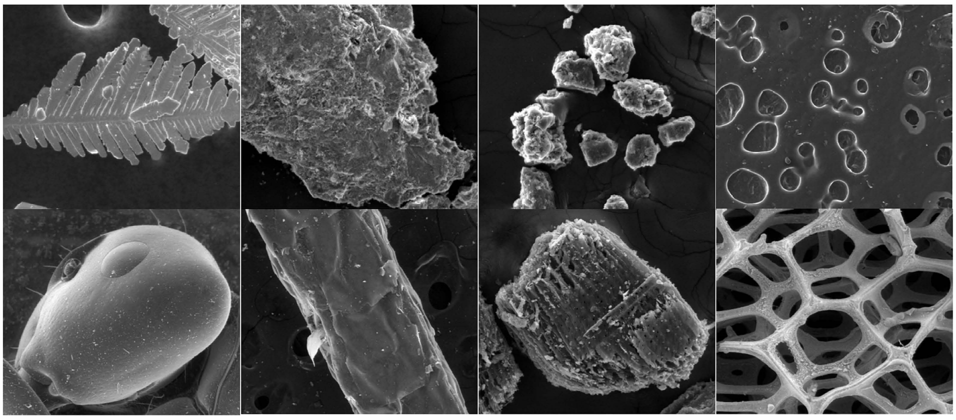


Figure 1. Examples of SEM images

image quality, but the validation of these measures is still lacking. The purpose of this paper is first to establish a SEM image database, which consists of 650 images of different types of distortion image. Then subjective experiment is conducted to collect the human scores. Finally, 150 blurred images are employed to evaluate the performances of 12 image sharpness/blur metrics and 8 general-purpose no-reference image quality metrics.

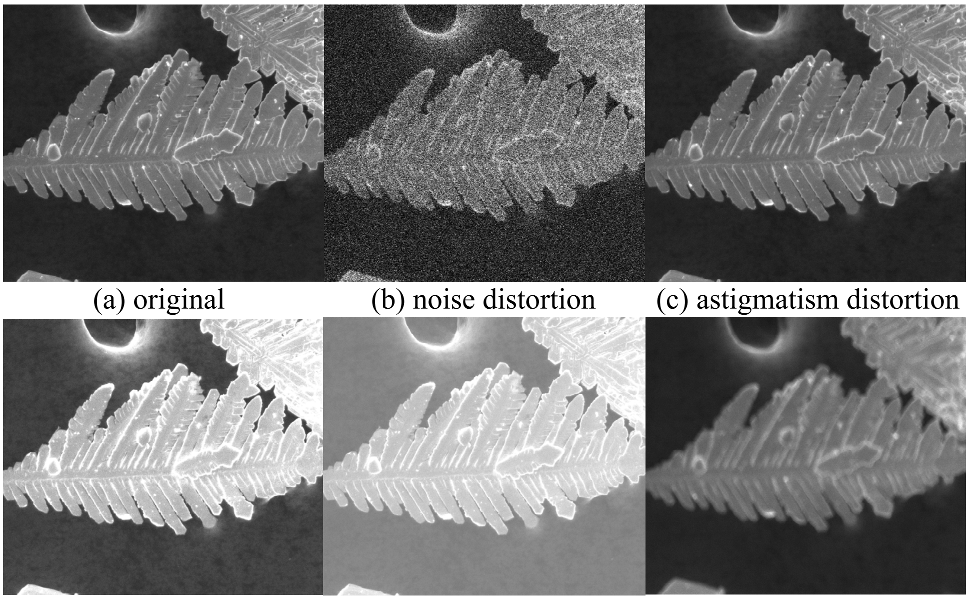
2. **SEM Image Database.** The IQA community has released a number of image quality databases. These databases play an important role in understanding how people perceive the quality of images, and they are crucial for designing and evaluating IQA models in terms of the consistency with human subjective evaluations. To the authors' best knowledge, such a database for SEM images is still lacking. So in this work, we first build a SEM database followed by user study.

The SEM database covers a variety of contents, including ants, metal, stamens, colloids, etc. There are 50 different images in total. Fig. 1 shows some of the sample images. A group of 13 pictures, the size of the picture is 1024*884, a total of 650 pictures. Each group of pictures, by adjusting the microscope parameters, can be taken out different distortion images, including noise, blur, contrast is too strong or weak, each group consists of an original image, an example of different distortions image in Fig. 2. For each image, 5 kinds of distortions are simulated, including noise distortion, astigmatism distortion, contrast distortion, brightness distortion and blur distortion. Table 1 describes each type of distortion.

