# Learning to integrate local and global features for a blind image quality measure

*Min Liu, Guangtao Zhai, Ke Gu, and Xiaokang Yang*
Insti. of Image Commu. & Infor. Proce., Shanghai Jiao Tong Univ., Shanghai, China, 200240
Shanghai Key Laboratory of Digital Media Processing and Transmissions
E-mail: {liumin_merry/zhaiguangtao/gukesjtuee/yangxiaokang}@sjtu.edu.cn

***Abstract*—In this paper, we present a new algorithm for blind/no-reference image quality assessment (BIQA/NR-IQA). Most existing measures are "opinion-aware", demanding human opinion scored images to map image features to them. The task of obtaining human scores of images is, however, commonly thought to be uneconomical, and thus we focus on "opinion free" (OF) quality metrics in this research. By integrating local and global features, this paper develops a learning-based BIQA approach with three steps by combining local and global features together. In the first step of extracting local features, we use the quality aware clustering with the centroid of each quality level trained by K-means, while we in the second step compute the global features based on the natural scene statistics. Finally, the third step uses the SVR to train a regression module from the above-mentioned local and global features to derive the overall image quality score. Experimental results on LIVE, TID2008, CSIQ, and TID2013 databases validate the effectiveness of our proposed metric (a general framework) as compared to popular no-, reduced- and full-reference IQA approaches.**

***Keywords*—Image quality assessment (IQA), local features, K-means, global features, support vector regression (SVR)***

## I. Introduction

Many efforts have been devoted to the study of image quality assessment for the ubiquitous use of digital images and networked handheld devices since last decade. It has merits of promoting qualities of digital images, which usually deteriorated by distortions like commonly encountered JPEG, JPEG2000, Gaussian blur, white noise, and etc. With the launch of high definition television (HDTV), Internet protocol IV (IPTV), and rapid development of internet service (Facebook, Google, Flicker, etc), the demanding of end user for more satisfactory quality of experience (QoE) increases. Better visual experience should be delivered to consumers by using objective measures of image quality.

Mentioning objective image quality assessment (IQA), we first come to full-reference (FR) IQA, which requires not only the distorted image, but also the 'clean' one. The distorted image is assessed with respect to the distance between itself and the associated 'clean' image. But the pristine images are usually not accessible in reality, so no-reference (NR) IQA is born, working under the situation that only information needed is from distorted images. As a tradeoff between FR-IQA and NR-IQA, there is reduced-reference (RR) IQA existing between them, with partial information of the reference image is required. A state-of-the-art RR-IQA algorithm maintains a good balance between the amount of RR features and the effectiveness of image quality prediction.

NR-IQA is believed to be the most tough category in object quality assessment. Current blind image quality assessment (BIQA) has two sorts: distortion-specific methods and distortion-independent (general-purpose) methods. In distortion-specific methods, Wang *et al.*, Gastaldo and Zunino, Brandão *et al.* proposed distortion specific BIQA algorithms for JPEG compression artifacts in [1], [2] and [3], by extracting features like blockiness and DCT coefficient statistics. Marziliano *et al.* and Sheikh *et al.* solved JPEG2000 compressed image quality by using blur, ringing, and wavelet coefficient statistics, as referred in [4] and [5]. In [6], the authors assessed contrast-changed images taking into account the information residual between the input and distorted images as well as the first four order statistics of the distorted image histogram. In [7], the authors blindly predicted the visual quality of tone-mapped images based on the concept of details preservation. In [8], the authors estimated the noise level and assessed noised images without the help of reference images. In [9], the authors resorted to the free energy based brain theory to measure the multiple distortions.

On the other hand, distortion-independent methods are more general for the application that the artifact information is unknown. The works in [10]-[18] are devoted to this problem successively. Sadd *et al.* proposed BLIINDS in [11], by claiming that image distortions will change the regularity of extracted features in the DCT domain. In [12], Mittal *et al.* released BRISQUE on the basis of the highly correlation of different features in the domain with human perception. Apart from the natural scene statistics (NSS) based IQA models, there is another group of training based methods. In [13], the authors integrated two RR-IQA models [19] and [20] to derive the effective referenceless quality measure. Ye *et al.* in [18] presented visual codebooks based on Gabor features extracted from local image patches.

