

NO-REFERENCE IMAGE QUALITY MEASURE FOR IMAGES WITH MULTIPLE DISTORTIONS USING MULTI-METHOD FUSION

KANJAR DE[✉] AND MASILAMANI V

Indian Institute of Information Technology, Design and Manufacturing Kancheepuram, Tamil Nadu, 600127, India

e-mail: kanjar.de@gmail.com, masila@iiitdm.ac.in

(Received April 26, 2016; revised February 12, 2018; accepted February 12, 2018)

ABSTRACT

Over the years image quality assessment is one of the active area of research in image processing. Distortion in images can be caused by various sources like noise, blur, transmission channel errors, compression artifacts etc. Image distortions can occur during the image acquisition process (blur/noise), image compression (ringing and blocking artifacts) or during the transmission process. A single image can be distorted by multiple sources and assessing quality of such images is an extremely challenging task. The human visual system can easily identify image quality in such cases, but for a computer algorithm performing the task of quality assessment is a very difficult. In this paper, we propose a new no-reference image quality assessment for images corrupted by more than one type of distortions. The proposed technique is compared with the best-known framework for image quality assessment for multiply distorted images and standard state of the art Full reference and No-reference image quality assessment techniques available.

Keywords: human visual system (HVS), image quality assessment (IQA), multiply distorted images, no-reference image quality assessment (NR-IQA).

INTRODUCTION

Over the years, image quality assessment is an active area of research followed by researchers. Distortions in images can be caused by many sources like during acquisition by faulty sensors or camera shake or can be introduced as an artifact of compression algorithms or can occur during transmission in a noisy channel. There can be multiple sources of distortion in a single image. Human visual system (HVS) can easily detect the quality of image under such conditions but designing an algorithm for performing this task is an extremely challenging task. Image quality assessment is very important in many image processing applications. With the invention of display devices like hand-held devices, high definition televisions, LED monitors, networked television (IPTV's) etc. there is a large amount of data in the form of visual signals which has increased exponentially over the years. Due to bandwidth constraints, data needs to be compressed for transmission and distortions can occur during compression process or can occur as noise or loss in the transmission channel.

Image quality assessment algorithms are classified broadly into three categories based on the requirement of a reference image.

1. Full reference image quality assessment
2. Reduced reference image quality assessment

3. No-reference image quality assessment algorithms (Wang and Bovik, 2006).

The first class of algorithms is known as full reference image quality assessment algorithms (FR-IQA) where the assessment of the quality of the image requires a reference image of the same scene which is assumed to be of good quality. The image quality assessment of a target image is performed by comparing it against the full information of the reference image. Peak signal to noise ratio (PSNR; Eskicioglu and Fisher, 1995) and structural similarity index measure (SSIM; Wang *et al.*, 2004) are the classical examples of FR-IQA. Recently a lot of new full reference techniques like Feature similarity index measure (FSIM; Zhang *et al.*, 2011), a measure based on gradient similarity (GSM; Liu *et al.*, 2012), a measure based on internal generative mechanism (IGM; Wu *et al.*, 2013) and a measure based on gradient magnitude and laplacian features (GMSD; Xue *et al.*, 2014) have been proposed in literature.

The next category of algorithms also require a reference image of the same scene but instead of the full reference image, only certain features of the reference image which give information about the visual quality are used to differentiate from a target image for the task of quality assessment using the quality aware features. These class of algorithms is known as reduced reference image quality assessment algorithms (RR-IQA). Reduced

reference entropic differencing (Soundararajan and Bovik, 2012) and reduced reference technique based on structural similarity estimation (Rehman and Wang, 2012) are some of the well-known techniques in this category.

Finally, the third category of image quality assessment algorithms which assess the quality of an image blindly without the requirement of a reference image of the same scene are known as no-reference image quality assessment algorithms (NR-IQA). NR-IQA is an active area of research in the field of image processing and it is the most challenging among the three strategies of image quality assessment. There are various approaches to no-reference image quality assessment. In some approaches, statistical features which give information about the visual quality of images are extracted and an image quality score is computed from these features. NIQE (Mittal *et al.*, 2013) is an example of NR-IQA where 'quality aware' statistical features are extracted from the images and an image quality score is proposed. The next class of no-reference image quality assessment techniques is machine learning based. Human psychophysics experiments are conducted on a dataset of images and subjective numerical human opinion scores are obtained in the form of differential mean opinion scores (DMOS). Machine learning algorithms like support vector regression (SVR), general regression neural networks (GRNN) etc. are trained using the features extracted from images that provide information of visual quality and DMOS scores (Xu *et al.*, 2015). Blind image quality index (BIQI; Moorthy and Bovik, 2010), blind image integrity notator using DCT Statistics (BLINDS; Saad *et al.*, 2012), distortion identification based image verity and integrity evaluation (DIIVINE; Moorthy and Bovik, 2011) and GRNN based no-reference image quality index (GRNN-NRQI; Li *et al.*, 2011) are some of the state of the art no-reference image quality assessment algorithms that use machine learning. A difference of Gaussian (DOG) features based full reference image quality assessment algorithm uses random forest regression (Pei and Chen, 2015). An image quality assessment based on contrast measure using random forests have been developed in literature (De and Masilamani, 2017b).

The next category of no-reference image quality assessment algorithms are distortion specific, these algorithms are designed for images corrupted by a single specific type of distortion, like additive white Gaussian noise (AWGN), blur, JPEG compression artifacts, JPEG2000 artifacts etc. Noise can occur in the images during image acquisition from faulty sensors. Pyatykh *et al.* (2013), Liu *et al.* (2013b),

Zoran and Weiss (2009) have all proposed algorithms which give an approximate estimate of amount or level of additive noise present in the distorted images. More the amount of noise present in the images, the poorer is the visual quality of the image. Blur in the image can occur from a variety of sources, for example during image acquisition process, due to camera shake or defocus. Analyzing image sharpness/blurriness is an active area of research among researchers working in the field of image quality assessment. Blur assessment techniques just noticeable blur measurement (JNBM; Ferzli and Karam, 2009), cumulative probability of blur detection (CPBD; Narvekar and Karam, 2011), local phase coherence (Hassen *et al.*, 2013), auto-regression space (Gu *et al.*, 2015a), Tchebichef moments (Li *et al.*, 2016) are a few of the recent well-known techniques for the assessment of image quality for blurred images. For transmission of visual data due to bandwidth constraints, it is required to compress the data using some good compression algorithms. However, compression of images can reduce the visual quality of the image by introducing some compression artifacts. Lossy compression can lead to loss of information from the image which in turn may lead to a reduction in visual quality of the image. JPEG and JPEG2000 are the most commonly used image compression schemes. JPEG Quality Estimator by Wang *et al.* (2002) and no-reference JPEG quality assessment (NJQA) proposed by Golestaneh (Golestaneh and Chandler, 2014) are standard image quality estimators for images compressed by JPEG algorithm. Sheikh, Bovik and Cormack have proposed an image quality assessment algorithm for JPEG2000 (Sheikh *et al.*, 2005).