Table 1. SEM database

Datebase	**Category**	**Size**	**Rank**
SEM	**Noise**	3	Weak, medium, strong
	Blur	3	Weak, medium, strong
	Contrast	2	Weak, strong
	Brightness	2	Weak, strong
	Astigmatism	2	X-axis, Y-axis

2.1. **Subjective Experiment.** Subjective experiment is performed on an Intel (R) Pentium (R) CPU G645 2.9GHz PC. Images are shown using a LCD display with the resolution of 1920 * 1080. The experiment was conducted in a normal light indoor workplace.



(a) original (b) noise distortion (c) astigmatism distortion
(d) contrast distortion (e) brightness distortion (f) blur distortion

FIGURE 2. An example of different distortions image

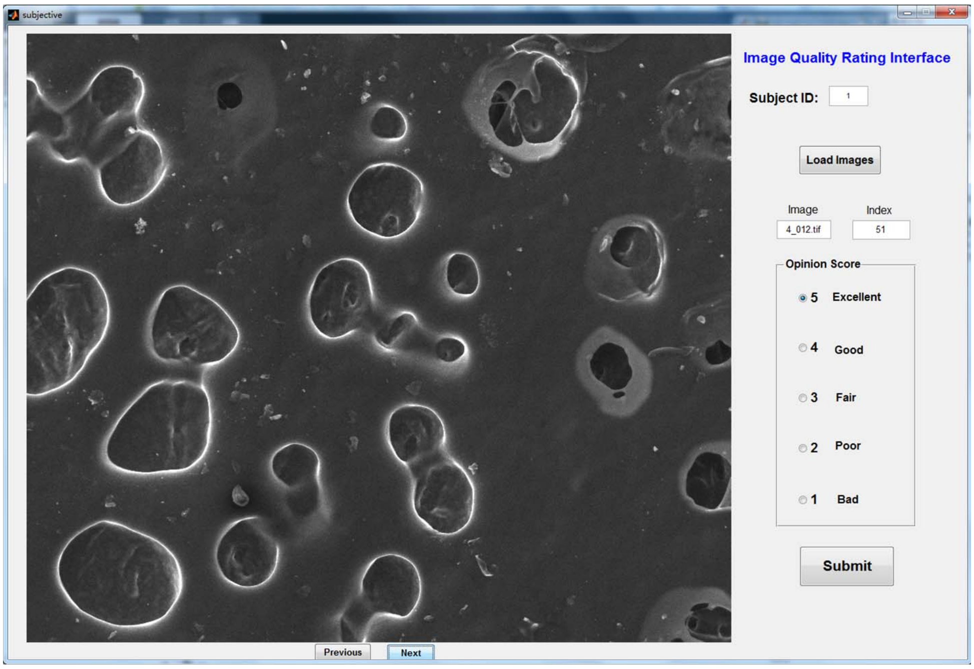


FIGURE 3. Illustration of the interactive system used in our subjective viewing experiments

Thirty subjects without explicit image processing background participated in the subjective experiment. We designed a MATLAB graphical user interface (GUI) to carry out the subjective test, as shown in Fig. 3. Before the experiment begins, the subjects are told how to use the interactive platform and how the quality of the image is divided. Subjects were required to consistently provide a holistic view of their quality in the range of 1 to 5. The subjective score can be automatically saved in a sheet after pressing the Submit button.

