

# Nonparametric Quality Assessment of Natural Images

Redzuan Abdul Manap  
*University of Sheffield*

Ling Shao  
*Northumbria University*

Alejandro F. Frangi  
*University of Sheffield*

The authors propose a two-stage nonparametric no-reference image quality assessment framework that can predict the image distortion type as well as the quality of local regions. The framework correlates well with human perception of image quality.

As multimedia and visual technologies have continued to advance in recent years, digital images have become ever more ubiquitous. Subsequently, a huge number of publicly available digital images have led to a surge of interest in the image processing and computer vision research areas. One particular area that has received significant research attention is image quality assessment (IQA). While subjective IQA measures are generally agreed to be the most reliable way to assess perceptual image quality, the fact that they are carried out by human observers makes them expensive and time-consuming. As such, an algorithm that can automatically provide an image quality measurement that is consistent with human perceptual measures is highly desired.

Objective IQA algorithms can generally be categorized into two main classes: full-reference (FR) and no-reference (NR). In the FR-IQA category, the quality of a distorted image is evaluated by comparing the entire information difference between the image and its corresponding undistorted reference image. Mean squared error and peak signal-to-noise ratio

(PSNR) are the simplest metrics to be implemented in this case. However, they have poor correlation with subjective quality measures. This has resulted in many other FR-IQA algorithms being developed where image quality is estimated based on various mechanisms such as the human visual system, image structure, or image statistics. SSIM<sup>1</sup> and FSIM<sup>2</sup> are examples of established high-performance FR-IQA algorithms that achieve high correlation with subjective IQA. However, in many situations, full information on the reference image is not available. For example, in photo and film restoration applications, it is possible that a degraded print is the only available record of a photo or a film. In such cases, an NR-IQA algorithm is preferred.

Rather than discovering suitable quality-predictive features, which have been intensively researched using parametric algorithms, our work attempts to look at an alternative framework to perform the NR-IQA task without having to undergo any training process. Following the feature extraction process, we propose a two-stage *nonparametric* NR-IQA framework. Our framework design is based on the observation that parametric NR-IQA algorithms are sensitive to different databases.<sup>3</sup> Once they are trained on one database, most of the algorithms would perform poorly when tested on another database, because they contain database-specific parameters. Considering that nonparametric models are more flexible and make fewer assumptions than their parametric counterparts, using a nonparametric framework should yield better performance across different databases. In addition, previous work<sup>4</sup> also indicates that the distribution of human differential mean opinion score (DMOS) values varies greatly between different distortion classes. Therefore, the introduction of a distortion identification stage in our framework should lead to a better selection of relevant training (labeled) samples to be used in predicting the quality of the test image.

This work is an extension of our previous work,<sup>5</sup> where we conducted experiments on a single IQA database and included only initial experimental results. For this article, we conducted further testing to fully show the potential of using the nonparametric approach to perform NR-IQA. Experimental results on the standard IQA databases demonstrate that the proposed algorithm achieves high correlation with human perceptual measures of image

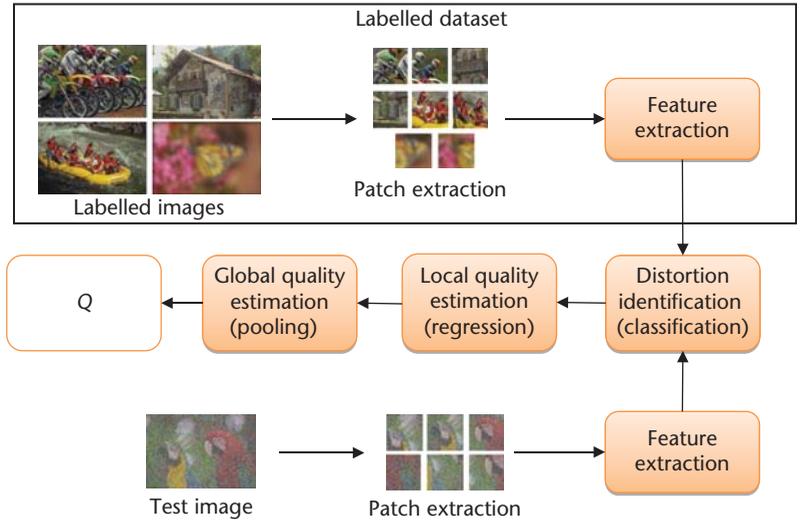
quality and provides comparable performance with state-of-the-art non-distortion-specific (NDS) NR-IQA algorithms.