In display devices, before the image reaches the end user it will undergo three stages. They are acquisition, compression, and transmission. Images can be corrupted in any of these three stages or in all the three stages. There can be more than one source of distortion in a single image. Performing image quality assessment of an image where multiple distortions are present is an extremely challenging task. A single image can have multiple distortions, common examples image is blurring while acquisition due to camera shake or defocus and then JPEG compression artifacts may be added to it during compression. Similarly, the image may be subjected to blur as well as additive noise during image acquisition process. Fig. 1 shows certain examples of images corrupted by blur-JPEG combination and blur-noise combination from the standard publicly available LIVE multi-distortion (LIVEMD; Jayaraman *et al.*, 2012) dataset. Single images can have more than two types of distortions also. Fig. 2 shows examples from the MDID2013 dataset (Gu *et al.*, 2014), here all the images are

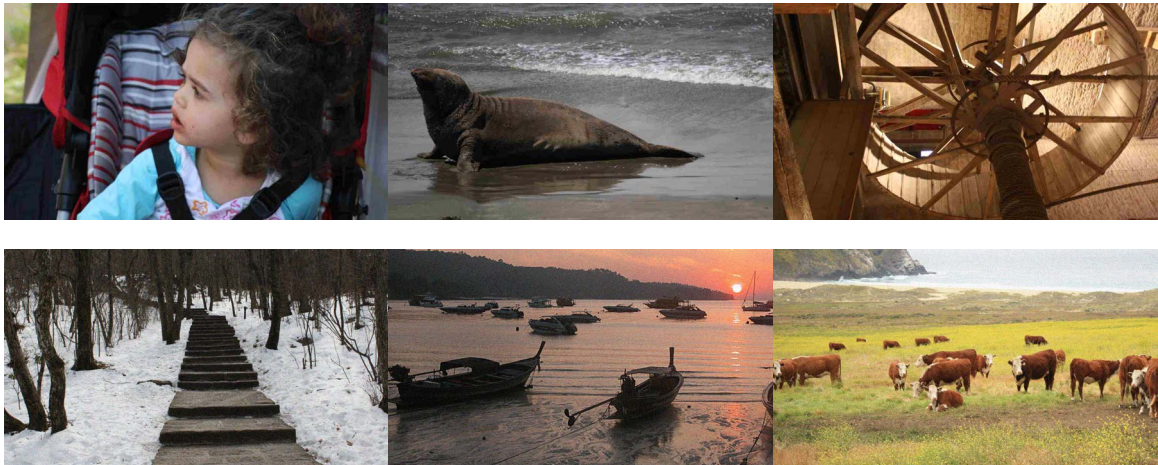


Fig. 1: Six images from LIVEMD database: Top three images are Blur + JPEG compressed, bottom three are Blur + additive white Gaussian noise.

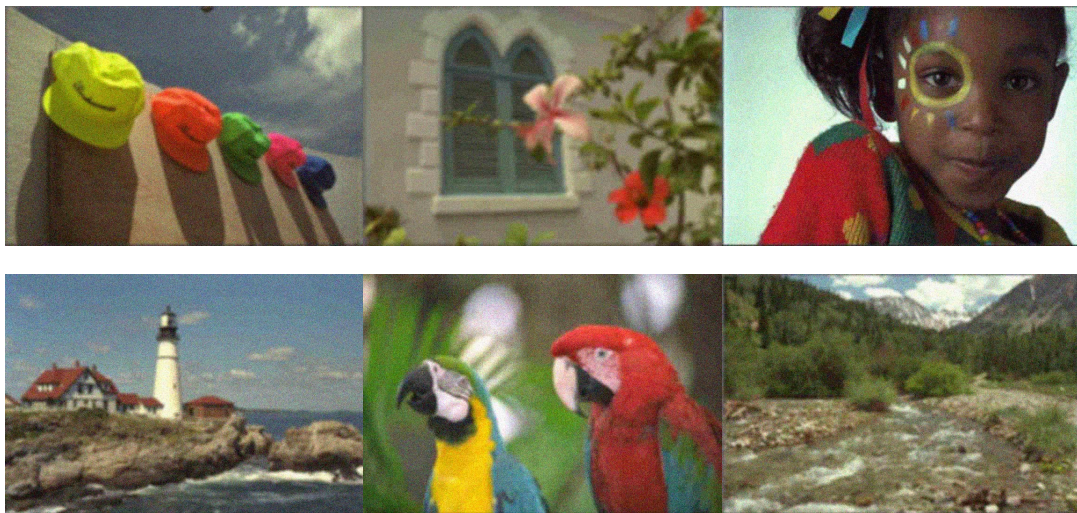


Fig. 2: Six images from MDID2013 database: Blur + JPEG compressed + additive white Gaussian noise.

subjected to all blur, JPEG compression artifacts, and additive noise distortions.

The earliest known work to tackle the challenge of image quality assessment for multiply distorted images was proposed by Gu *et al.* (2013) where they proposed a five-step blind metric for quality assessment of multiply distorted images (FISBLIM). A new framework was proposed where firstly, it was checked whether additive noise was present in the distorted image and if present then the noise was estimated by the algorithm proposed by Zoran and Weiss (2009) followed by denoising using the state of the art BM3D algorithm (Dabov *et al.*, 2007) and then blur analysis was done using perceptual blur metric (Marziliano *et al.*, 2002) and JPEG analysis

was done using technique proposed by Wang *et al.* (2002). Later Gu *et al.* (2014) modified the five step blind metric to a more advanced six step blind metric (SISBLIM) where they added another step to incorporate the concept of free energy theory. The concept of free energy in the problem of image quality assessment is explained in detail in Gu *et al.* (2015b). SISBLIM to best of our knowledge is the current best-known technique for image quality assessment of multiply distorted images. In this framework, different combinations of noise and blur metrics have been tested and results are presented in Gu *et al.* (2014). In this framework for image quality assessment first it is checked whether additive noise is present or not in the image, then if it is present then additive noise

is measured and removed from the image and then the quality assessment for other distortions is performed on the denoised image and finally the different quality scores for different distortions are fused into a single image quality score.

The rest of the paper is organized as follows the section *Methods* describes the proposed technique to assess quality of images degraded by multiple distortions in detail, followed by *Results* section, where we provide the details of the experiments involved and then we validate the results with human visual system and present a detailed comparative study of proposed technique against the state of the art image quality assessment algorithms and explain it in *Discussion* section and finally we conclude in section *Conclusion*.