The above-mentioned IQA algorithms all require human subjective quality scores, i.e. "opinion-aware" (OA) BIQA models. However, achieving MOS values are usually too time-consuming and expensive. Consequently, "opinion-free" (OF) BIQA which do not require human opinion scores for training are desirable. In [21], Mittal presented a notion that certain latent characteristics differ between original images. NIQE in [22] is a perceptual model estimating distance of multivariate Gaussian fit of distorted images between the original ones. Xue *et al.* proposed in [23] a series of quality-aware centroids of each quality level as codebooks to infer image quality of each patch. In this paper, we further tackle this problem by proposing a three-step learning based BIQA (TSBIQA)

Fig. 1. Ten train images from Berkeley Segmentation database for training.

algorithm by combining local and global features, describing local image neighborhoods and the entire image respectively.

Previous works usually treating local and global features individually, we therefore in this paper make an innovation to combine the bottom-up and top-down statistics together. To specify, our approach consists of the following three components. (1) local feature extraction: according to Xue *et al.*'s work above, we achieve local quality of each image patch by referring to a class of quality-aware clustering (QAC) centroids. (2) global feature extraction: shape and variance of mean subtracted contrast normalized (MSCN) coefficient distribution which varies with different distortions. (3) support vector regression (SVR) based regression model. It needs to highlight here that our work is a general model, which can be adopted to various local and global features.

The remainder of this paper is organized as follows. Section II illustrates the details of the proposed method. Experimental implementations and results are described in Section III. Finally, Section IV concludes the paper .

## II. OUR PROPOSED METHOD

Human intuition of images is the synthesis of local and global perception. The average distortion influences the whole human perception to some degree, while a terrible local part will lower the overall image quality severely. The theory has also been adopted in some IQA and saliency detection literatures [24]-[25]. This part describes the details of our work, including K-means based local feature extraction, and NSS-based global feature extraction, along with Minkowski pooling strategy and SVR regression.

### A. Local feature extraction

Previous BIQA works rarely throughly consider image local features, we therefore in this paper extract image local information by estimating quality of overlapped image patches, referring to the work of Xue *et al.* in [23], for its simpleness and convenience. Out of the restriction of "opinion-free", we randomly select ten images from Berkeley Segmentation Database (can be observed in Fig. 1), and generate four kinds of distortions (JPEG, JPEG2000, white noise, and Gaussian blur), and each at ten distortion levels for every single raw image.

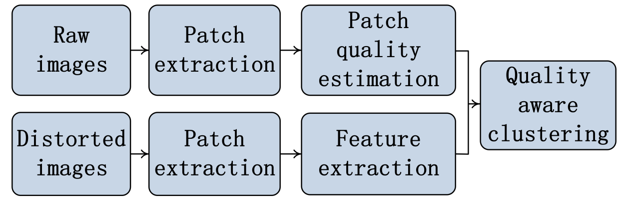


Fig. 2. The block of quality aware clustering procedure.

Given an image $\mathbf{x}$ and its corresponding distorted image $\mathbf{y}$, local patches $\mathbf{x}_i$ and $\mathbf{y}_i$ are generated by partitioning them into overlapped patches. To specify local quality of $\mathbf{y}_i$, we use FSIM [26] in this paper for its effectiveness and efficiency. The quality of the $i$-th patch is obtained as $s_i$.

$$\begin{aligned} s_i(x_i, y_i) &= FSIM(x_i, y_i) \\ &= \frac{2PC(x_i)PC(y_i) + T_1}{PC(x_i)^2 + PC(y_i)^2 + T_1} \times \frac{2G(x_i)G(y_i) + T_2}{G(x_i)^2 + G(y_i)^2 + T_2} \end{aligned} \tag{1}$$

where $T_1$ and $T_2$ are positive constants for numerical stability, and PC($x_i$) and G($x_i$) are the phase congruency and gradient magnitude of patch $x_i$. Further, the patches with worse quality have been found to have higher correlation with human perception, which has been proved in [27]. We hence take the $10\%$ lowest quality patches. The normalized quality of patch $i$ is set as $c_i$.

