
Article:
Manap, R.A., Shao, L. and Frangi, A.F. orcid.org/0000-0002-2675-528X (2016) Nonparametric Quality Assessment of Natural Images. IEEE Multimedia, 23 (4). pp. 22-30. ISSN 1070-986X

https://doi.org/10.1109/MMUL.2016.2

Non-Parametric Quality Assessment of Natural Images

Redzuan Abdul Manap, Department of Electronic and Electrical Engineering, University of Sheffield, UK
Ling Shao, Department of Computer Science and Digital Technologies, Northumbria University, UK
Alejandro F. Frangi, Department of Electronic and Electrical Engineering, University of Sheffield, UK

Abstract

In this paper, we attempt to explore an alternative way to perform the no-reference image quality assessment (NR-IQA) task. Following a feature extraction stage in which spatial domain statistics are utilized as our features, a two-stage non-parametric NR-IQA framework is proposed. No training phase is required and it also enables prediction of the image distortion type as well as local regions' quality, which is not available in most of the previous algorithms. Experimental results on the IQA databases show that the proposed framework achieves high correlation to human perception of image quality and delivers competitive performance to state-of-the-art NR-IQA algorithms.

Keywords – image processing and computer vision; image quality assessment; non-parametric classification and regression

I. Introduction

As multimedia and visual technologies keep advancing in recent years, so does the presence of digital images in our life. 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 as the most reliable judgement in assessing 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 measurement of an image quality 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 with its corresponding undistorted reference image. Mean squared error (MSE) 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 results in many other FR-IQA algorithms being developed where the image quality is estimated based on various mechanisms such as human visual system, image structure or image statistics. SSIM [1] and FSIM [2] are examples of established high performance FR-IQA algorithms. Higher correlation with the subjective assessment of image quality is achieved by these FR-IQA

algorithms. However, in many situations, full information of 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 case, an NR-IQA algorithm is preferred.

Present NR-IQA algorithms can be further classified into two major categories [3]: distortion-specific (DS) and non-distortion-specific (NDS). In the DS cases, the distortion type contained in an image is assumed to be known beforehand. A specific distortion model is then employed to estimate the quality of the image. However, these DS algorithms can only be employed in specific application domains due to this assumption. Meanwhile, no prior knowledge of the type of distortion affecting the image is required by the NDS NR-IQA algorithms. 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 differential mean opinion scores (DMOS) or human mean opinion scores (MOS) 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 BIQI [4], DIIVINE [5] and NSS-GS/NSS-TS [6] in the wavelet domain while BLIINDS-II [7] in the DCT domain. To reduce expensive computational costs due to the image transformation procedure, other NSS-based algorithms utilized features that are extracted in the spatial domain. Well known example of this approach is BRISQUE [8].

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 [9]. In [10], GMLOG extracts features based on statistical properties of local contrast features. In addition, the magnitude, the variance and the entropy of the wavelet coefficients are utilized to design the features for SRNSS [11]. Meanwhile, other algorithms propose their features to be learned directly from raw image pixels. The approach is first presented by CORNIA [12] and its success leads to the introduction of other algorithm, CNN [13]. 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 termed as parametric methods.

Rather than discovering suitable quality predictive features, which have been intensively researched by these 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, a two-stage non-parametric NR-IQA framework is proposed. At the first stage of the framework, the distortion type that degrades the test image is identified. Based on the intuition that images that are affected by the same type of distortion should have similar quality predictive feature properties, a nearest-neighbour (NN) based classifier is utilized to determine the distortion class. This is done by employing Image-to-Class (I2C) distance computation between test image patches and a set of labelled patches. Once the distortion class is identified, the quality of the test image at patch level is then predicted

through k-NN regression that utilizes the DMOS values associated with the labelled patches within the identified class. The predicted scores from all test patches are then combined together to yield the image-level quality score of the test image.

Our framework design is based on the observation that the parametric NR-IQA algorithms are sensitive to different databases [11]. Once they are trained on one database, most of the algorithms would perform poorly when tested on another database. This is because they contain database-specific parameters. Considering that the non-parametric models are more flexible and make less assumption than their parametric counterparts, the use of a non-parametric framework should yield better performance across different databases. In addition, previous work in [10] also indicates that the distribution of the DMOS values is greatly varied between different distortion classes. Therefore, the introduction of a distortion identification stage in our framework should lead to a better selection of relevant training (labelled) samples to be used in predicting the quality of the test image.