METHODS

In this section, we explain the different concepts used to implement our proposed scheme. In literature, different techniques related to image quality assessment of images for different types of distortion like additive noise, blur and JPEG compression artifacts have been proposed. We use few of the state of the art distortion specific and general image quality measures to propose a new scheme which uses multi-method fusion to assess the quality of the images distorted by more than one type of distortion. Initially different distortion specific measures were chosen to form a feature vector, then only the best performing image quality measures were chosen to form the final feature vector. The best features chosen using the sequential forward search feature selection (SFFS) algorithm which is run on a set of image quality measures. The SFFS algorithm has been used in image quality assessment algorithms successfully (Liu *et al.*, 2013a; De and Masilamani, 2017a) is shown in Fig. 3. The sequential forward search selection algorithm is run based on maximization of an optimization function J which for the proposed work is Spearman rank correlation coefficient given in Eq. 1 between the objective image quality measure and subjective image quality measure (DMOS) obtained from the database. Let the total number of image quality measures to begin with be denoted by T . We propose a new technique which uses random forests technique for multi-method fusion of different no reference image quality assessment techniques.

Algorithm 1 Sequential Forward Feature Selection

- 1: Start with empty feature set $F_0 = \phi$ and $k = 0$.
- 2: Select the next best feature $f^* =$

$$\arg \max_{f \in F - F_k} J(F_k \cup \{f\})$$

- 3: Update $F_{k+1} = F_k \cup \{f^*\}; k = k + 1$
 - 4: Go to 2 till $k \leq T$.
-

Fig. 3: Sequential forward search selection algorithm.

PROPOSED TECHNIQUE

The block diagram of proposed image quality assessment scheme is shown in Fig. 4. Generally machine learning based techniques have two stages:

- Training stage - In training phase we train a model from training data by giving input in the form of $\{X, y\}$ pair where $X = x_1, x_2, \dots, x_n$ is the input feature vector and y is the corresponding subjective image quality score obtained by running psychophysics experiments on humans.
- Testing stage - In testing phase we used the trained model to test on inputs which were not part of the training set and at the output we get an objective image quality score which is expected to be closer to human opinion scores.

Table 1: Image quality measures as features.

Algorithm	Type of Distortion
<i>SINE</i> (Zoran and Weiss, 2009)	Noise
<i>WPTNE</i> (Liu <i>et al.</i> , 2013b)	Noise
<i>S₃</i> (Vu <i>et al.</i> , 2012)	Blur
<i>JNBM</i> (Ferzli and Karam, 2009)	Blur
<i>LPCSI</i> (Hassen <i>et al.</i> , 2013)	Blur
<i>ARISIM</i> (Gu <i>et al.</i> , 2015a)	Blur
<i>Q-metric</i> (Zhu and Milanfar, 2010)	Blur
<i>FISH</i> (Vu and Chandler, 2012)	Blur
<i>FISH_{bb}</i> (Vu and Chandler, 2012)	Blur
<i>Sharpness_index</i> (Leclaire and Moisan, 2015)	Blur
<i>S_index</i> (Blanchet and Moisan, 2012)	Blur
<i>BIBLE</i> (Li <i>et al.</i> , 2016)	Blur
<i>JPEGQ</i> (Wang <i>et al.</i> , 2002)	JPEG
<i>BIQ</i> (Gabarda and Cristóbal, 2007)	Image Quality

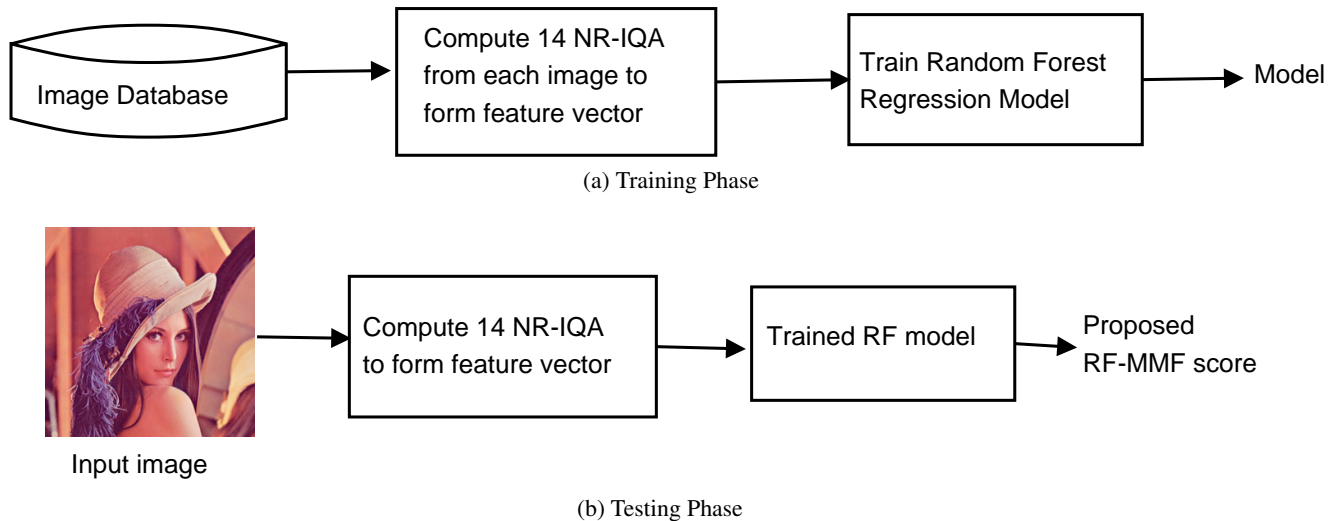


Fig. 4: Proposed IQA System: (a) training the model. (b) testing the model.

We are proposing a new technique for image quality assessment for images which are corrupted by multiple distortions by fusing or combining different image quality measures which are already existing in literature for different types of distortions. The proposed method is a multi-method fusion technique and we are denoting it as random forest-multi-method fusion (*RF-MMF*). Table 1 shows the list of measures which we have considered for our proposed technique. These 14 image quality measures are used as a feature vector for training a random forest regressor (Breiman, 2001).

Measures for Noise

In the last section, we have introduced certain measures which are proposed in the literature for the purpose of detecting the amount of noise present in the image. In the proposed method we have used two techniques available in the literature as features of random forest regression. These techniques are:

1. Zoran and Weiss have proposed a technique describing noise estimation in images. This technique uses kurtosis values for the purpose of noise estimation as kurtosis values are scale invariant in natural images. Details of the implementation are available in Zoran and Weiss (2009). This technique is referred as *SINE* in the rest of the paper.
2. Liu et. al has proposed a technique for noise level estimation using weak textured patches of a single noisy image. This technique will be referred as *WPTNE* in the rest of the paper. The details of implementation of this noise level estimator are available in Liu *et al.* (2013b).

Measures for Blur

As mentioned earlier, analysis of image sharpness/blurriness is an active area of research. Most of the no-reference techniques work reasonably well when images are corrupted by blur distortion only but in presence of any other distortion along with blur, like noise or JPEG compression the performance of these techniques reduces. In our proposed work ten of the best-known image sharpness measures available in the literature are used as features in a feature vector for random forest regressor. These blur measures individually do not work well, but in presence of distortion but combination gives good results as discussed in Section *Results*.