We then uniformly distribute $c_i$ into $L$ levels, and group patches having equal quality into the same quality group, w.r.t $G_l$, $l = 1, 2, ..., L$. So far, a set of quality aware clustering is generated. A high pass filter is then utilized to enhance the clustering accuracy, as follows

$$\mathbf{h}_\sigma(r) = l_{r=0} - \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{r^2}{2\sigma^2}) \tag{2}$$

where $\sigma$ is the scale parameter to control the shape of the filter. Image details will be enhanced by convolving $\mathbf{h}_\sigma$ with the image. The filter is a special case of DoG filter when the support size of the first Gaussian shrinks to 1.

Unsupervised feature learning method K-means is performed to learn a set of quality-aware centroids $\{\mathbf{m}_{l,k}\}$, $k = 1, 2, ...K$. At last, $L$ sets of centroids on $L$ different quality levels is obtained. The flowchart of getting the quality-aware centroids is presented in Fig. 2.

### B. Feature pooling

For test image $\mathbf{y}$, the feature vector of overlapped patch $\mathbf{y}_i$ is $\mathbf{f}_i^y$, $i = 1, 2, ..., N$. The nearest centroid of feature vector $\mathbf{f}_i^y$ at quality level $l$ is $\mathbf{m}_{l,k_i}$. And the distance between them is $\delta_{l,i} = ||\mathbf{f}_i^y - \mathbf{m}_{l,k_i}||^2$. Obviously, the shorter the distance is, the more likely $\mathbf{y}_i$ should have the same quality with centroid $\mathbf{m}_{l,k_i}$. A weighted average rule is used to determine the final quality of $\mathbf{y}_i$

$$z_i = \frac{\sum_{l=1}^{L} q_l exp(-\delta_{l,i}/\lambda)}{\sum_{l=1}^{L} exp(-\delta_{l,i}/\lambda)} \tag{3}$$

where $\lambda$ is a parameter to control the decay rate of weight $exp(-\delta_{l,i}/\lambda)$.

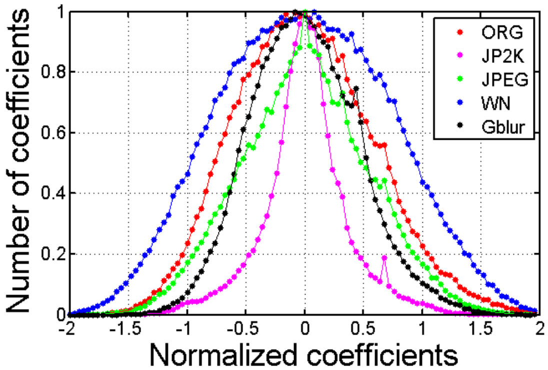


Fig. 3. Histogram of normalized coefficients for a natural undistorted image and its various distorted versions from the TID2008 database. ORG: the original image. JP2K: JPEG2000 compression. JPEG: JPEG compression. WN: additive white Gaussian noise. Gblur: Gaussian blur..

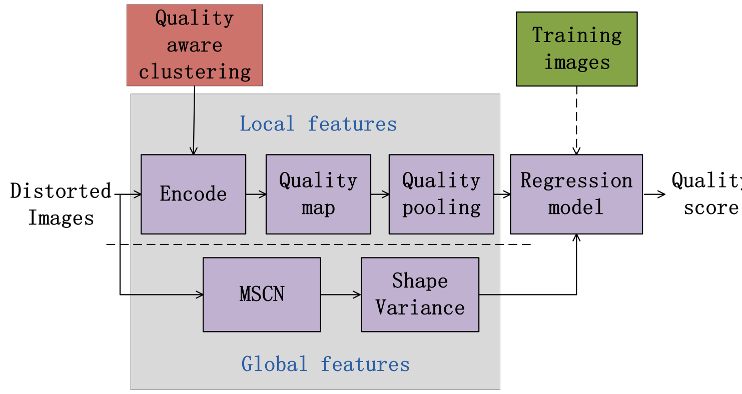


Fig. 4. The framework of our proposed general model TSBIQA.