### Current NR-IQA Algorithms

Present NR-IQA algorithms can be classified into two major categories:<sup>6</sup> distortion-specific (DS) and NDS. In DS cases, the distortion type contained in an image is assumed to be known beforehand. A specific distortion model is then employed to estimate image quality. However, these DS algorithms can be employed only in specific application domains due to this assumption. Meanwhile, NDS NR-IQA algorithms require no prior knowledge of the type of distortion affecting the image. Instead, the image quality score is obtained based on an assumption that the image has similar distortion to images in the standard IQA databases. Using the database image examples, whose human DMOSs or human mean opinion scores (MOSs) are provided, these NDS algorithms are then trained to predict the quality of a given image.

A two-stage framework is usually employed when designing these algorithms: feature extraction followed by learning a regression model from human perceptual measures of training images. In the first stage, the extracted quality-predictive features can be either handcrafted or determined via machine learning approaches. Most of the handcrafted quality-predictive features designed for the NDS NR-IQA task are based on natural scene statistical (NSS) properties. Some NSS-based algorithms had their features derived in image transformation domains, such as BIQL,<sup>7</sup> DIIVINE,<sup>8</sup> and NSS-GS/NSS-TS<sup>9</sup> in the wavelet domain and BLINDS-II<sup>10</sup> in the discrete cosine transform (DCT) domain. To reduce expensive computational costs due to the image transformation procedure, other NSS-based algorithms used features that were extracted in the spatial domain. A well-known example of this approach is BRISQUE.<sup>11</sup>

The NSS-based algorithms can also be differentiated by their types of quality-predictive features. For example, statistical properties of distortion textures, natural image, and blur/noise are used to derive the features for LBIQ.<sup>12</sup> The GMLOG algorithm (explained in detail later) extracts features based on statistical properties of local contrast features.<sup>4</sup> In addition, the magnitude, variance, and entropy of the wavelet coefficients are utilized to design the features for the SRNSS algorithm.<sup>3</sup> Meanwhile,



**Figure 1.** The proposed no-reference image quality assessment (NR-IQA) framework. In addition to estimating the image quality, the framework can also identify image distortion.

other algorithms involve features being learned directly from raw image pixels. This approach was first presented by CORNIA (the Codebook Representation for No-Reference Image Assessment),<sup>13</sup> and its success has led to the introduction of another algorithm that uses convolutional neural networks (CNN).<sup>14</sup> The extracted features are then used to learn the mapping between the feature space and the image quality through a regression algorithm. Kernel-based learning methods are used in most cases—in particular, support vector machine (SVM) and support vector regression (SVR) with linear/radial basis functions. In this case, all these NDS NR-IQA algorithms can be referred to as *parametric* methods.

### Our Proposed Framework

Our proposed nonparametric IQA framework is illustrated in Figure 1. It consists of five major components: local feature extraction, labeled dataset construction, distortion identification, local (patch-level) quality estimation, and pooling for overall (image-level) quality estimation.

#### Local Feature Extraction

Because the features are extracted from local image patches instead of from a whole image, it is essential to use quality-predictive features that have low computational requirements. As such, we chose to use features from the spatial domain that alleviate expensive computation encountered by image-transform-based features. In this

---

**We have adopted two  
local spatial contrast  
features, gradient  
magnitude and Laplacian  
of Gaussian, to perform  
the NR-IQA task.**

---

work, we have adopted two local spatial contrast features, gradient magnitude (GM) and Laplacian of Gaussian (LOG), to perform the NR-IQA task. This is based on the observation that they can characterize image semantic structures such as edges and corners, which in turn are closely related to human perception of image quality. Thus, we chose four joint statistical properties of these features (as implemented elsewhere<sup>4</sup>) as the quality-predictive features to be extracted from the images.

Specifically, the GM map of an image  $\mathbf{I}$  can be computed as

$$\mathbf{G}_I = \sqrt{[\mathbf{I} \otimes \mathbf{h}_x]^2 + [\mathbf{I} \otimes \mathbf{h}_y]^2},$$

where  $\mathbf{h}_x$  and  $\mathbf{h}_y$  are the Gaussian partial derivative filters applied along horizontal and vertical directions respectively. Meanwhile, the LOG of the image is given by

$$\begin{aligned} \mathbf{L}_I &= \mathbf{I} \otimes \mathbf{h}_{\text{LOG}} \\ \mathbf{h}_{\text{LOG}}(x, y | \sigma) &= \frac{\partial^2}{\partial x^2} \mathbf{g}(x, y | \sigma) + \frac{\partial^2}{\partial y^2} \mathbf{g}(x, y | \sigma), \end{aligned}$$

where  $\mathbf{g}(x, y | \sigma)$  is the isotropic Gaussian function with scale parameter  $\sigma$ . The computed GM and LOG operators are then normalized to achieve stable statistical image representations:

$$\bar{\mathbf{G}}_I = \frac{\mathbf{G}_I}{(N_I + \varepsilon)}, \bar{\mathbf{L}}_I = \frac{\mathbf{L}_I}{(N_I + \varepsilon)}. \quad (1)$$

The locally adaptive normalization factor  $N_I$  in Equation 1 is computed at each location  $(i, j)$  as

$$N_I(i, j) = \sqrt{\sum_{(l, k) \in \Omega_{ij}} \omega(l, k) \mathbf{F}_I^2(l, k)},$$

where  $\Omega_{ij}$  is a local window centered at  $(i, j)$ ,  $\omega(l, k)$  are weights, and  $\mathbf{F}_I(i, j) = \sqrt{\mathbf{G}_I^2(i, j) + \mathbf{L}_I^2(i, j)}$ .