1. S_3 proposed by Vu *et al.* (2012) is a sharpness metric which is a combination of the spatial and spectral measure of sharpness in an image. For implementation details of this technique refer Vu *et al.* (2012).
2. *JNBM* stands for just noticeable blur measure is one of the most popular image sharpness measures in literature proposed by Ferzli and Karam (2009).
3. *LPCSI* stands for Local Phase Coherence Sharpness index - a metric developed for measuring sharpness in an image using the concept of local phase coherence. For implementation details refer Hassen *et al.* (2013).
4. *ARISM* stands for auto regressive based image sharpness metric proposed by Gu et. al, which finds an estimate of sharpness in autoregressive parameter space and it is the next sharpness measure used as a feature in our proposed method. For implementation details of *ARISM* technique refer Gu *et al.* (2015a).

5. *Q-metric* proposed by Zhu and Milanfar is a no-reference image content measure which was developed for assessing the performance of denoising algorithms. The details of implementation are available in Zhu and Milanfar (2010).
6. *FISH* and *FISH_{bb}* stand for fast image sharpness which has two variants one which uses the full image for sharpness estimation (*FISH*) and the second variation divides the images into blocks for image sharpness estimation (*FISH_{bb}*). Both these variants are used as two separate features for random forest regressor in the proposed technique. For implementation details of this wavelet based technique refer Vu and Chandler (2012).
7. Leclaire and Moisan developed a no-reference sharpness metric for deblurring using Fourier Phase information. The details of implementation are available in Leclaire and Moisan (2015).
8. Blanchet and Moisan proposed a no reference sharpness metric based on global phase coherence. For implementation details refer Blanchet and Moisan (2012).
9. *BIBLE* stands for Blind Image Blur Evaluation algorithm which is one of the recent image sharpness measure based on Tchebichef moments. The details of the proposed measure are available in Li *et al.* (2016).

Measures for JPEG

Wang proposed one of the most widely used no-reference image quality assessment algorithm for JPEG compressed images. We used this measure as a feature in the feature vector for training random forest regression model. This measure has three components blockiness, the average absolute difference among in-block image samples and zero-crossing rate. In this paper, we refer this measure as *JPEGO*. The details of implementation of this measure are available in Wang *et al.* (2002).

Measures for Image Quality

The final feature vector is an anisotropy based image quality measure proposed by Gabarda and Cristóbal (2007). The algorithm and its corresponding implementation details are available in Gabarda and Cristóbal (2007). In the rest of the paper, this measure will be referred as *BIQ*. This measure was designed as a general image quality assessment measure which works for different types of distortions present in the image.

Random Forest Regression

Random forests (Breiman, 2001) regression is one of the widely used regression technique for many machine learning applications. Random forests are an ensemble learning technique which uses multiple decision trees for performing classification or regression operations. In our proposed technique random forest regression technique is used to combine different IQA into a single score which will give an estimate of the quality of the image. We have generated a 14-dimensional vector with image quality measures mentioned in Table 1 as features and then we train the system using random forest regression model with these measures as the feature vector for training the regression model and the corresponding subjective image quality score available from the standard image quality datasets are used as the corresponding outputs to the input feature vectors. The implementation of random forest proposed in Jaiantilal (2009) is used for this work. The features are normalized to the values between 0 and 1 before applying random forest regression. Multi-method fusion (Liu *et al.*, 2013a) is one of the latest directions in image quality assessment. The purpose of using random forests regression here is to map the feature vector to a predicted quality score. In this proposed technique we perform multi-method fusion using random forests and we denote the proposed measure as Random Forest Multi-method Fusion (*RF-MMF*).

RESULTS

In this section, we try to validate the proposed method by doing comparative studies against the state of the art image quality assessment algorithms. The proposed model performance is compared by the human visual system (HVS). The human opinion scores are available with standard datasets as differential mean opinion scores (DMOS). Let D denote a database, N denotes the number of images in the database. To evaluate prediction monotonicity, we use two statistical measures Spearman rank order correlation coefficient (*SROCC*) and Kendall rank correlation coefficient (*KRCC*).

Spearman rank order correlation coefficient is defined as

$$SROCC = 1 - \frac{6}{N(N^2 - 1)} \sum_{i=1}^N d_i^2, \quad (1)$$

where d_i is the difference between the subjective DMOS rank and objective image quality score rank of the i^{th} image of the database, $i = 1, 2, \dots, N$.

Kendall rank correlation coefficient is defined as

$$KRCC = \frac{N_c - N_d}{\frac{1}{2}N(N-1)}, \quad (2)$$

where N_c and N_d denote the number of concordant and discordant pairs in the dataset D respectively, N denotes the number of images in the dataset D . To evaluate prediction accuracy we use two statistical quantities: Pearson linear correlation coefficient ($PLCC$) and root mean square error ($RMSE$).

Pearson linear correlation coefficient is defined as

$$PLCC = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 (y_i - \bar{y})^2}}, \quad (3)$$

where x_i is the subjective DMOS score of the i^{th} image and y_i is the objective image quality score of the i^{th} image, \bar{x} is the mean subjective DMOS score of the dataset D and \bar{y} denote the mean objective image quality score of dataset D , $i = 1, 2, \dots, N$, $N =$ total number of images in the dataset D .

Root mean square error is defined as

$$RMSE = \sqrt{\frac{1}{N} \sum (x_i - y_i)^2}, \quad (4)$$

where N is the total number of images in dataset D , x_i is the subjective DMOS score of the i^{th} image and y_i is the objective image quality score of the i^{th} image. Before calculating the correlations the calculated image quality measure is passed through a non-linear regression function as per recommendations given by video quality experts group (VQEG). (Rohaly *et al.*, 2000). The non linear regression function is given by

$$Q_p(\beta) = \frac{\beta_1 - \beta_2}{1 + \exp(Q - \frac{\beta - \beta_3}{\beta_4})} + \beta_2, \quad (5)$$

where $\beta_1, \beta_2, \beta_3, \beta_4$ are regression parameters, Q and Q_p are the predicted image quality scores before and after the non linear regression respectively.