The final quality map of $\mathbf{y}$ can be inferred with each estimated quality of patch $\mathbf{y}_i$ is available. Common used strategies like max pooling and percentile pooling have been released before. We adopt Minkowski pooling in this paper, for the reason that Xue *et al.* claimed in [28] that image local quality degradation can better reflect quality of image. The Minkowski summation is given by

$$z = \frac{1}{N}\sum_{i=1}^{N} z_i^p \tag{4}$$

where $N$ is the number of overlapped patches, and $p$ is the Minkowski exponent. We set $p$ = {1, 2, 3, 4} here. More focus will be shifted to lower quality regions with $p$ increases. This makes sense intuitively because human eyes tend to higher distortion regions when most distortion regions concentrate in a small region, and the subjective score will be lower than direct average of the quality map.

### C. Global feature extraction

In 1994, Ruderman claimed in his literature [29] the regularity of natural scene statistics (NSS). To further explore the characteristics of images in the spatial domain, we achieve normalized luminance by subtracting the local mean and performing divisive normalization, which is as follows.

$$\hat{\mathbf{y}}(i,j) = \frac{\mathbf{y}(i,j) - \mu(i,j)}{\sigma(i,j) + T_3} \tag{5}$$

where $T_3$ is a constant that prevents instabilities when denominator tends to zero. Besides, $i \in$ 1, 2,..., $H$, $j \in$ 1, 2,..., $W$, and $H$ and $W$ are the height and width of image. The mean $\mu$ and variance $\sigma$ are specified as

$$\mu_y(i,j) = \sum_{p=-P}^{P}\sum_{q=-Q}^{Q} w_{p,q}\mathbf{y}_{p,q}(i,j) \tag{6}$$

$$\sigma_y(i,j) = (\sum_{p=-P}^{P}\sum_{q=-Q}^{Q} w_{p,q}(\mathbf{y}_{p,q}(i,j) - \mu_y(i,j))^2)^{1/2} \tag{7}$$

where $w$ = $\{w_{p,q}|$p = -P,...,P, q = -Q,...,Q$\}$ is a 2D circularly-symmetric Gaussian weighting function sampled out to 3 standard deviations and rescaled to unit volume.

The derived $\hat{\mathbf{y}}(i,j)$ is defined as mean subtracted contrast normalized (MSCN) coefficients. To better describe the variation of MSCN to different distortions, we stochastically choose an image from TID2008 [32] database, and visualize the histogram of MSCN coefficients with various distortions, as can be seen in Fig. 3. This discovery of variation of natural statistics to distortions enables the implementation of BIQA without the presence of reference images. Through observation, generalized Gaussian distribution (GGD) is found to be able to effectively capture a broader spectrum of distorted image statistics, and the GGD with zero mean is given by

$$f(\mathbf{y};\alpha,\sigma^2) = \frac{\alpha}{2\beta\gamma(1/\alpha)} exp(-(\frac{|\mathbf{y}|}{\beta})^{\alpha}) \tag{8}$$

where

$$\beta = \sigma\sqrt{(\frac{\gamma(1/\alpha)}{\gamma(3/\alpha)})} \tag{9}$$

among which, $\gamma$ is the gamma function,

$$\gamma = \int_0^{\infty} t^{\alpha-1}e^{-t}dt \qquad \alpha > 0 \tag{10}$$

The shape parameter $\alpha$ takes control of the shape of the distribution, while $\sigma^2$ controls the variance. They are used to capture image distortion. Natural images are thought to possess many statistical properties in spatial domain. We adopt two basic characters shape and variance in this paper.

### D. SVR based regression

Support Vector Regression (SVR) is utilized here for regression, which is used to learn a mapping from from feature space to quality scores, generating a image quality measure. SVR possess qualities like high performance in high-dimensional spaces, over-fitting avoidance, and good generalization. We use the LIBSVM [30] package to implement the SVR with a radial basis function (RBF) kernel. The framework of the proposed general model TSBIQA is shown in Fig. 4.