The marginal probability functions of the jointly normalized GM and LOG operators,

denoted by  $P_{\bar{\mathbf{G}}_I}$  and  $P_{\bar{\mathbf{L}}_I}$  respectively, are then computed and selected as the first two quality-predictive features:

$$\begin{aligned} P_{\bar{\mathbf{G}}_I}(\bar{\mathbf{G}}_I = g_m) &= \sum_{n=1}^N \mathbf{K}_{m,n}, P_{\bar{\mathbf{L}}_I}(\bar{\mathbf{L}}_I = l_n) \\ &= \sum_{m=1}^M \mathbf{K}_{m,n}, \end{aligned}$$

where  $\mathbf{K}_{m,n} = P(\bar{\mathbf{G}}_I = g_m, \bar{\mathbf{L}}_I = l_n)$  is the joint empirical probability function of  $\bar{\mathbf{G}}_I$  and  $\bar{\mathbf{L}}_I$ , while  $m = 1, \dots, M$  and  $n = 1, \dots, N$  are the quantization levels of  $\bar{\mathbf{G}}_I$  and  $\bar{\mathbf{L}}_I$ .

Considering the fact that there are dependencies between the GM and LOG features, the two remaining quality-predictive features, known as independency distributions, are then computed. They can be represented as

$$\begin{aligned} Q_{\bar{\mathbf{G}}_I}(\bar{\mathbf{G}}_I = g_m) &= \frac{1}{N} \sum_{n=1}^N P(\bar{\mathbf{G}}_I = g_m | \bar{\mathbf{L}}_I = l_n) \\ Q_{\bar{\mathbf{L}}_I}(\bar{\mathbf{L}}_I = l_n) &= \frac{1}{M} \sum_{m=1}^M P(\bar{\mathbf{L}}_I = l_n | \bar{\mathbf{G}}_I = g_m). \end{aligned}$$

These four quality-predictive features are then combined to produce the final GMLOG feature vector for an image:

$$\mathbf{GMLOG}_I = [P_{\bar{\mathbf{G}}_I}, P_{\bar{\mathbf{L}}_I}, Q_{\bar{\mathbf{G}}_I}, Q_{\bar{\mathbf{L}}_I}].$$

### Labeled Dataset Construction

Considering that most parametric NR-IQA algorithms use 80:20 train-test ratios to train their regression models, we follow the same strategy to construct the labeled dataset. In other words, the dataset is constructed based on 80 percent of the randomly sampled reference images and their associated distorted images from a selected standard IQA database. To this end, let the total number of images in the labeled dataset be denoted as  $L_T$ . Given one labeled image, it is first divided into  $L$  nonoverlapped patches of  $B \times B$  size. The GMLOG feature vector is extracted from each of these patches. They are then combined over all labeled images to form the dataset. Consequently, we can represent the size of the feature matrix  $\mathbf{GMLOG}_D$  of the dataset as

$$\mathbf{GMLOG}_D = \left[ \left( \sum_{i=1}^{L_T} \mathbf{L}_i \right) \times 4M \right].$$

In the dataset, two different labels are provided for those selected patches. The first label is the distortion class. Each patch is assigned a label of the distortion type that is affecting its associated source image. The second label is the DMOS where each patch is assigned with its corresponding source images' DMOS. Although

this assignment might be questionable, it is acceptable in this case because the distortion level is uniform across the image. An example of this dataset construction on one reference image and its associated distorted images is shown in Figure 2.

### Image Distortion Identification

The next stage is to identify (classify) the distortion class of the test image. Prior to this, a test image  $\mathbf{Y}$  is first partitioned into  $P$  nonoverlapped patches  $y_i$ . The GMLOG feature vector  $\mathbf{GMLOG}_{y_i}$  is then computed for each  $y_i$ ,  $i = 1, 2, \dots, P$ , before being combined to form the test image feature matrix  $\mathbf{GMLOG}_{\mathbf{Y}}$ .

In nonparametric classification, it has been shown that under the Naïve-Bayes assumption, the optimal distance to be used is the image-to-class (I2C) distance rather than the typically used image-to-image distance. Thus, we have adopted the Naïve-Bayes Nearest Neighbor (NN) algorithm<sup>15</sup> to design the nonparametric classifier in this work.