PERFORMANCE WITH LIVE MULITDISTORTION DATASET

Comparison with human visual system

LIVE multidistortion (LIVE-MD) dataset (Jayaraman *et al.*, 2012) proposed by Laboratory of Image and Video engineering, University of Texas, Austin is one of the standard datasets available for the problem of image quality assessment for multiply distorted images. The dataset is partitioned into two

classes, in the first class, images are first blurred and then compressed by JPEG algorithm and in the second class, the images are first blurred and then additive Gaussian noise is added in the images. Total of 225 images generated from 15 reference images are available in each class. We have trained a random forest regression model on the 405 distorted images out of 450 images of the LIVEMD dataset in a 80/20 ratio where 80 percent of images are used for training and 20 percent for testing. 1000 different combinations of train/test sets are generated and for these trials, experiments are performed. The median performance scores of 1000 trials are presented in Table 2. The median scores are used to avoid the effect of outliers and performance bias in the conducted experiments. We have compared it against the latest known IQA, *FISBLIM*, *SISBLIM* (which has 4 variants sfb,sm,wfb,wm. Refer Gu *et al.* (2014) for details) are the algorithms developed for multiply distorted images, *PSNR* and *SSIM* are classical full reference image quality assessment and we have considered some latest full-reference image quality assessment like *IGM*, *GSM*, *FSIM* and *GMSD*. Finally we considered two popular no reference techniques *BRISQUE* and *NIQE*. The proposed image quality assessment technique performs in accordance with human visual system as the results are compared with DMOS scores which are obtained by running psychophysics experiments on human subjects.

Statistical significance and hypothesis testing

Fig. 5 shows the plot of mean Spearman rank order correlation coefficient (SROCC) values across 1000 random train-test trials and standard error bars for the competing image quality assessment algorithms. Statistical significance of each of the algorithm needs to be evaluated, for this purpose hypothesis testing using one-sided t -test (Sheskin, 2004) is performed on the SROCC values generated from the 1000 random train-test trials and the results are shown in Table 3.

Null hypothesis: Mean Correlation for the algorithm in the row is equal to the mean correlation for the algorithm in the column with a confidence of 95 %.

Alternate hypothesis: Mean correlation for the algorithm in the row is greater than or less than the mean correlation of the algorithm in the column.

A value of '1' in the table denotes that the row algorithm is statistically superior to the column algorithm, on the other hand '-1' in the table denotes that the row algorithm is statistically worse than the column algorithm. A value of '0' inside the table denotes that the algorithm in the row and algorithm in the column are statistically equivalent (or indistinguishable) which means we failed to reject the

Table 2: Performance evaluation of our proposed IQA measure and its comparison with state of the art IQA for LIVE multidistortion dataset for 1000 train-test combination trials.

Name	Type	Median SROCC	Median KRCC	Median PLCC	Median RMSE
<i>PSNR</i>	FR	0.6792	0.5050	0.7524	12.4257
<i>SSIM</i>	FR	0.6479	0.4695	0.7475	12.4999
<i>FSIM</i>	FR	0.8650	0.6805	0.8974	8.2872
<i>IGM</i>	FR	0.8542	0.6693	0.8891	8.6039
<i>GSM</i>	FR	0.8447	0.6605	0.8848	8.7627
<i>GMSD</i>	FR	0.8451	0.6596	0.8847	8.7699
<i>BRISQUE</i>	NR	0.5945	0.4160	0.6104	14.8563
<i>NIQE</i>	NR	0.7688	0.5806	0.8436	19.1263
<i>FSBLIM</i>	NR	0.8565	0.6723	0.8858	8.7796
<i>SISBLIM_{sfb}</i>	NR	0.8534	0.6713	0.8691	9.2745
<i>SISBLIM_{sm}</i>	NR	0.8763	0.6963	0.8988	8.2364
<i>SISBLIM_{wfb}</i>	NR	0.8581	0.6659	0.8711	9.1825
<i>SISBLIM_{wm}</i>	NR	0.8761	0.6946	0.8972	8.2964
<i>RF-MMF</i>	NR	0.8948	0.7235	0.9235	7.2118

Table 3: Results of one sided t -test performed between SROCC values of various IQA algorithms on LIVE multidistortion dataset. A value of '1' denotes that row algorithm is statistically superior to column algorithm, '-1' denotes that the row algorithm is worse than the column algorithm. A value of '0' indicates that the two algorithms are statistically indistinguishable.

	<i>PSNR</i>	<i>SSIM</i>	<i>FSIM</i>	<i>IGM</i>	<i>GSM</i>	<i>GMSD</i>	<i>FSBLIM</i>	<i>SISBLIM_{sfb}</i>	<i>SISBLIM_{sm}</i>	<i>SISBLIM_{wfb}</i>	<i>SISBLIM_{wm}</i>	<i>BRISQUE</i>	<i>NIQE</i>	<i>RF-MMF</i>
<i>PSNR</i>	0	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	-1	-1
<i>SSIM</i>	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	-1	-1
<i>FSIM</i>	1	1	0	1	1	1	1	1	-1	1	-1	1	1	-1
<i>IGM</i>	1	1	-1	0	1	1	-1	-1	-1	-1	-1	1	1	-1
<i>GSM</i>	1	1	-1	-1	0	1	-1	-1	-1	-1	-1	1	1	-1
<i>GMSD</i>	1	1	-1	-1	-1	0	-1	-1	-1	-1	-1	1	1	-1
<i>FSBLIM</i>	1	1	-1	1	1	1	0	1	-1	-1	-1	1	1	-1
<i>SISBLIM_{sfb}</i>	1	1	-1	-1	1	1	-1	0	-1	-1	-1	1	1	-1
<i>SISBLIM_{sm}</i>	1	1	1	1	1	1	1	1	0	1	1	1	1	-1
<i>SISBLIM_{wfb}</i>	1	1	-1	1	1	1	1	1	-1	0	-1	1	1	-1
<i>SISBLIM_{wm}</i>	1	1	1	1	1	1	1	1	-1	1	0	1	1	-1
<i>BRISQUE</i>	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	-1	-1
<i>NIQE</i>	1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	0	-1
<i>RF-MMF</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	0

null hypothesis at the 95% confidence level (Mittal *et al.*, 2012). The proposed technique is statistically better than all the competing metrics for the problem of image quality measures for multiply distorted images in the LIVE multidistortion database which has two categories of images (Category 1: blur + JPEG compression artifacts and Category 2: blur + additive noise).

PERFORMANCE WITH MDID2013 DATABASE

Comparison with human visual system

MDID2013 (Gu *et al.*, 2014) is one of the new databases available for image quality assessment for multiply distorted images. In this dataset blur, JPEG artifacts and additive noise are present in a single image. 324 images with varying levels of noise,

compression artifacts and blur are generated from 27 reference images in this dataset. Similar to LIVE-MD, we have trained a random forest regression model on the 324 images of the MDID2013 dataset in an 80/20 ratio where 80 percent of images are used for training and 20 percent for testing. 1000 different combinations of train/test sets are generated and for these trials, experiments are performed. The median performance scores of 1000 trials are presented in Table 4. For MDID2013, we have also compared our proposed technique against the well known full reference and no reference image quality assessment algorithms and presented the correlation results in Table 4. The proposed measure *RF-MMF* performs better than all the competing standard image quality measures (both full reference and no reference) for the MDID2013 dataset.

Table 4: Performance evaluation of our proposed IQA measure and its comparison with state of the art IQA for MDID2013 dataset for 1000 train-test combination 1000 trials.