TABLE I. SROCC AND PLCC ON LIVE DATABASE.

| SROCC | JP2K | JPEG | WN | BLUR | ALL |
|---|---|---|---|---|---|
| PSNR | 0.870 | 0.885 | 0.942 | 0.761 | 0.867 |
| SSIM | 0.939 | 0.946 | 0.964 | 0.907 | 0.910 |
| QAC | 0.868 | 0.938 | 0.952 | **0.918** | 0.877 |
| TSBIQA | **0.927** | **0.953** | **0.976** | 0.902 | **0.929** |
| PLCC | JP2K | JPEG | WN | BLUR | ALL |
| PSNR | 0.873 | 0.876 | 0.926 | 0.766 | 0.853 |
| SSIM | 0.921 | 0.955 | 0.982 | 0.891 | 0.900 |
| QAC | 0.851 | 0.943 | 0.924 | **0.919** | 0.863 |
| TSBIQA | **0.930** | **0.964** | **0.982** | 0.886 | **0.932** |

TABLE II. SROCC AND PLCC ON TID2008 DATABASE.

| SROCC | JP2K | JPEG | WN | BLUR | ALL |
|---|---|---|---|---|---|
| PSNR | 0.838 | 0.887 | 0.917 | 0.929 | 0.869 |
| SSIM | 0.962 | 0.932 | 0.847 | 0.959 | 0.905 |
| QAC | 0.890 | 0.887 | 0.717 | **0.856** | 0.861 |
| TSBIQA | **0.907** | **0.900** | **0.860** | 0.830 | **0.906** |
| PLCC | JP2K | JPEG | WN | BLUR | ALL |
| PSNR | 0.888 | 0.880 | 0.945 | 0.914 | 0.845 |
| SSIM | 0.971 | 0.964 | 0.816 | 0.954 | 0.902 |
| QAC | 0.878 | 0.917 | 0.736 | **0.842** | 0.842 |
| TSBIQA | **0.909** | **0.939** | **0.856** | 0.828 | **0.901** |

TABLE III. SROCC AND PLCC ON CSIQ DATABASE.

| SROCC | JP2K | JPEG | WN | BLUR | ALL |
|---|---|---|---|---|---|
| PSNR | 0.910 | 0.891 | 0.933 | 0.809 | 0.885 |
| SSIM | 0.962 | 0.954 | 0.912 | 0.960 | 0.934 |
| QAC | 0.888 | **0.912** | 0.865 | 0.852 | 0.858 |
| TSBIQA | **0.908** | 0.898 | **0.915** | **0.914** | **0.907** |
| PLCC | JP2K | JPEG | WN | BLUR | ALL |
| PSNR | 0.861 | 0.887 | 0.946 | 0.771 | 0.856 |
| SSIM | 0.906 | 0.982 | 0.910 | 0.945 | 0.930 |
| QAC | 0.896 | 0.947 | 0.911 | 0.861 | 0.890 |
| TSBIQA | **0.930** | **0.957** | **0.929** | **0.932** | **0.935** |

TABLE IV. SROCC AND PLCC ON TID2013 DATABASE.

| SROCC | JP2K | JPEG | WN | BLUR | ALL |
|---|---|---|---|---|---|
| PSNR | 0.884 | 0.919 | 0.923 | 0.915 | 0.907 |
| SSIM | 0.905 | 0.910 | 0.853 | 0.963 | 0.850 |
| QAC | 0.790 | 0.837 | 0.7414 | 0.846 | 0.805 |
| TSBIQA | **0.918** | **0.886** | **0.907** | **0.874** | **0.902** |
| PLCC | JP2K | JPEG | WN | BLUR | ALL |
| PSNR | 0.917 | 0.917 | 0.949 | 0.914 | 0.891 |
| SSIM | 0.915 | 0.931 | 0.859 | 0.958 | 0.829 |
| QAC | 0.809 | 0.869 | 0.794 | 0.847 | 0.805 |
| TSBIQA | **0.935** | **0.938** | **0.910** | **0.868** | **0.907** |

# III. EXPERIMENTAL RESULTS

## A. Protocol

Four databases LIVE [31], TID2008 [32], CSIQ [33] and TID2013 [34] are used as testing beds to validate the effectiveness of our method.