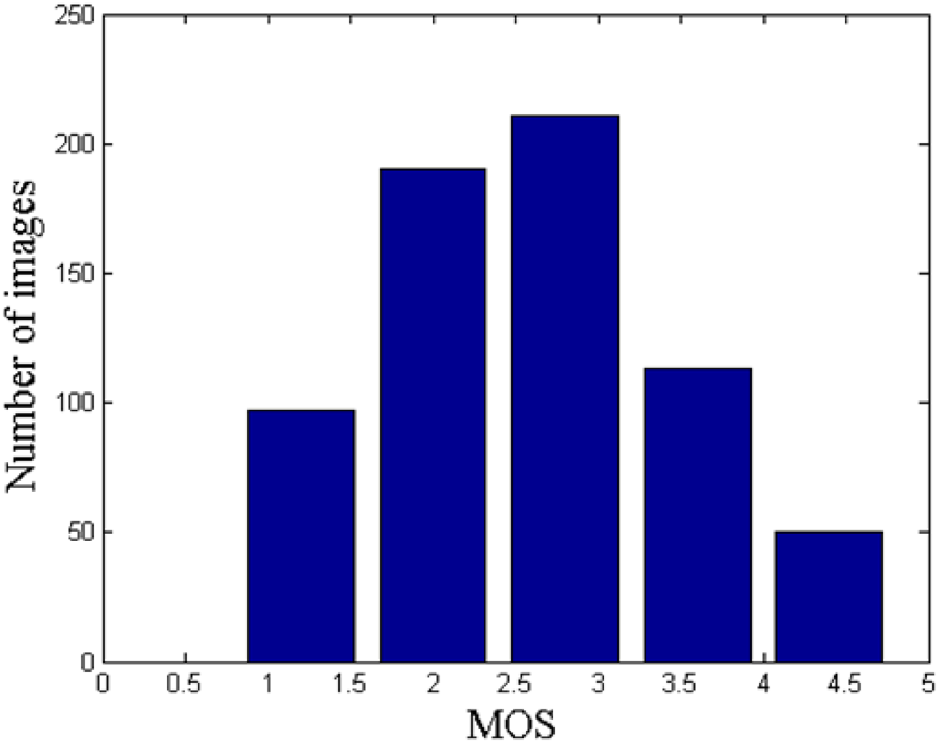


FIGURE 4. Histogram distribution of the MOS values

2.2. **Processing and Analysis of Subjective Scores.** In order to reduce the error of experimental results, five outliers were removed from 30 subjective evaluation data of each image, leaving only 25 valid test results. The mean opinion score (MOS) was calculated as the subjective quality. The histogram distribution of the MOS values is shown in Fig. 4. It can be found from the figure that the MOS values cover the whole range. This means that database contains SEM images with different distortion levels.

3. **Performance Evaluation of Image Sharpness/Blur Metrics.** In recent years, a number of blurred image evaluation algorithms have been proposed. In [7], Marziliano *et al.* firstly proposed to use Sobel operator for detecting image edge, and then edge width was computed as the image sharpness score. In [8], Ferzli *et al.* proposed the Just Noticeable Blur (JNB) algorithm, which combines the concept of JNB with the probability sum model to predict the relative amount of the image of different content. Niranjan *et al.* [9] studied the JNB algorithm by Calculating the blurred Probability Blur Detecting (CPBD) algorithm based on the blurred ambiguity of different contrast values. In [10], Vu *et al.* proposed Spectral and Spatial Sharpness (S3) algorithm, and evaluated the image clarity by using local amplitude spectrum slope and total variational combination. In [11], Vu *et al.* proposed the use of Discrete Wavelet Transform (DWT) to decompose the image, and calculate the weighted average of the wavelet coefficients as the sharpness score. In [12], Hassen *et al.* proposed that image sharpness can be identified by strong Local Phase Coherence (LPC). In [13], Sang *et al.* proposed a no-reference image blur metric based on Singular Value Curve (SVC). Bahrami *et al.* [14] defines the Maximum Local Variation (MLV) for each pixel as the maximum intensity difference between the pixel and its 8-neighborhood, and the standard deviation of the MLV distribution for each pixel is metric of sharpness. In [15], Gu *et al.* proposed an image sharpness algorithm based on Auto Regressive (AR) model parameter analysis. In [16], Li *et al.* proposed the use of discrete Tchebichef moments to measure the change of image shape caused by

image blur distortion. In [17], Li *et al.* proposed a no-reference blurred quality model based on Sparse Representation (SR), and proposed to eliminate the influence of image content by variance normalized sparse coefficient energy. In [18], Li *et al.* proposed a no-reference and robust image sharpness evaluation (RISE) method by learning multiscale features extracted in both the spatial and spectral domains.

To evaluate the rationality of using the existing sharpness/blur metrics for SEM images, we test their performances using 150 blurred SEM images. Three common criteria are used to evaluate the performance of image quality metrics. Specifically, Pearson linear correlation coefficient (PLCC) and root mean square error (RMSE) are used to describe the prediction accuracy. Spearmans rand-order correlation coefficient (SRCC) measures prediction monotonicity.

In order to calculate these indexes, the nonlinear regression equation is used to logically map the predicted scores and subjective scores, so that they are linearly related in the same scale space. The nonlinear regression used is as follows [19]:

$$f\left(x\right) = \tau_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\tau_2 (x - \tau_3)}} \right) + \tau_4 x + \tau_5, \tag{1}$$

where τ_i, i=1,2,3,4,5 are the parameters to be fitted.