Name	Type	Median SROCC	Median KRCC	Median PLCC	Median RMSE
<i>PSNR</i>	FR	0.5557	0.3938	0.5578	0.0416
<i>SSIM</i>	FR	0.4996	0.3532	0.5203	0.0428
<i>FSIM</i>	FR	0.5930	0.4038	0.5874	0.0407
<i>IGM</i>	FR	0.8195	0.6250	0.8207	0.0288
<i>GSM</i>	FR	0.6598	0.4623	0.6513	0.0383
<i>GMSD</i>	FR	0.8273	0.6300	0.8371	0.0275
<i>BRISQUE</i>	NR	0.2187	0.1637	0.1469	0.0496
<i>NIQE</i>	NR	0.5446	0.3819	0.5619	0.0415
<i>FSBLIM</i>	NR	0.7683	0.5714	0.7563	0.0328
<i>SISBLIM_{sfb}</i>	NR	0.6873	0.4823	0.7025	0.0359
<i>SISBLIM_{sm}</i>	NR	0.8051	0.6161	0.8135	0.0292
<i>SISBLIM_{wfb}</i>	NR	0.6905	0.4970	0.7001	0.0362
<i>SISBLIM_{wm}</i>	NR	0.7937	0.6042	0.8007	0.0300
<i>RF-MMF</i>	NR	0.8667	0.6815	0.8840	0.0235

Table 5: Results of one sided t -test performed between SROCC values of various IQA algorithms on MDID2013 dataset. A value of '1' denotes that row algorithm is statistically superior to column algorithm, '-1' denotes that the row algorithm is worse than the column algorithm. A value of '0' indicates that the two algorithms are statistically indistinguishable.

	<i>PSNR</i>	<i>SSIM</i>	<i>FSIM</i>	<i>IGM</i>	<i>GSM</i>	<i>GMSD</i>	<i>FSBLIM</i>	<i>SISBLIM_{sfb}</i>	<i>SISBLIM_{sm}</i>	<i>SISBLIM_{wfb}</i>	<i>SISBLIM_{wm}</i>	<i>BRISQUE</i>	<i>NIQE</i>	<i>RF-MMF</i>
<i>SSIM</i>	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	-1	-1
<i>FSIM</i>	1	1	0	-1	-1	-1	-1	-1	-1	-1	-1	1	1	-1
<i>IGM</i>	1	1	1	0	1	-1	1	1	1	1	1	1	1	-1
<i>GSM</i>	1	1	1	-1	0	-1	-1	-1	-1	-1	-1	1	1	-1
<i>GMSD</i>	1	1	1	1	1	0	1	1	1	1	1	1	1	-1
<i>FSBLIM</i>	1	1	1	-1	1	-1	0	1	-1	1	-1	1	1	-1
<i>SISBLIM_{sfb}</i>	1	1	1	-1	1	-1	-1	0	-1	-1	-1	1	1	-1
<i>SISBLIM_{sm}</i>	1	1	1	-1	1	-1	1	1	0	1	1	1	1	-1
<i>SISBLIM_{wfb}</i>	1	1	1	-1	1	-1	-1	1	-1	0	-1	1	1	-1
<i>SISBLIM_{wm}</i>	1	1	1	-1	1	-1	1	1	-1	1	0	1	1	-1
<i>BRISQUE</i>	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	-1	-1
<i>NIQE</i>	-1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	0	-1
<i>RF-MMF</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	0

Statistical significance and hypothesis testing

Similar to LIVE multi-distortion dataset, statistical significance and hypothesis testing is performed using one-sided t -test with the same null and alternate hypothesis to find the statistical relevance of the difference in SROCC values. The results are presented in Table 5 and the plot of mean SROCC values and standard error bars for each algorithm across 1000 trials for the MDID2013 database is shown in Fig. 6. The proposed technique is statistically better than all the competing IQA techniques for the task of assessing image quality for multiply distorted images in the MDID2013 database. We have compared our proposed scheme against state of the art full reference and no reference image quality assessment algorithms for our study and we observe proposed algorithm outperforms the competing measures in the MDID2013 dataset

which consists of images which are distorted by three types of distortions (blur, additive noise and JPEG compression artifacts) in a single image.

DISCUSSION

In this paper we propose a new technique in which we use random forests, a well-known machine learning technique for fusing multiple image quality assessment measures to perform the task of image quality assessment for images corrupted by more than one type of distortions. Multi-method fusion (Liu *et al.*, 2013a) approaches are an active area of research for performing the task of image quality assessment. We have presented the results to demonstrate the effectiveness of our proposed model. We have used two separate datasets for validating our work. Firstly, we use the LIVE multidistortion dataset (Jayaraman

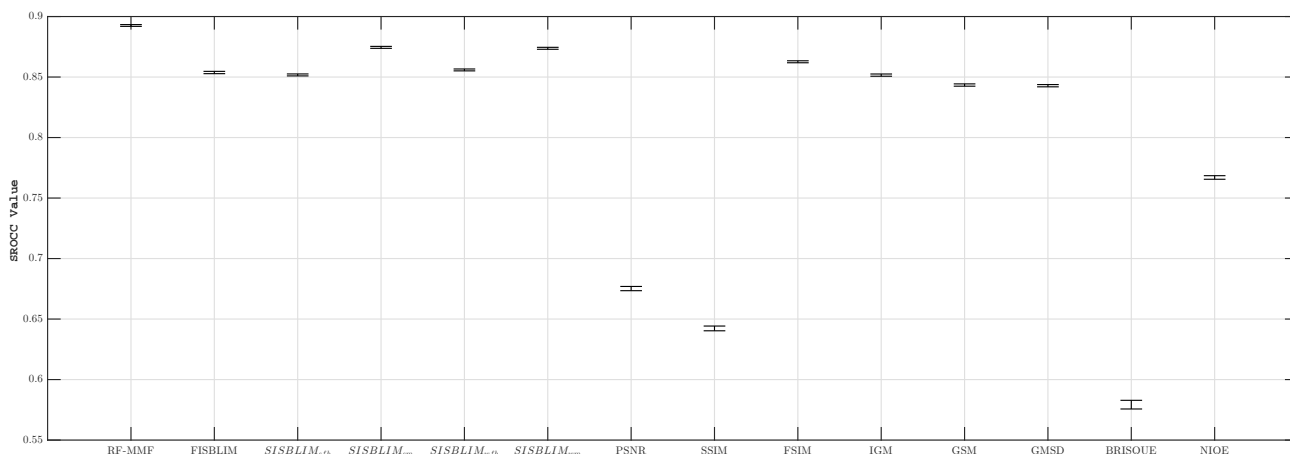


Fig. 5: Mean SROCC and standard error bars of various competing algorithms across 1000 random train-test trials on LIVEMD database.