The proposed framework has the following advantages. First, no training phase is required. One major disadvantage of the previous parametric methods is that re-training of regression parameters is required when samples of new distortion types are added to the training set making them impractical for on-line learning. In this framework, no training is necessary as the new samples can be simply added into the labelled dataset. Second, over-fitting of parameters, which can be an issue with parametric approaches, is also avoided. In addition, it also enables the prediction of distortion type and image quality of local regions, which is not available in most of the previous algorithms.

This work is the extension of our previous work presented in [14]. The previous experiments were conducted on a single IQA database and only initial experimental results were included. In this paper, further testing is conducted to fully show the potential of non-parametric approach to performing the NR-IQA task. Experimental results on the standard IQA databases demonstrate that the proposed algorithm achieves high correlation to human perceptual measures of image quality and provides comparable performance with state-of-the-art NDS NR-IQA algorithms.

The remainder of this paper is structured as follows. The proposed non-parametric framework of the algorithm is presented in detail in Section II. In Section III, experimental results and subsequent analysis are presented before the paper is concluded in Section IV.

II. Non-Parametric Framework for IQA

The proposed framework is illustrated in Figure 1. It consists of five major components which are: 1) Local feature extraction; 2) Labelled dataset construction; 3) Distortion identification; 4) Local (patch-level) quality estimation, and 5) Pooling for overall (image-level) quality estimation.

A. Local feature extraction

Since 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 choose to use features from the spatial domain which alleviate expensive computation encountered by image transform based features. In this work,

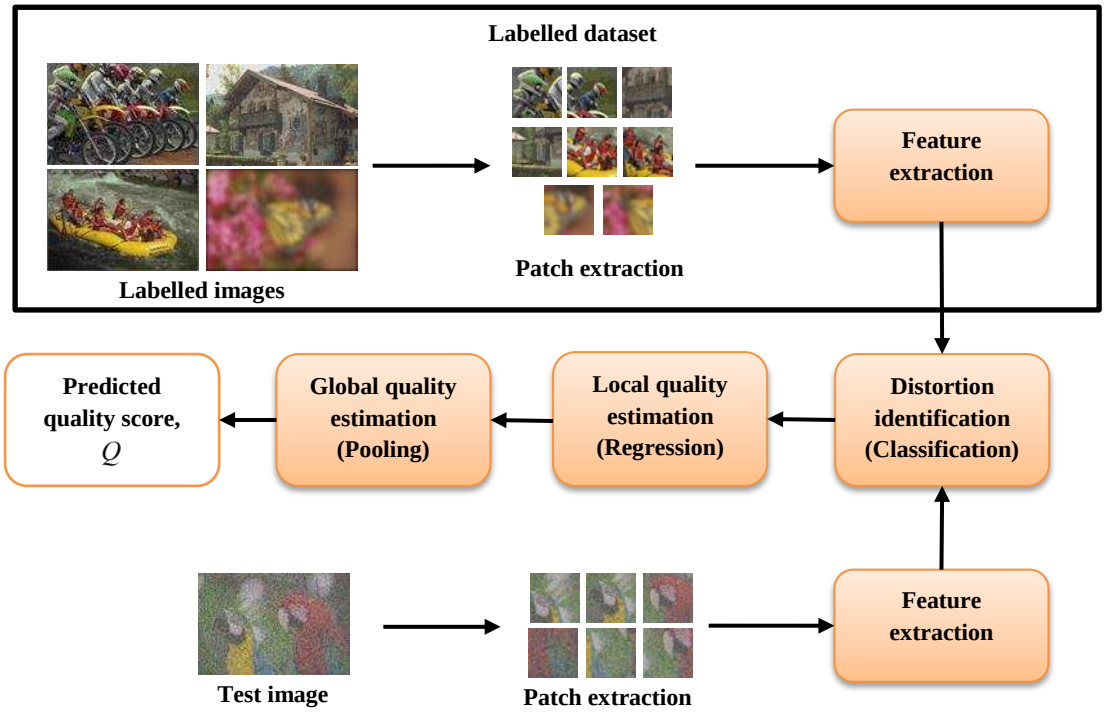

Figure 1. Proposed NR-IQA framework.

two local spatial contrast features: gradient magnitude (GM) and Laplacian of Gaussian (LOG), are adopted to perform the NR-IQA task. It 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. As such, four joint statistical properties of these features as implemented in [10] are chosen as the quality predictive features to be extracted from the images.