1) LIVE database: it consists of 29 reference images each with five different types of distortions - JPEG, JPEG2000 (JP2K), white gaussian noise (WN), Gaussian blur (BLUR) and fast fading (FF). Each distortion has 5 to 6 distortion levels. Image subjective score (DMOS) value is in the range of [0, 100].

2) TID2008 database: it involves of 25 reference images and 1700 distorted images with 17 different distortions at 4 levels. MOS values are provided to every distorted image. The higher MOS value, the better image quality.

3) CSIQ database: 30 reference images and derived 6 different types of distortions at 4 to 5 different levels are consisted in this database. A total of 900 images are included.

4) TID2013 database: it consists of 25 original images and 3000 distorted images with 24 different distortions at 5 levels. Similar to TID2008 database, MOS values are also included for every distorted image.

The overlapped distortion types among these four databases are JPEG, JP2K, WN and BLUR, so we only consider these four distortion types in our experiment. They are also the most commonly encountered distortions in digital images.

## B. Implementation details

The results are obtained by 1000 train-test iterations with randomly selected 10 images from Berkeley Segmentation database [35], along with generated JPEG, JP2K, WN, and BLUR distorted images with distortions at ten different levels. LIVE, TID2008, CSIQ and TID2013 databases are used as testing sets. We use linear $\epsilon$-SVR for regression in this part. For both training and testing images, only distorted images are used. Additionally, with 1000 times training and testing, the median results are reported finally.

## C. Evaluation

Person linear correlation coefficient (PLCC), Spearman rank-order correlation coefficient (SROCC) are used to evaluate our system performance. PLCC can be considered as a measure of prediction accuracy, while SROCC computes the monotonicity by ignoring the relative distance between the data. The higher SROCC and PLCC values indicate better performance in terms of correlation with human opinion. A four-parameter logistic function is chosen to fit the scores of our method to subjective scores [36]

$$Quality(z) = \frac{\beta_1 - \beta_2}{1 + exp(-(z - \beta_3)/\beta_4)} + \beta_2 \qquad (11)$$

where $z$ is the input score, and $Quality(z)$ is the mapped score, and $\beta_1$ to $\beta_4$ are free parameters to be determined during the curve fitting process. Two traditional FR-IQA metrics peak-signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [37], and one OF-BIQA metric QAC [21] are used as reference to validate TSBIQA. Experiment results of SROCC and PLCC values on LIVE, TID2008, CSIQ, and TID2013 databases on four specific distortions and the whole datasets are tabulated in Table I - Table IV respectively.

From Table I-IV, we can clearly see that our model achieves remarkable results on the testing four subjective image databases. Note that FR-IQA algorithms are hardly competed with BIQA due to the requirement of the lossless

TABLE V. THE PERCENTAGE OF PERFORMANCE IMPROVEMENT OF TSBIQA THAN QAC ACCORDING TO SROCC (%).

| | JP2K | JPEG | WN | BLUR | ALL |
|---|---|---|---|---|---|
| LIVE | 6.797 | 1.599 | 2.521 | -1.742 | 5.929 |
| TID2008 | 1.910 | 1.466 | 19.944 | -3.037 | 5.227 |
| CSIQ | 2.252 | -1.535 | 5.780 | 7.277 | 5.711 |
| TID2013 | 16.203 | 5.854 | 22.336 | 3.310 | 12.050 |

image, but the proposed TSBIQA model is superior to FR PSNR, and matchable with the popular FR SSIM. Further, TSBIQA has achieved remarkable improvements of $5\%$ higher than QAC on LIVE, TID2008, CSIQ and TID2013 databases in terms of SROCC value, as listed in Table V, especially performing significantly better on TID2013 database. From the performance evaluations, it can be found that QAC works better on TID2008 than TID2013, which is probably explained by the fact that QAC is not good at capturing low image distortions which are contained in TID2013 database. For each single distortion type, TSBIQA also attains favorable results. It only achieves a little bit lower than QAC on BLUR datasets on LIVE, TID2008 databases, and on JPEG datasets on CSIQ database. However, we conclude that even with little features provided, our paradigm achieves promising performance.