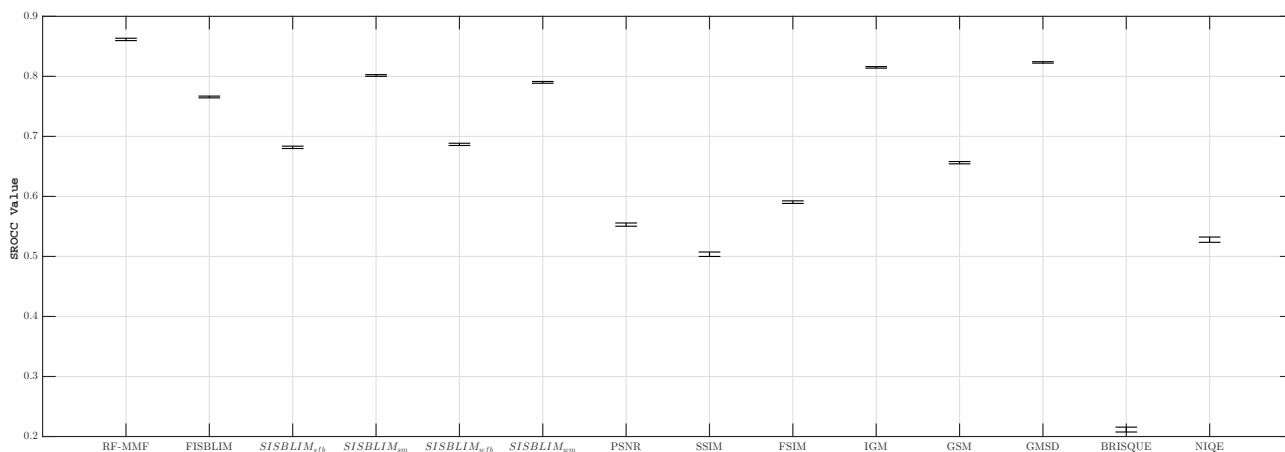


Fig. 6: Mean SROCC and standard error bars of various competing algorithms across 1000 random train-test trials on MDID2013 database.

et al., 2012) which consists of two subcategories of images, images of the first class are distorted by blur followed artifacts of JPEG compression and the second class of images which are distorted by blur followed by additive noise. We have used both categories of images to model our proposed technique and demonstrated that our proposed technique works better than the state of the art image quality assessment techniques. We have used Spearman rank correlation coefficient (*SROCC*) and Kendall rank correlation coefficient (*KRCC*) for validating prediction monotonicity and for showing prediction accuracy we use Pearson linear correlation coefficient (*PLCC*) and Root mean square error (*RMSE*). We have run 1000 independent trials on the database where we have randomly used 80% of the dataset for training the model and 20% for testing.

To remove bias and outliers we have presented the median scores of *SROCC*, *KRCC*, *PLCC* and *RMSE* in Table 2. To assess whether the differences between the median correlation values of different algorithms are statistically relevant or not, we have performed statistical significance and hypothesis testing using one-sided *t*-test across 1000 trials on the *SROCC* values to show the performance of the proposed technique. We have performed the one-sided *t*-test on *SROCC* values of the proposed measure and state of the art measures available in literature across 1000 trials and observed that proposed technique is statistically superior to all the competing measures which can be inferred from Table 3.

We have used the MDID2013 database which

consists of 324 images, where each image is distorted by blur, additive noise and JPEG compression artifacts. We have run 1000 independent trials of the experiment using 80-20 train-test split of the data like we did for the LIVE MD dataset and provided the results in Table 4. The main observation from Table 4 is that for this dataset also the performance of our proposed technique is better than the competing image quality assessment measures. We have also performed the statistical significance and hypothesis testing using one-sided t -test and observed that the proposed technique is statistically superior to most of the standard measures. The proposed technique does not need to modify the image to perform the task of image quality assessment like *SISBLIM* (Gu *et al.*, 2014) and *FISBLIM* (Gu *et al.*, 2013) which removes the noise after calculating noise. Modification of images may lead to the addition of more blur which may reduce the accuracy of image quality assessment.

We have compared the results of our proposed method against well known full reference image quality assessment algorithms, which use a reference image of the same scene to assess the quality of target image. We have compared against techniques like PSNR, SSIM (Wang *et al.*, 2004), FSIM (Zhang *et al.*, 2011), GSM (Liu *et al.*, 2012), IGM (Wu *et al.*, 2013), GMSD (Xue *et al.*, 2014) and observed that it performs better than full reference techniques. We have observed that for both datasets the proposed scheme works better than the Full reference image quality assessment techniques. The proposed scheme is highly efficient as without the use of reference image, it is able to give very good results and able to compete with full reference techniques. Hence, the proposed method can be used in applications where the reference image of the same scene is not available.

CONCLUSION

In this paper, we have proposed a machine learning based image quality measure for multiply distorted datasets. It is a very challenging problem and still it is not solved very accurately. Random forest regression is a type of ensemble technique which uses multiple decision trees for performing classification and regression tasks. Currently, the best-known framework for solving the problem of image quality assessment for multiply distorted images is *SISBLIM*. This framework uses BM3D denoising algorithm to remove noise from the image which is not a desirable as it modifies the image. Image Quality assessment algorithms must try to assess the quality of an image without modifying the image. The proposed

algorithm shows better performance than the *SISBLIM* framework and the four variants of the measure. We have performed separate experiments on *LIVEMD* database and the *MDID2013* database as both the datasets have different types of multiply distorted images and compared our proposed technique against both full reference and no reference image quality assessment algorithms. *LIVEMD* dataset has images with two combinations of distortions (blur + JPEG or blur + additive noise), and *MDID2013* dataset has images with three combinations of distortions (blur + noise + JPEG). The DMOS scores in the two datasets are not in the same range, hence, separate experiments were performed. The problem of multiply distorted image quality assessment is very challenging and one of the challenges faced is the lack of standard datasets available for this problem. Datasets which have a variable number of multiple distortions must be made available publicly with human psychophysics experiment DMOS scores. We have compared the results of our proposed scheme against the state of the art image quality assessment algorithms both full-reference and no-reference techniques and we conclude that the proposed image quality assessment algorithm performs better than most of the standard measures proposed till date for multiply distorted images.

REFERENCES

- Blanchet G, Moisan L (2012). An explicit sharpness index related to global phase coherence. In: Proc 2012 IEEE Int Conf Acoust Speech Signal Process (ICASSP) 1065–8.
- Breiman L (2001). Random forests. *Mach Learn* 45:5–32.
- Dabov K, Foi A, Katkovnik V, Egiazarian K (2007). Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE T Image Process* 16:2080–95.
- De K, Masilamani V (2017a). Image quality assessment for blurred images using nonsubsampling contourlet transform features. *J Computer* 12:156–64.
- De K, Masilamani V (2017b). No-reference image contrast measure using image statistics and random forest. *Multimed Tools Appl* 76:18641.
- Eskicioglu AM, Fisher PS (1995). Image quality measures and their performance. *IEEE T Commun* 43:2959–65.
- Ferzli R, Karam LJ (2009). A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB). *IEEE T Image Process* 18:717–28.
- Gabarda S, Cristóbal G (2007). Blind image quality assessment through anisotropy. *J Opt Soc Am A* 24:B42–51.