Specifically, the GM map of an image **I** can be computed as:

$$\mathbf{G_I} = \sqrt{[\mathbf{I} \otimes \mathbf{h}_x]^2 + [\mathbf{I} \otimes \mathbf{h}_y]^2}\,, \tag{1}$$

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:

$$\mathbf{L_I} = \mathbf{I} \otimes \mathbf{h}_{\mathrm{LOG}}\,, \tag{2}$$

$$\mathbf{h}_{\mathrm{LOG}}(x, y|\sigma) = \frac{\partial^2}{\partial x^2}\mathbf{g}(x, y|\sigma) + \frac{\partial^2}{\partial y^2}\mathbf{g}(x, y|\sigma)\,, \tag{3}$$

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

$$\bar{\mathbf{G}}_{\mathbf{I}} = \frac{\mathbf{G}_{\mathbf{I}}}{(\mathbf{N}_{\mathbf{I}} + \varepsilon)}, \bar{\mathbf{L}}_{\mathbf{I}} = \frac{\mathbf{L}_{\mathbf{I}}}{(\mathbf{N}_{\mathbf{I}} + \varepsilon)} \ . \tag{4}$$

The locally adaptive normalization factor $\mathbf{N}_{\mathbf{I}}$ in Equation (4) is computed at each location (i, j) as:

$$\mathbf{N}_{\mathbf{I}}(i, j) = \sqrt{\sum\sum_{(l,k)\in\Omega_{i,j}} \omega(l,k)\mathbf{F}_{\mathbf{I}}^{2}(l,k)}, \tag{5}$$

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

The marginal probability functions of the jointly normalized GM and LOG operators, denoted by $P_{\bar{\mathbf{G}}_{\mathbf{I}}}$ and $P_{\bar{\mathbf{L}}_{\mathbf{I}}}$ respectively, are then computed and selected as the first two quality predictive features.

$$P_{\bar{\mathbf{G}}_{\mathbf{I}}}(\bar{\mathbf{G}}_{\mathbf{I}} = g_m) = \sum_{n=1}^{N} \mathbf{K}_{m,n} \ , P_{\bar{\mathbf{L}}_{\mathbf{I}}}(\bar{\mathbf{L}}_{\mathbf{I}} = l_n) = \sum_{m=1}^{M} \mathbf{K}_{m,n} \tag{6}$$

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

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:

$$Q_{\bar{\mathbf{G}}_{\mathbf{I}}}(\bar{\mathbf{G}}_{\mathbf{I}} = g_m) = \frac{1}{N}\sum_{n=1}^{N} P(\bar{\mathbf{G}}_{\mathbf{I}} = g_m | \bar{\mathbf{L}}_{\mathbf{I}} = l_n), \tag{7}$$

$$Q_{\bar{\mathbf{L}}_{\mathbf{I}}}(\bar{\mathbf{L}}_{\mathbf{I}} = l_n) = \frac{1}{M}\sum_{m=1}^{M} P(\bar{\mathbf{L}}_{\mathbf{I}} = l_n | \bar{\mathbf{G}}_{\mathbf{I}} = g_m). \tag{8}$$

These four quality predictive features are then combined to produce the final feature vector for an image. Further details of these statistical features can be found in [10].

B. Labelled dataset construction

Considering that most of the parametric NR-IQA algorithms use 80:20 train-test ratios to train their regression model, the same strategy is followed to construct the labelled dataset. In other words, the dataset is constructed based on 80% 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 labelled dataset be denoted as L_T. Given one labelled image, it is first divided into L non-overlapped patches of $B \times B$ size. The GMLOG feature vector is

extracted from each of these patches. They are then combined over all labelled images to form the dataset. Consequently, the size of the feature matrix $\mathbf{GMLOG_D}$ of the dataset can be represented as:

$$\mathbf{GMLOG_D} = \left[\left(\sum_{i=1}^{L_T} \mathbf{L_i}\right) \times 4M\right]. \tag{9}$$

In the dataset, two different labels are provided for those selected patches. The first label is the distortion class. Each patch is assigned with a label of the distortion type that affecting its associated source image. The second label is the DMOS where each patch is assigned with its corresponding source images' DMOS. Though this assignment may be questionable, it is acceptable in this case as 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.