Based on this algorithm, the predicted class for the test image is found as

$$C_p = \operatorname{argmin}_C \|\mathbf{GMLOG}_{\mathbf{Y}} - NN_C(\mathbf{GMLOG}_{\mathbf{Y}})\|^2,$$

where  $NN_C(\mathbf{GMLOG}_{\mathbf{Y}})$  is the NN-descriptor of  $\mathbf{GMLOG}_{\mathbf{Y}}$  in the distortion class  $C$ .

### Local Quality Estimation

Once we determine the distortion type affecting the test image, we then use the labeled patches within the identified class to estimate the quality of the test image patches. In this work, we employ a typical  $k$ -NN regression algorithm. First, we compute the Euclidean distance between the test patch and the labeled patches. Then we rearrange the labeled patches in ascending order according to the computed distances. We then empirically choose the first  $K$  nearest labeled patches to estimate the test patch quality score. At this point, rather than using a common inverse distance weighting scheme where the selected patches are assigned weights according to the inverse of their computed distances, we estimate the quality score of the test patch through a simple linear regression. In this case, the predicted score is

$$Q_{y_i} = \mathbf{w}(\mathbf{GMLOG}_{y_i}),$$

where  $\mathbf{w}$  is the optimized weights for the test patch feature vector.

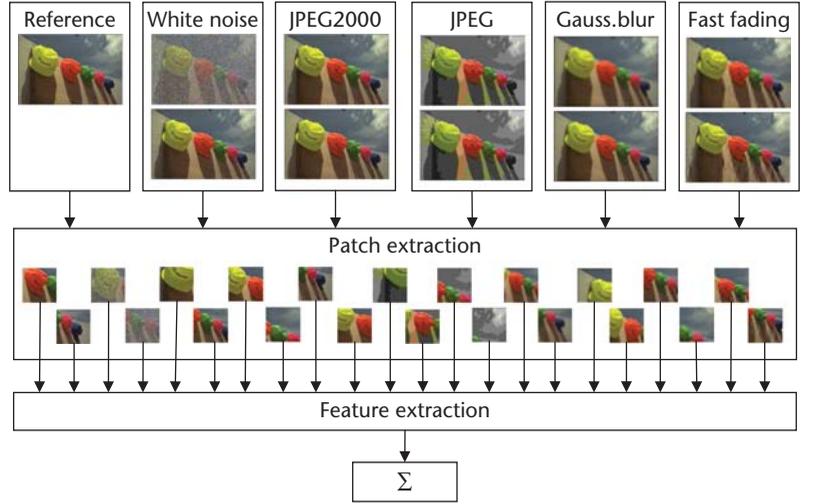


Figure 2. An example of dataset construction based on one reference image and its distorted versions. The dataset consists of feature vectors extracted from patches of labeled images.

### Global Quality Estimation

We can then infer the image-level quality of the test image. In this work, rather than using a simple average pooling, we employ an inverse distance weighting rule where each predicted local score is assigned a weight based on the minimum Euclidean distance  $d_i$  computed in the previous local quality estimation stage. As such, the global quality score for the test image  $\mathbf{Y}$  is given as

$$Q_{\mathbf{Y}} = \frac{\sum_{i=1}^P w_i Q_{y_i}}{\sum_{i=1}^P w_i},$$

where

$$w_i = \left( \sum_{i=1}^P d_i \right) / d_i.$$

## Experiments and Discussion

In this section, we present the experiments, evaluate the performance, and discuss the properties of our algorithm.

### Protocols

The performance of an NR-IQA algorithm is usually evaluated using subjective image databases. There are several established subjective image evaluation databases within the IQA research area. For this work, we used two publicly available databases, LIVE<sup>16</sup> (developed by the Laboratory for Image & Video Engineering

**Table 1. Overall performance for the non-distortion-specific experiment. (The top three NR-IQA algorithms appear in bold.)**

Algorithm	LIVE		CSIQ	
	SROCC	LCC	SROCC	LCC
PSNR	0.8659	0.8561	0.9292	0.8562
SSIM	0.9126	0.9064	0.9362	0.9347
FSIM	0.9639	0.9602	0.9629	0.9675
BIQI	0.8204	0.8200	0.7598	0.8353
DIIVINE	0.9156	0.9166	0.8697	0.9010
BLIINDS-II	0.9312	0.9296	0.9003	0.9282
BRISQUE	0.9400	<b>0.9418</b>	<b>0.9085</b>	<b>0.9356</b>
GMLOG	<b>0.9511</b>	<b>0.9551</b>	<b>0.9243</b>	<b>0.9457</b>
CORNIA	<b>0.9416</b>	0.9347	0.8845	0.9241
Proposed	<b>0.9408</b>	<b>0.9414</b>	<b>0.9384</b>	<b>0.9535</b>