- Golestaneh SA, Chandler DM (2014). No-reference quality assessment of JPEG images via a quality relevance map. *IEEE Signal Proc Lett* 21:155–8.
- Gu K, Zhai G, Lin W, Yang X, Zhang W (2015a). No-reference image sharpness assessment in autoregressive parameter space. *IEEE T Image Process* 24:3218–31.
- Gu K, Zhai G, Liu M, Yang X, Zhang W, Sun X, Chen W, Zuo Y (2013). Fisblim: A five-step blind metric for quality assessment of multiply distorted images. In: *Proc 2013 IEEE Worksh Signal Process Syst (SiPS)* 241–6.
- Gu K, Zhai G, Yang X, Zhang W (2014). Hybrid no-reference quality metric for singly and multiply distorted images. *IEEE T Broadcast* 60:555–67.
- Gu K, Zhai G, Yang X, Zhang W (2015b). Using free energy principle for blind image quality assessment. *IEEE T Multimedia* 17:50–63.
- Hassen R, Wang Z, Salama MM, *et al.* (2013). Image sharpness assessment based on local phase coherence. *IEEE T Image Process* 22:2798–810.
- Jaiantilal A (2009). Classification and regression by randomforest-matlab. [Online] Available: <http://code.google.com/p/randomforest-matlab/>.
- Jayaraman D, Mittal A, Moorthy AK, Bovik AC (2012). Objective quality assessment of multiply distorted images. In: *2012 Conf Rec 46th Asilomar Conf Signal Syst Computer (ASILOMAR)* 1693–7.
- Leclaire A, Moisan L (2015). No-reference image quality assessment and blind deblurring with sharpness metrics exploiting fourier phase information. *J Math Imaging Vis* 52:145–72.
- Li C, Bovik AC, Wu X (2011). Blind image quality assessment using a general regression neural network. *IEEE T Neural Networ* 22:793–9.
- Li L, Lin W, Wang X, Yang G, Bahrami K, Kot AC (2016). No-reference image blur assessment based on discrete orthogonal moments. *IEEE T Cybernetics* 46:39–50.
- Liu A, Lin W, Narwaria M (2012). Image quality assessment based on gradient similarity. *IEEE T Image Process* 21:1500–12.
- Liu TJ, Lin W, Kuo CCJ (2013a). Image quality assessment using multi-method fusion. *IEEE T Image Process* 22:1793–07.
- Liu X, Tanaka M, Okutomi M (2013b). Single-image noise level estimation for blind denoising. *IEEE T Image Process* 22:5226–37.
- Marziliano P, Dufaux F, Winkler S, Ebrahimi T (2002). A no-reference perceptual blur metric. In: *Proc 2002 IEEE Int Conf Image Proces* 3:57–60.
- Mittal A, Moorthy AK, Bovik AC (2012). No-reference image quality assessment in the spatial domain. *IEEE T Image Process* 21:4695–08.
- Mittal A, Soundararajan R, Bovik AC (2013). Making a “completely blind” image quality analyzer. *IEEE Signal Proc Lett* 20:209–12.
- Moorthy AK, Bovik AC (2010). A two-step framework for constructing blind image quality indices. *IEEE Signal Proc Lett* 17:513–6.
- Moorthy AK, Bovik AC (2011). Blind image quality assessment: From natural scene statistics to perceptual quality. *IEEE T Image Process* 20:3350–64.
- Narvekar ND, Karam LJ (2011). A no-reference image blur metric based on the cumulative probability of blur detection (CPBD). *IEEE T Image Process* 20:2678–83.
- Pei SC, Chen LH (2015). Image quality assessment using human visual DOG model fused with random forest. *IEEE T Image Process* 24:3282–92.
- Pyatykh S, Hesser J, Zheng L (2013). Image noise level estimation by principal component analysis. *IEEE T Image Process* 22:687–99.
- Rehman A, Wang Z (2012). Reduced-reference image quality assessment by structural similarity estimation. *IEEE T Image Process* 21:3378–89.
- Rohaly AM, Libert J, Corriveau P, Webster A, eds (2000). Final report from the video quality experts group on the validation of objective models of video quality assessment. [Online] Available: <http://www.vqeg.org/>.
- Saad MA, Bovik AC, Charrier C (2012). Blind image quality assessment: A natural scene statistics approach in the DCT domain. *IEEE T Image Process* 21:3339–52.
- Sheikh HR, Bovik AC, Cormack L (2005). No-reference quality assessment using natural scene statistics: JPEG2000. *IEEE T Image Process* 14:1918–27.
- Sheskin DJ (2004). Handbook of parametric and nonparametric statistical procedures, 3rd Ed. Boca Raton: Chapman & Hall/CRC.
- Soundararajan R, Bovik AC (2012). RRED indices: Reduced reference entropic differencing for image quality assessment. *IEEE T Image Process* 21:517–26.
- Vu CT, Phan TD, Chandler DM (2012). S_3 : A spectral and spatial measure of local perceived sharpness in natural images. *IEEE T Image Process* 21:934–45.
- Vu PV, Chandler DM (2012). A fast wavelet-based algorithm for global and local image sharpness estimation. *IEEE Signal Proc Lett* 19:423–6.
- Wang Z, Bovik AC (2006). Modern image quality assessment. San Rafael: Morgan & Claypool.
- Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004). Image quality assessment: from error visibility to structural similarity. *IEEE T Image Process* 13:600–12.
- Wang Z, Sheikh HR, Bovik AC (2002). No-reference perceptual quality assessment of JPEG compressed

- images. In: Proc 2002 IEEE Int Conf Image Process I:477–80.
- Wu J, Lin W, Shi G, Liu A (2013). Perceptual quality metric with internal generative mechanism. *IEEE T Image Process* 22:43–54.
- Xu L, Lin W, Kuo CCJ (2015). *Visual quality assessment by machine learning*. Singapore: Springer.
- Xue W, Zhang L, Mou X, Bovik AC (2014). Gradient magnitude similarity deviation: a highly efficient perceptual image quality index. *IEEE T Image Process* 23:684–95.
- Zhang L, Zhang L, Mou X, Zhang D (2011). FSIM: a feature similarity index for image quality assessment. *IEEE T Image Process* 20:2378–86.
- Zhu X, Milanfar P (2010). Automatic parameter selection for denoising algorithms using a no-reference measure of image content. *IEEE T Image Process* 19:3116–32.
- Zoran D, Weiss Y (2009). Scale invariance and noise in natural images. In: Proc 12th IEEE Int Conf Comput Vision 2209–16.