C. 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 non-overlapped patches y_i. The GMLOG feature vector $\mathbf{GMLOG}_{y_i}$ is then computed for each $y_i, i = 1,2,\ldots,P$ before being combined to form the test image feature matrix $\mathbf{GMLOG_Y}$.

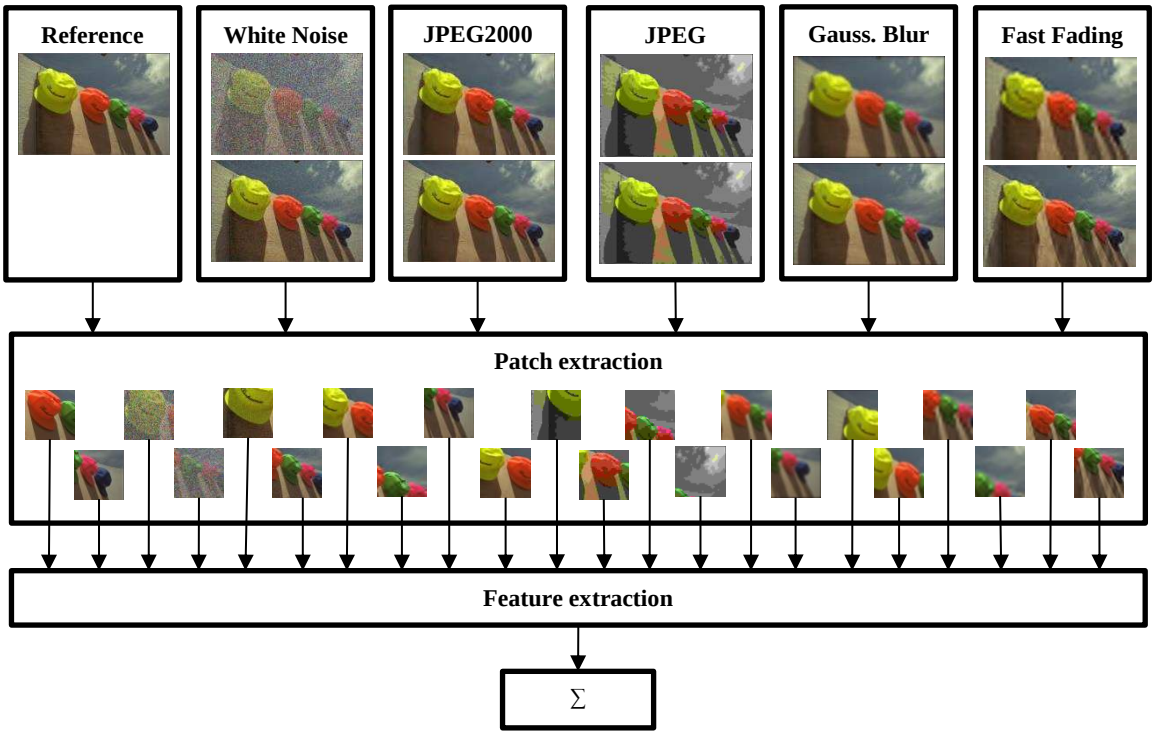

Figure 2. An example of dataset construction based on one reference image and its distorted versions.

In non-parametric classification, it has been shown that under the Naïve-Bayes assumption, the optimal distance to be used is I2C distance rather than the usually used image-to-image (I2I) distance. Thus, the Naïve-Bayes Nearest Neighbour (NBNN) algorithm introduced in [15] is adopted to design the non-parametric classifier in this work. Based on this algorithm, the predicted class for the test image is found as:

$$C_p = \arg\min_C \left\| \mathbf{GMLOG_Y} - NN_C\left(\mathbf{GMLOG_Y}\right) \right\|^2, \tag{10}$$

where $NN_C\left(\mathbf{GMLOG_Y}\right)$ is the NN-descriptor of $\mathbf{GMLOG_Y}$ in the distortion class C.