at the University of Texas at Austin) and CSIQ<sup>17</sup> (developed by the Laboratory of Computational and Subjective Image Quality at Shizuoka University). The LIVE database is probably the most widely used database for evaluating the performance of IQA algorithms. It consists of 29 undistorted reference images. Each of these reference images is then subjected to five or six degradation levels in five different distortion types: JPEG2000 compression (JP2K), JPEG compression (JPEG), additive white noise (WN), Gaussian blur (GB), and simulated fast fading channel (FF), yielding a total of 779 distorted images. These distorted images are provided with DMOS values in the range between 0 and 100. Meanwhile, the CSIQ database comprises 866 distorted images. They are generated when a total of six different types of distortions are applied to 30 reference images at four or five levels. In contrast to the LIVE database, each distorted image is assigned a DMOS value between 0 and 1. In both databases, an image with a lower distortion level is assigned with a lower DMOS value.

The scale parameter  $\sigma$  is set at 0.5, while the quantization level  $M = N$  is set at 10.<sup>4</sup> The patch size  $B$  and the number of NN labeled patches  $K$  are empirically set at 96 and 1,000, respectively.

The performance of the NR-IQA algorithms is measured by their ability to predict image quality as close as possible to human visual system performance. These algorithms commonly use two metrics that measure the consistency between the predicted quality score of the

image and its corresponding DMOS/MOS: the Spearman rank order correlation coefficient (SROCC) and the linear correlation coefficient (LCC). The SROCC is used to represent algorithm prediction monotonicity, while the LCC is used to evaluate the prediction accuracy of the algorithm. For both SROCC and LCC metrics, a correlation score that is close to 1 (or  $-1$ ) indicates good performance by the algorithm.

### Evaluation on LIVE Database

We compared our proposed algorithm to three FR-IQA algorithms: PSNR, SSIM,<sup>1</sup> and FSIM.<sup>2</sup> We also chose six recent NDS NR-IQA algorithms for comparison, where their codes are publicly available. The six algorithms chosen were Blind Image Quality Indices (BIQI),<sup>7</sup> Distortion Identification-based Image Verity and INtegrity Evaluation (DIIVINE),<sup>8</sup> Blind Image Integrity notator using DCT Statistics-II (BLIINDS-II),<sup>10</sup> Blind/Referenceless Image Spatial QUALity Evaluator (BRISQUE),<sup>11</sup> Gradient Magnitude and Laplacian of Gaussian (GMLOG),<sup>4</sup> and CORNIA.<sup>13</sup> These algorithms' databases are partitioned into two parts: 80 percent of the reference images and their distorted versions are randomly selected as a training set, and the remaining 20 percent is for testing, thus ensuring there is no overlap. In our case, we used the same training set to construct the required labeled dataset, and we used LIBSVM<sup>18</sup> to perform regression for these algorithms. For a fair comparison, we determined their respective regression parameters through cross-validation in accordance with their published research papers.

We conducted two experiments: an NDS experiment and a DS experiment. In the NDS experiment, we performed the train-test (labeled-test, in our research) run across all distorted images regardless of their distortions. In the DS experiment, we conducted the run on a single type of distortion to evaluate how well the algorithm performs in one particular distortion. We repeated the train-test procedure 1,000 times; the median results are reported in Tables 1 and 2. For brevity, Table 2 shows only the SROCC results for the DS experiment. Similar conclusions can be made for LCC results. The top three NR-IQA algorithms are highlighted in bold.

For the NDS experiment, our framework clearly outperforms BIQI, DIIVINE, and BLIINDS-II when tested on both the LIVE and CSIQ databases. In addition, it also achieves similar performance as BRISQUE and CORNIA

Table 2. Overall performance for the distortion-specific experiment. (The top NR-IQA algorithms appear in bold.)