## IV. CONCLUSION

We propose a new effective and efficient three step (TS-BIQA) method in this paper by combining local and global features. K-means based quality aware clustering constructs centroids of each image level to infer image quality of each patch. Additionally, two natural scene statistics, shape and variance of MSCN, are used as global features to express the overall quality of the image. SVR is adopted at last to generate a regression model. Importantly, the measure we propose in this paper is a general framework, and in other words, the local or global features can be flexibility replaced in this kind of application.


## ACKNOWLEDGMENT

This work was supported in part by NSFC (61025005, 61371146, 61221001, 61390514), 973 Program (2010CB731401) and FANEDD (201339).


## REFERENCES


[1] Z. Wang, H. R. Sheikh, and A. C. Bovik, “No-reference perceptual quality assessment of JPEG compressed images,” *in Proc. IEEE Int. Conf. Image Process.*, Rochester, NY, Sep. 2002, vol. 1, pp. I-477-I-480.

[2] P. Gastaldo and R. Zunino, “Neural networks for the no-reference assessment of perceived quality,” *J. Electron. Imag.*, vol. 14, no. 3, p. 033004, Aug. 2005.

[3] T. Brandão and M. P. Queluz, “No-reference image quality assessment based on DCT domain statistics,” *Signal Process.*, vol. 88, no. 4, pp. 822-833, Apr. 2008.

[4] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, “Perceptual blur and ringing metrics: Application to JPEG2000,” *Signal Process. Image Commun.*, vol. 19, no. 2, pp. 163-172, Feb. 2004.

[5] H. R. Sheikh, A. C. Bovik, and L. Cormack, “No-reference quality assessment using natural scene statistics: JPEG2000,” *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1918-1927, Nov. 2005.

[6] K. Gu, G. Zhai, X. Yang, W. Zhang, and M. Liu, “Subjective and objective quality assessment for images with contrast change,” *in Proc. IEEE Int. Conf. Image Process.*, pp. 383-387, Sept. 2013.

[7] K. Gu, G. Zhai, M. Liu, X. Yang, and W. Zhang, “Details preservation inspired blind quality metric of tone mapping methods,” *in Proc. IEEE Int. Symp. Circuits and Syst.*, pp. 518-521, Jun. 2014.

[8] M. Liu, G. Zhai, Z. Zhang, Y. Sun, K. Gu, and X. Yang, “Blind image quality assessment for noise,” *in Proc. IEEE Int. Symp. Broadband Multimedia Systems and Broadcasting*, pp. 1-5, Jun. 2014.

[9] K. Gu, G. Zhai, X. Yang, and W. Zhang, “Hybrid no-reference quality metric for singly and multiply distorted images,” *Trans. Broadcasting*, vol. 60 , no. 3, pp. 555-567, Sept. 2014.

[10] A. K. Moorthy and A. C. Bovik, “A two-step framework for constructing blind image quality indices,” *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 513-516, May 2010.

[11] M. A. Saad, A. C. Bovik, and C. Charrier, “A DCT statistics-based blind image quality index,” *IEEE Signal Process. Lett.*, vol. 17, no. 6, pp. 583-586, Jun. 2010.

[12] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik, “No-Reference Image Quality Assessment in the Spatial Domain,” *IEEE Trans. Image Process.*, vol. 21, no.12, December 2012.

[13] K. Gu, G. Zhai, X. Yang, W. Zhang, and L. Liang, “No-reference image quality assessment metric by combining free energy theory and structural degradation model,” *in Proc. IEEE Int. Conf. Multimedia and Expo*, pp. 1-6, Jul. 2013.

[14] K. Gu, G. Zhai, X. Yang, and W. Zhang, “Deep learning network for blind image quality assessment,” *in Proc. IEEE Int. Conf. Image Process.*, Oct. 2014, to appear.