D. Local quality estimation

Once the distortion type affecting the test image is determined, the labelled patches within the identified class are then used to estimate the quality of the test image patches. In this work, a typical k-NN regression algorithm is employed. The Euclidean distance between the test patch and the labelled patches is first computed. The labelled patches are then rearranged in ascending order according to the computed distances. The first K nearest labelled patches are then empirically chosen 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, the quality score of the test patch is estimated through a simple linear regression. In this case, the predicted score is:

$$Q_{y_i} = \mathbf{w}\left(\mathbf{GMLOG}_{y_i}\right). \tag{11}$$

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

E. Global quality estimation

The image-level quality of the test image can then be inferred. In this work, rather than using a simple average pooling, an inverse distance weighting rule is employed where each of the 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:

$$Q_{\mathbf{Y}} = \frac{\sum_{i=1}^{P} w_i Q_{y_i}}{\sum_{i=1}^{P} w_i}, \tag{13}$$

with

$$w_i = \left(\sum_{i=1}^{P} d_i\right) \Big/ d_i . \tag{14}$$

III. Experiments and Discussion

A. Protocols

Databases: The performance of an NR-IQA algorithm is usually evaluated by using subjective image databases. There are several established subjective image evaluation databases within the IQA research area. In this work, two publicly available databases are utilized: LIVE [16] and CSIQ [17]. The LIVE database is probably the most widely used database in evaluating the performance of IQA algorithms. It consists of 29 undistorted reference images. Each of these reference images is then subjected to 5 to 6 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 which are in the range between 0 and 100. Meanwhile, the CSIQ database is composed of 866 distorted images. They are generated when a total of 6 different types of distortions are applied to 30 reference images at 4 to 5 levels. In contrast to the LIVE database, each distorted image is assigned with a DMOS value in the range between 0 and 1. In both databases, an image with a lower distortion level is assigned with a lower DMOS value.

Parameter setting: The scale parameter σ is set at 0.5 while the quantization level M = N is 10 as in [10]. The patch size, B and the number of NN labelled patches, K is empirically set at 96 and 1000 respectively.

Performance metrics: The performance of the NR-IQA algorithms is measured by their ability to predict image quality as close as possible to HVS. Two metrics that measure the consistency between the predicted quality score of the image and its corresponding DMOS/MOS are commonly used by these algorithms: the Spearman rank order correlation coefficient (SROCC) and the linear correlation coefficient (LCC). The SROCC is used to represent the 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.

B. Evaluation on LIVE database

The proposed algorithm is compared to three FR-IQA algorithms: PSNR, SSIM and FSIM. Six recent NDS NR-IQA algorithms: BIQI, DIIVINE, BLIINDS-II, BRISQUE, GMLOG, and CORNIA are also selected for comparison where their codes are publicly available. For these algorithms, the databases are partitioned into two parts. 80% of the reference images and their distorted versions are randomly selected as a training set. The remaining 20% is for testing, thus ensuring there is no overlap between them. In our case, the same training set is used to construct the required labelled dataset. LIBSVM [18] is used to perform regression for these algorithms. For fair comparison, their regression parameters are determined through cross validation in accordance to their papers.

Two experiments are conducted: NDS experiment and DS experiment. In the NDS experiment, the train-test (labelled-test, in this work) run is performed across all distorted images regardless their distortions. In the DS experiment, the run is conducted on a single type of distortion to evaluate how well the algorithm performs in one particular distortion. The train-test procedure is repeated 1000 times and the median results are reported in Tables I

Table I. Overall Performance for NDS Experiment

	LIVE		CSIQ	
Algorithm	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	**0.9418**	**0.9085**	**0.9356**
GMLOG	**0.9511**	**0.9551**	**0.9243**	**0.9457**
CORNIA	**0.9416**	0.9347	0.8845	0.9241
Proposed	**0.9408**	**0.9414**	**0.9384**	**0.9535**

Table II. Overall Performance for DS Experiment

	LIVE					CSIQ			
Algorithm	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	**0.9837**	0.9212	0.8632	0.8692	0.8843	0.8131	0.8756
BLIINDS-II	**0.9288**	0.9420	0.9687	0.9232	0.8886	0.8870	0.9115	0.8863	**0.9152**
BRISQUE	0.9135	**0.9645**	0.9789	**0.9509**	0.8774	0.8934	**0.9253**	**0.9310**	**0.9143**
GMLOG	**0.9283**	**0.9659**	**0.9849**	0.9395	**0.9008**	**0.9172**	**0.9328**	**0.9406**	0.9070
CORNIA	0.9271	**0.9437**	0.9608	**0.9553**	**0.9103**	**0.8950**	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**

and II. For brevity, only the SROCC results for the DS experiment are shown. Similar conclusions can be made for LCC results. Note that 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 LIVE and CSIQ databases. In addition, it also achieves similar performance as BRISQUE and CORNIA while approaching 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 the use of a non-parametric 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 approaching 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 while comes second for JPEG. This is due to the fact that our prediction performance depends on what types of features are being used. Since we are using statistical features as in GMLOG algorithm, it can be seen that the prediction patterns for both our framework and GMLOG are similar over the two databases. Different algorithms' features could be used in our framework to achieve better performance in other distortion classes.