Algorithm	LIVE					CSIQ			
	JP2K	JPEG	WN	GB	FF	JP2K	JPEG	WN	GB
PSNR	0.8954	0.8809	0.9854	0.7823	0.8907	0.9363	0.8882	0.9363	0.9289
SSIM	0.9614	0.9764	0.9694	0.9517	0.9556	0.9606	0.9546	0.8974	0.9609
FSIM	0.9724	0.9840	0.9716	0.9708	0.9519	0.9704	0.9664	0.9359	0.9729
BIQI	0.7989	0.8911	0.9507	0.8457	0.7073	0.7573	0.8384	0.6000	0.8160
DIIVINE	0.9128	0.9096	<b>0.9837</b>	0.9212	0.8632	0.8692	0.8843	0.8131	0.8756
BLIINDS-II	<b>0.9288</b>	0.9420	0.9687	0.9232	0.8886	0.8870	0.9115	0.8863	<b>0.9152</b>
BRISQUE	0.9135	<b>0.9645</b>	0.9789	<b>0.9509</b>	0.8774	0.8934	<b>0.9253</b>	<b>0.9310</b>	<b>0.9143</b>
GMLOG	<b>0.9283</b>	<b>0.9659</b>	<b>0.9849</b>	0.9395	<b>0.9008</b>	<b>0.9172</b>	<b>0.9328</b>	<b>0.9406</b>	0.9070
CORNIA	0.9271	<b>0.9437</b>	0.9608	<b>0.9553</b>	<b>0.9103</b>	<b>0.8950</b>	0.8845	0.7980	0.9006
Proposed	0.9342	0.9412	0.9853	0.9433	0.8910	0.9395	0.9314	0.9591	0.9230

Table 3. The Spearman rank order correlation coefficient (SROCC) values and the linear correlation coefficient (LCC) values for different patch sizes.

Size	16	32	48	64	80	96	112	128
SROCC	0.5184	0.8162	0.9271	0.9367	0.9370	0.9408	0.9368	0.9387
LCC	0.5733	0.8131	0.9283	0.9376	0.9386	0.9414	0.9366	0.9379

while approaching the performance of state-of-the-art GMLOG on the LIVE database. However, when tested on the CSIQ database, our framework has better prediction performance than all of the competing NR-IQA algorithms. These results support our intuition that using a nonparametric framework can work better across different databases. This also indicates that our framework is robust and has good generalization capability. When compared to the FR-IQA algorithms, our framework also outperforms PSNR and SSIM, and approaches FSIM.

Meanwhile, for the DS experiment, our framework has the best prediction performance for images affected by JP2K and WN distortions on the LIVE database. It is also among the top three NR-IQA algorithms for GB and FF cases while giving comparable performance for JPEG. When tested on the CSIQ database, our framework performs the best for JP2K, WN, and GB cases and comes second for JPEG. This is due to the fact that our prediction performance depends on what types of features are being used. Because we are using statistical features as in the GMLOG algorithm, the prediction patterns for both our framework and GMLOG are similar over the two databases. Different algo-

rithms' features could be used in our framework to achieve better performance in other distortion classes.

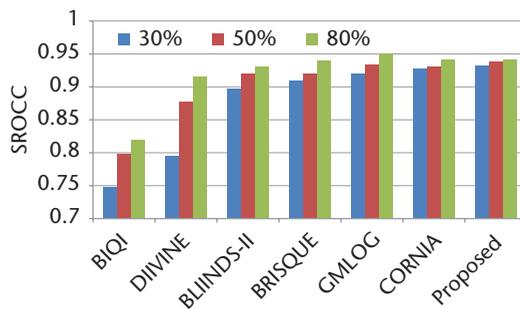
#### Effects of Algorithm Parameters

Because the patches are sampled in a nonoverlapping way, the number of patches for each image is directly affected by the patch size. Table 3 shows the changes in performance with respect to patch size while fixing the labeled images at 80 percent ratios. In general, a larger patch size results in better performance; top performance is achieved when the patch size is set at 96. There is no significant difference in performance when the patch size is increased more than 96. Meanwhile, to investigate the effect of varying the number of images in the labeled dataset on the performance of the proposed framework, we partitioned the databases under three different settings: we used 80 percent, 50 percent, and 30 percent of the images to construct the labeled dataset while using the remaining images for testing. Similarly, we evaluated all the other six competing NR-IQA algorithms under the same settings.

Table 4 and Figure 3 show the SROCC results for the NDS experiment under various training

**Table 4. SROCC comparison for different training (labeled) sample ratios. (The top performing algorithms appear in bold.)**

<b>LIVE database</b>	Ratio	BIQI	DIIVINE	BLIINDS-II	BRISQUE	GMLOG	CORNIA	Proposed
	30%	0.7484	0.7954	0.8973	0.9094	0.9208	0.9277	<b>0.9320</b>
	50%	0.7993	0.8768	0.9198	0.9213	0.9343	0.9314	<b>0.9375</b>
	80%	0.8204	0.9156	0.9312	0.9400	<b>0.9511</b>	0.9416	0.9408
<b>CSIQ database</b>	30%	0.6721	0.7838	0.8465	0.8628	0.8949	0.8605	<b>0.9143</b>
	50%	0.7208	0.8246	0.8832	0.8857	0.9109	0.8706	<b>0.9295</b>
	80%	0.7598	0.8697	0.9003	0.9085	0.9243	0.8845	<b>0.9384</b>



**Figure 3. SROCC comparison over different training ratios on the LIVE database. The proposed algorithm performs well, particularly in small sample cases.**

(labeled) ratios. As expected, the performance of all competing algorithms decreased as the number of samples was reduced. On both databases, our framework constantly performed better than the competing algorithms at the 30 percent and 50 percent rates. At the 80 percent rate, it also performed the best on the CSIQ database but yielded slightly lower SROCC values than CORNIA and GMLOG on the LIVE database. Thus, we can say that our framework works better in situations where the number of samples is small. Note that the top performing algorithms in Table 4 appear in bold.