The experimental results are summarized in Table 2, where the best results are marked in boldface. The experimental results show that even the best performing metric RISE can only achieve moderate correlation with subjective scores. Therefore, there is still much room for the design of SEM image sharpness metrics.

TABLE 2. Performances of blurred image algorithms on database

Algorithm	**PLCC**	**SROCC**	**RMSE**
RISE[18]	0.8415	0.8190	0.3705
LPC[12]	0.7531	0.6828	0.4552
MLV[14]	0.7155	0.6931	0.4833
BIBLE[16]	0.4244	0.4077	0.6264
S3[10]	0.3919	0.3610	0.6365
JNB[8]	0.3210	0.2869	0.6552
SPARISH[17]	0.2942	0.2513	0.6612
SVC[13]	0.2919	0.2640	0.6617
PBRM[7]	0.2680	0.2446	0.6665
FISH[11]	0.2435	0.2194	0.6710
CPBD[9]	0.1399	0.6850	0.4466
ARISM[15]	0.0960	0.0756	0.6886

4. **Evaluation of Existing General-Purpose No-Reference IQA Metrics.** We have also tested the performances of the existing general-purpose NR image quality metrcis, which are designed for the quality assessment of distortions without any prior information. The tested NR image quality metrics include Blind/Referenceless Image Spatial QUality Evaluator (BRISQUE) [20], Natural Image Quality Evaluator (NIQE) [21], Blind Image Quality Index (BIQI) [22], Blind Image Integrity Notator using DCT Statistics (BLIINDS-II) [23], Derivative Statistics-based Image QUality Evaluator (DESIQUE) [24], Distortion Identification-based Image Verity and Integrity Evaluation (DIIVINE) Index [25], Quality Aware Clustering (QAC) model [26] and SpatialSpectral Entropy-based

Quality (SSEQ) index [27]. The experimental results are summarized in Table 3, where the best results are marked in boldface.

TABLE 3. Performances of blurred image algorithms on database

Algorithm	**PLCC**	**SROCC**	**RMSE**
BRISQUE[20]	0.8276	0.8102	0.3812
DESIQUE[24]	0.8247	0.8059	0.3843
BIQI[22]	0.8127	0.7867	0.3983
DIIVINE[25]	0.7766	0.7461	0.4276
BLLINDS-II[23]	0.7698	0.7442	0.4320
NIQE[21]	0.7391	0.7311	0.4660
SSEQ[27]	0.6398	0.6323	0.5141
QAC[26]	0.2513	0.1616	0.6696

It is observed from the table that the general-purpose NR quality metrics generally perform better than the aforementioned image sharpness/blur metrics. It should be noted that most of the general-purpose NR quality metrics are learning-based. Training and test in such a small-scale image database is very likely to cause over-fitting. From this perspective, the performances of the above NR metrics are still not that ideal. This also indicates that there is still much room for the design of quality metrics specifically for SEM images.

5. **Conclusion.** In this paper, a SEM image database is first established, and then the subjective experiment and data analysis are carried out. The obtained MOS values were used to evaluate 12 blurred image quality metrics and 8 general-purpose no-reference image quality metrics to check their suitability in evaluating SEM images. The experimental results show that the existing IQA metrics are rather limited for SEM image quality evaluation. This indicates that quality models specifically designed for microscopic images are highly needed, which will be our future work.

Acknowledgment. This work was supported in part by the National Natural Science Foundation of China under Grants 61771473, 61379143 and 51604271, in part by the Six Talent Peaks High-level Talents in Jiangsu Province under Grant XYDXX-063, and in part by the Qing Lan Project.

REFERENCES

[1] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, 2004.

[2] J. J. Wu, W. S. Lin, G. M. Shi, L. D. Li, and Y. M. Fang, "Orientation selectivity based visual pattern for reduced-reference image quality assessment," *Information Sciences*, vol. 351, pp. 18-29, 2016.

[3] L. D. Li, H. C. Zhu, G. B. Yang, and J. S. Qian "Referenceless Measure of Blocking Artifacts by Tchebichef Kernel Analysis," *IEEE Signal Processing Letters*, vol. 21, no. 1, pp. 122-125, 2014.

[4] L. D. Li, Y. Zhou, W. S. Lin, J. J. Wu, X. F. Zhang, and B. J. Chen, "No-reference quality assessment of deblocked images," *Neurocomputing*, vol. 177, pp. 572-584, 2016.