C. Effects of algorithm parameters

Since the patches are sampled in a non-overlapping way, the number of patches for each image is directly affected by the patch size. The changes of performance with respect to patch size while fixing the labelled images at 80% ratios are shown in Table III. In general, a larger patch size result in better performance with the 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 labelled dataset to the performance of the proposed framework, the databases are partitioned under three different settings: 80%, 50%, and 30% of the images are used to construct the labelled dataset while the remaining is used for testing. Similarly, all the other six competing NR-IQA algorithms are also evaluated under the same settings.

The SROCC results for the NDS experiment under various training (labelled) ratios are shown in Table IV and Figure 3. As expected, the performance of all competing algorithms decreases as the number of samples is reduced. On both databases, our framework constantly performs better than the competing algorithms at the 30% and 50% rates. At 80% rate, it also performs the best on the CSIQ database while yields slightly lower SROCC value than

Table III. SROCC and 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

Table IV. SROCC Comparison for Different Training (Labelled) Samples Ratios

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

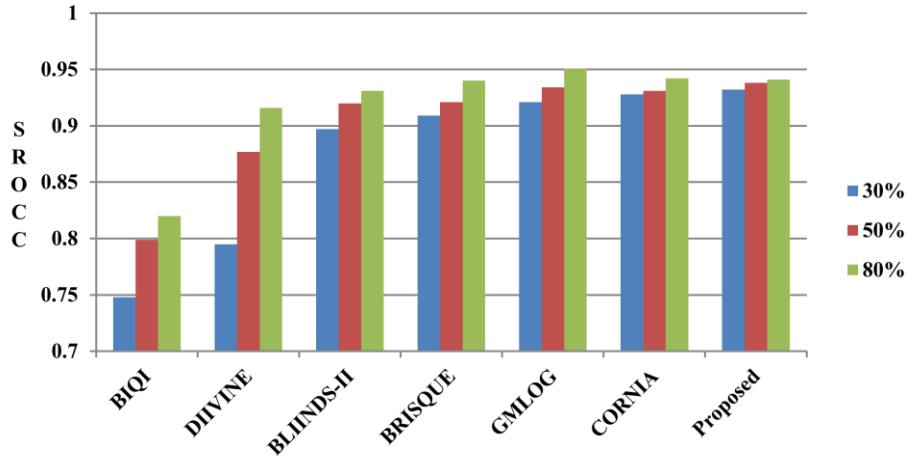

Figure 3. SROCC comparison over different training ratios on LIVE database.

CORNIA and GMLOG on the LIVE database. Thus, it can be said that our framework has better robustness to the number of training (labelled) samples and it works better in situations where the number of samples are small.

D. Computational complexity

In this sub-chapter, the computational complexity of the proposed algorithm is analysed. The computation time required by our algorithm to estimate the image quality in a typical 512×768 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. Since the features are to be extracted locally at the patch level, higher number of test patches will lead to longer feature extraction time. However, by employing non-overlap sampling strategy and increasing the patch size, the number of test patches can be reduced. Using the parameter setting described in Section III.A, about 0.09 seconds is required to extract the GMLOG features for the whole test image patches.

It can be seen that there is a clear trade-off between the prediction performance and the I2C distance computation. As indicated in Table IV, larger dataset size leads to better prediction accuracy. However, longer computation time is required to compute the I2C distance between the test patches and the labelled patches. Using the 80% training (labelled) rate, another 0.04 seconds is needed to compute the I2C distances for test patches in one test image during the distortion identification stage. Finally, an extra 0.06 seconds is also needed to perform regression for the local quality estimation.

In all, the overall computation time required by the proposed algorithm to compute image quality estimation for one 512×768 test image is about 0.19 seconds. This is achieved using an un-optimised MATLAB code on a computer with an Intel i5 2.60 GHz processor. The dataset construction time is not considered here as it is already constructed prior to the testing stage. Comparison of the average run-time of the competing algorithms is shown in

Table V. Average Run-time (Second)

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

Table V. Although BIQI is the fastest, it has the worst prediction accuracy compared to other algorithms. BLIINDS-II is the slowest, followed by DIIVINE and CORNIA respectively. While the proposed algorithm is slower than M3 and BRISQUE, it still can process up to 5 images per second thus providing alternative solution to real-time IQA applications.