### Computational Complexity

The computation time required by our algorithm to estimate image quality in a typical  $512 \times 768$  test image is dominated by three major processes: feature extraction, I2C distance computation, and quality estimation. The feature extraction stage is the most time consuming part of the algorithm. Because the features are to be extracted locally at the patch level, a higher number of test patches will lead to a longer feature extraction time. However, by employing a non-overlap sampling strategy and increasing the patch size, we can reduce

the number of test patches. Using the parameter setting described earlier, about 0.09 seconds is required to extract the GMLOG features for the whole set of test image patches.

There is a clear tradeoff between prediction performance and I2C distance computation. As indicated in Table 4, a larger dataset size leads to better prediction accuracy. However, a longer computation time is required to compute the I2C distance between the test patches and the labeled patches. Using the 80 percent training (labeled) rate, another 0.04 second is needed to compute the I2C distances for the test patches in one test image during the distortion identification stage. Finally, an extra 0.06 second is also needed to perform regression for local quality estimation.

In all, the overall computation time required by the proposed algorithm to compute image quality estimation for one  $512 \times 768$  test image is about 0.19 second. This is achieved using unoptimized Matlab code on a computer with an Intel i5 2.60 GHz processor. We did not consider the dataset construction time here because it was already constructed prior to the testing stage. Table 5 compares the average runtimes of the competing algorithms. Although BIQI is the fastest, it has the worst prediction accuracy compared to the other algorithms. BLIINDS-II is the slowest, followed by DIIVINE and CORNIA respectively. While the proposed algorithm is slower than GMLOG and BRISQUE, it can still process up to five images per second, thus providing an alternative solution to real-time IQA applications.

The fact that our proposed framework correlates well with human perceptual measures of image quality across various kinds of image distortions and performs comparably to other algorithms is encouraging, given that our

Table 5. Average runtimes (seconds) of the algorithms we tested.

	BIQI	DIIVINE	BLINDS-II	BRISQUE	GMLOG	CORNIA	Proposed
Runtime	0.08	28.20	95.24	0.18	0.10	2.43	0.19

proposed framework need not undergo any prior training or learning phase, as parametric NR-IQA models require.

For our future work, we can take further steps to improve the performance of the proposed framework. First, we can use saliency detection to guide the patch sampling process in the framework. We can first generate a visual saliency map that weighs the importance of the image's local patches to the human perceptual measures of image quality and then use it to select appropriate patches for the test image. Second, obtaining accurate distortion classes of the test patches can also help us select better candidates to be used for regression in the quality estimation stage. As such, other I2C-based classifiers can also be tested for better classification accuracy. Third, we can also consider integration of a nonparametric incremental learning technique to construct the labeled dataset when dealing with an increasing number of new distortion classes. Finally, similar to most of the previous NR-IQA methods, our current work only focuses on images degraded by a single type of distortion. Encouraged by the promising results, we plan to extend our framework to include images with mixed distortions. **MM**