[5] L. D. Li, W. S. Lin, and H. C. Zhu, "Learning Structural Regularity for Evaluating Blocking Artifacts in JPEG Images," *IEEE Signal Processing Letters*, vol. 21, no. 8, pp. 918-922, 2014.

[6] T. Yin, G. B. Yang, L. D. Li, and D. Y. Zhang, "Detecting seam carving based image resizing using local binary patterns," *Computers & Security*, vol. 55, pp. 130-141, 2015.

[7] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric," *International Conference on Image Processing*, vol. 3, no. 4, pp. III-57 - III-60, 2002.

[8] R. Ferzli, and L. J. Karam, "A No-Reference Objective Image Sharpness Metric Based on the Notion of Just Noticeable Blur (JNB)," *IEEE Transactions on Image Processing*, vol. 18, no. 4, pp. 717-28, 2009.

[9] N. D. Narvekar, and L. J. Karam, "A No-Reference Image Blur Metric Based on the Cumulative Probability of Blur Detection (CPBD)," *IEEE Trans Image Process*, vol. 20, no. 9, pp. 2678-83, 2011.

[10] C. T. Vu, and D. M. Chandler, "S3: A Spectral and Spatial Sharpness Measure," *International Conference on Advances in Multimedia*, pp. 37-43, 2009.

[11] P. V. Vu, and D. M. Chandler, "A Fast Wavelet-Based Algorithm for Global and Local Image Sharpness Estimation," *IEEE Signal Processing Letters*, vol. 19, no. 7, pp. 423-426, 2012.

[12] R. Hassen, Z. Wang, and M. M. A. Salama, "Image Sharpness Assessment Based on Local Phase Coherence," *IEEE Transactions on Image Processing*, vol. 22, no. 7, pp. 2798 - 2810, 2013.

[13] Q. B. Sang, H. X. Qi, X. J. Wu, C. F. Li, and A. C. Bovik, "No-reference image blur index based on singular value curve," *Journal of Visual Communication* & *Image Representation*, vol. 25, no. 7, pp. 1625-1630, 2014.

[14] K. Bahrami, and A. C. Kot, "A Fast Approach for No-Reference Image Sharpness Assessment Based on Maximum Local Variation," *IEEE Signal Processing Letters*, vol. 21, no. 6, pp. 751-755, 2014.

[15] K. Gu, G. T. Zhai, W. S. Lin, X. K. Yang, and W. J. Zhang, "No-Reference Image Sharpness Assessment in Autoregressive Parameter Space," *IEEE Transactions on Image Processing*, vol. 24, no. 10, pp. 3218 - 3231, 2015.

[16] L. D. Li, W. S. Lin, X. S. Wang, G. B. Yang, K. Bahrami, and A. C. Kot, "No-Reference Image Blur Assessment Based on Discrete Orthogonal Moments," *IEEE Transactions on Cybernetics*, vol. 46, no. 1, pp. 39 - 50, 2016.

[17] L. D. Li, D. Wu, J. J. Wu, H. L. Li, W. S. Lin, and A. C. Kot, "Image Sharpness Assessment by Sparse Representation," *IEEE Transactions on Multimedia*, vol. 18, no. 6, pp.1085-1097, 2016.

[18] L. D. Li, W. H. Xia, W. S. Lin, Y. M. Fang, and S. Q. Wang, "No-Reference and Robust Image Sharpness Evaluation Based on Multiscale Spatial and Spectral Features," *IEEE Transactions on Multimedia*, vol. 19, no. 5, pp. 1030-1040, 2017.

[19] 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.

[20] 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.

[21] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a Completely Blind Image Quality Analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209-212, 2013.

[22] A. K. Moorthy, and A. C. Bovik, "A Two-Step Framework for Constructing Blind Image Quality Indices," *IEEE Signal Processing Letters*, vol. 17, no. 5, pp. 513-516, 2010.

[23] M. A. 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.

[24] Y. Zhang, and D. M. Chandler, "An algorithm for no-reference image quality assessment based on log-derivative statistics of natural scenes," *Image Quality and System Performance X*, pp. 1-11, 2013.

[25] 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.

[26] W. F. Xue, L. Zhang, and X. Q. Mou, "Learning without Human Scores for Blind Image Quality Assessment," *Computer Vision and Pattern Recognition*, pp. 995-1002., 2013.

[27] L. X. Liu, B. Liu, H. Huang, and A. C. Bovik, "No-reference image quality assessment based on spatial and spectral entropies," *Signal Processing Image Communication*, vol. 29, no. 8, pp. 856-863, 2014.