[15] W. Lu, K. Zeng, D. Tao, Y. Yuan, and X. Gao, “No-reference image quality assessment in contourlet domain,” *Neurocomputing*, vol. 73, no. 4-6, pp. 784-794, Jan. 2010.

[16] C. Li, A. C. Bovik, and X.Wu, “Blind image quality assessment using a general regression neural network,” *IEEE Trans. Neural Netw.*, vol. 22, no. 5, pp. 793-799, May 2011.

[17] Z. Wang and A. C. Bovik, “Reduced- and no-reference image quality assessment: The natural scene statistic model approach,” *IEEE Signal Process. Mag.*, vol. 28, pp. 29-40, Nov. 2011.

[18] P. Ye and D. Doermann, “No-reference image quality assessment using visual codebooks,” *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3129C3138, Jul. 2012.

[19] G. Zhai, X. Wu, X. Yang, W. Lin, and W. Zhang, “A psychovisual quality metric in free-energy principle,” *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 41-52, Jan. 2012.

[20] K. Gu, G. Zhai, X. Yang, and W. Zhang, “A new reduced-reference image quality assessment using structural degradation model,” *in Proc. IEEE Int. Symp. Circuits and Systems*, pp. 1095-1098, May 2013.

[21] A. Mittal, G.S. Muralidhar, J. Ghosh, and A.C. Bovik, “Blind image quality assessment without human training using latent quality factors,” *IEEE Signal Process. Letters*, vol. 19, no. 2, pp. 75-78, Feb. 2012.

[22] A.K. Moorthy and A.C. Bovik, “Making a “completely blind” image quality analyzer,” *IEEE Signal Process. Letters*, vol. 20, no. 3, pp. 209-212, Mar. 2013.

[23] Wufeng Xue, Lei Zhang, Xuanqin Mou, “Learning without human scores for blind image quality assessment,” *Computer Vision and Pattern Recognition (CVPR), IEEE,* 2013.

[24] K. Gu, G. Zhai, X. Yang, and W. Zhang, “An efficient color image quality metric with local-tuned-global model,” *in Proc. IEEE Int. Conf. Image Process.*, Oct. 2014, to appear.

[25] S. Goferman, L. Zelnik-Manor, and A. Tal, “Context-aware saliency detection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 1915-1926, Oct. 2012.

[26] L. Zhang, L. Zhang, X. Mou, and D. Zhang. “Fsim: A feature similarity index for image quality assessment”. *IEEE Trans. Image Process.*, 20(8):2378-2386, 2011.

[27] A. Moorthy and A. Bovik. “Visual importance pooling for image quality assessment”. *IEEE Journal of Selected Topics in Signal Processing*, 3(2):193C201, 2009.

[28] W. Xue, L. Zhang, X. Mou and A. Bovik, “Gradient Magnitude Similarity Deviation: A Highly Efficient Perceptual Image Quality Index,” *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 684, 695, Feb. 2014.

[29] D. L. Ruderman, “The statistics of natural images,” *Netw. Comput. Neural Syst.*, vol. 5, no. 4, pp. 517-548, 1994.

[30] C. C. Chang and C. J. Lin. (2001). LIBSVM: A Library for Support Vector Machines [Online]. Available: http://www.csie.ntu.edu.tw/∼ cjlin/libsvm/

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

[32] 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*, vol. 10, pp. 30-45, 2009.

[33] E.C. Larson and D.M. Chandler, “Most apparent distortion: Full-reference image quality assessment and the role of strategy,” *J. Electr. Imaging*, vol. 19, no. 1, pp. 001006: 3440-3451, Mar. 2010.

[34] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, and F. Battisti, “Color image database TID2013: Peculiarities and preliminary results,” *in 4th European Workshop on Visual Information Processing*, 2013.

[35] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. *In ICCV*, 2001.

[36] VQEG, “Final report from the video quality experts group on the validation of objective models of video quality assessment,” March 2000, http://www.vqeg.org/.

[37] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Trans. Image Process.*, 2004.