## References

1. Z. Wang et al., "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, 2004, pp. 600–612.
2. L. Zhang, X. Mou, and D. Zhang, "FSIM: A Feature Similarity Index for Image Quality Assessment," *IEEE Trans. Image Processing*, vol. 20, no. 8, 2011, pp. 2378–2386.
3. L. He et al., "Sparse Representation for Blind Image Quality Assessment," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 1146–1153.
4. W. Xue et al., "Blind Image Quality Assessment Using Joint Statistics of Gradient Magnitude and Laplacian Features," *IEEE Trans. Image Processing*, vol. 23, no. 11, 2014, pp. 4850–4862.
5. R.A. Manap, L. Shao, and A.F. Frangi, "A Nonparametric Framework for No-Reference Image Quality Assessment," *Proc. IEEE Global Conf. Signal and Information Processing*, 2015, pp. 562–566.
6. R.A. Manap and L. Shao, "Non-Distortion-Specific No-Reference Image Quality Assessment: A Survey," *Information Sciences*, vol. 301, Apr. 2015, pp. 141–160.
7. 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, 2010, pp. 513–516.
8. A.K. Moorthy and A.C. Bovik, "Blind Image Quality Assessment: From Natural Scene Statistics to Perceptual Quality," *IEEE Trans. Image Processing*, vol. 20, no. 12, 2011, pp. 3350–3364.
9. X. Gao et al., "Universal Blind Image Quality Assessment Metrics via Natural Scene Statistics and Multiple Kernel Learning," *IEEE Trans. Neural Network Learning Systems*, vol. 24, no. 12, 2013, pp. 2013–2026.
10. M.A. Saad, A.C. Bovik, and C. Charrier, "Blind Image Quality Assessment: A Natural Scene Statistics Approach in the DCT Domain," *IEEE Trans. Image Processing*, vol. 21, no. 8, 2012, pp. 3339–3352.
11. A. Mittal, A.K. Moorthy, and A.C. Bovik, "No-Reference Image Quality Assessment in the Spatial Domain," *IEEE Trans. Image Processing*, vol. 21, no. 12, 2012, pp. 4695–4708.
12. H. Tang, N. Joshi, and A. Kapoor, "Learning a Blind Measure of Perceptual Image Quality," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 305–312.
13. P. Ye et al., "Unsupervised Feature Learning Framework for No-Reference Image Quality Assessment," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 1098–1105.
14. L. Kang et al., "Convolutional Neural Networks for No-Reference Image Quality Assessment," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 1733–1740.
15. O. Boiman, E. Shechtman, and M. Irani, "In Defense of Nearest-Neighbor Based Image Classification," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2008, pp. 1–8.
16. H.R. Sheikh et al., *LIVE Image Quality Assessment Database Release 2*; <http://live.ece.utexas.edu/research/quality>.

17. E.C. Larson and D.M. Chandler, "Most Apparent Distortion: Full-Reference Image Quality Assessment and the Role of Strategy," *J. Electronic Imaging*, vol. 19, no. 1, 2010, pp. 011006.1–011006.21.
18. C.-C. Chang and C.-J. Lin, "LIBSVM: A Library for Support Vector Machines," *ACM Trans. Intelligent Systems and Technology*, vol. 2, no. 3, 2011, pp. 27:1–27:27.

**Redzuan Abdul Manap** is a lecturer in the Faculty of Electronic and Computer Engineering at Universiti Teknikal Malaysia Melaka (UTeM), Malaysia. Currently, he is on a study leave pursuing his PhD at the University of Sheffield under Malaysian Ministry of Higher Education scholarship awards. His research interests include image processing, visual perception, machine learning, and computer vision. He is an IEEE member. Contact him at redzuan@utem.edu.my or rabdulmanap1@sheffield.ac.uk.

**Ling Shao** is a professor at Northumbria University and an Advanced Visiting Fellow at the University of

Sheffield. His research interests include computer vision, pattern recognition, and machine learning. He is a Fellow of the British Computer Society, a Fellow of the Institution of Engineering and Technology, and a senior member of IEEE. Contact him at ling.shao@ieee.org.

**Alejandro F. Frangi** is a professor of biomedical image computing at the University of Sheffield and director of the Centre for Computational Imaging and Simulation Technologies in Biomedicine. His main research interests are in medical image computing, medical imaging, and image-based computational physiology. He is a Fellow of IEEE and a member of SPIE and SIAM. Contact him at a.frangi@sheffield.ac.uk.

 Selected CS articles and columns are also available for free at <http://ComputingNow.computer.org>.

## ADVERTISER INFORMATION

### Advertising Personnel

Marian Anderson: Sr. Advertising Coordinator  
Email: manderson@computer.org  
Phone: +1 714 816 2139 | Fax: +1 714 821 4010

Sandy Brown: Sr. Business Development Mgr.  
Email: sbrown@computer.org  
Phone: +1 714 816 2144 | Fax: +1 714 821 4010

### Advertising Sales Representatives (display)

Central, Northwest, Far East:  
Eric Kincaid  
Email: e.kincaid@computer.org  
Phone: +1 214 673 3742  
Fax: +1 888 886 8599

Northeast, Midwest, Europe, Middle East:  
Ann & David Schissler  
Email: a.schissler@computer.org, d.schissler@computer.org  
Phone: +1 508 394 4026  
Fax: +1 508 394 1707

Southwest, California:  
Mike Hughes  
Email: mikehughes@computer.org  
Phone: +1 805 529 6790

Southeast:  
Heather Buonadies  
Email: h.buonadies@computer.org  
Phone: +1 973 304 4123  
Fax: +1 973 585 7071

### Advertising Sales Representatives (Classified Line)

Heather Buonadies  
Email: h.buonadies@computer.org  
Phone: +1 973 304 4123  
Fax: +1 973 585 7071

### Advertising Sales Representatives (Jobs Board)

Heather Buonadies  
Email: h.buonadies@computer.org  
Phone: +1 973 304 4123  
Fax: +1 973 585 7071