IV. Conclusion and Future Work

In this paper, a simple yet effective NDS NR-IQA algorithm has been presented where a non-parametric framework consisting of an I2C distance computation between the test image patches and a labelled dataset as well as k-NN regression is employed to predict the quality of the test image. Experimental results on standard IQA databases indicate that the proposed framework has high correlation with human perceptual measure of image quality across various kinds of image distortions and produces comparable performance with recent algorithms. This is encouraging enough taking into account that our proposed work does not have to undergo any prior training or learning phase as required by the parametric NR-IQA models.

For our future work, there are further steps can be taken to improve the performance of the proposed framework. First, saliency detection can be used to guide the patch sampling process in the framework. A visual saliency map that weights the importance of the image local patches to the human perceptual measures of an image quality can be first generated and then used to select appropriate patches for the test image. Second, obtaining accurate distortion class of the test patches can also help selecting 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, an integration of a non-parametric incremental learning technique in constructing the labelled dataset can also be considered 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.

References

[1] 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., vol. 13, no. 4, pp. 600-612, Apr. 2004.

[2] L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," IEEE Trans. Image Process., vol. 20, no. 8, pp. 2378-2386, Aug. 2011.

[3] R. A. Manap and L. Shao, "Non-distortion-specific no-reference image quality assessment: A survey," Information Sciences, vol. 301, pp. 141-160, Apr. 2015.

[4] 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.

[5] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: from natural scene statistics to perceptual quality," IEEE Trans. Image Process., vol. 20, no. 12, pp. 3350-3364, Dec. 2011.

[6] X. Gao, F. Gao, D. Tao, and X. Li, "Universal blind image quality assessment metrics via natural scene statistics and multiple kernel learning," IEEE Trans. Neural Network Learn. Syst., vol. 24, no. 12, pp. 2013-2026, Dec. 2013.
[7] 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 Process., vol. 21, no. 8, pp. 3339-3352, Aug. 2012.
[8] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," IEEE Trans. Image Process., vol. 21, no. 12, pp. 4695-4708, Dec. 2012.
[9] H. Tang, N. Joshi, and A. Kapoor, "Learning a blind measure of perceptual image quality," in Proc. IEEE CVPR, Colorado Springs, Colorado, June 2011, pp. 305-312.
[10] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features," IEEE Trans. Image Process., vol. 23, no. 11, pp. 4850-4862, Nov. 2014.
[11] L. He, D. Tao, X. Li, and X. Gao, "Sparse representation for blind image quality assessment," in Proc. IEEE CVPR, Providence, Rhode Island, June 2012, pp. 1146-1153.
[12] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in Proc. IEEE CVPR, Providence, Rhode Island, June 2012, pp. 1098-1105.
[13] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in Proc. IEEE CVPR, Columbus, Ohio, June 2014, pp. 1733-1740.
[14] R. A. Manap, L. Shao, and A. F. Frangi, "A non-parametric framework for no-reference image quality assessment," to be published in Proc. IEEE Global SIP, Orlando, Florida, Dec. 2015.
[15] O. Boiman, E. Shechtman, and M. Irani, "In defense of nearest-neighbor based image classification," in Proc. IEEE CVPR, Anchorage, Alaska, June 2008, pp. 1-8.
[16] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, LIVE Image Quality Assessment Database Release 2 [Online]. Available: http://live.ece.utexas.edu/research/quality
[17] 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," Adv. Modern Radio Electron., vol. 10, no. 4, pp. 30-45, 2009.
[18] C. –C. Chang and C. –J. Lin, "LIBSVM: A library for support vector machines," ACM Trans. Intel. Syst. Tech., vol. 2, pp. 27:1-27:27, May 2011.

Redzuan Abdul Manap is a Lecturer at Faculty of Electronic and Computer Engineering, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia. Currently, he is on study leave pursuing his PhD degree 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.
redzuan@utem.edu.my, rabdulmanap1@sheffield.ac.uk

Ling Shao currently is Chair in Computer Vision and Machine Intelligence 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 IET and a Senior Member of the IEEE.
ling.shao@northumbria.ac.uk

Alejandro F. Frangi is a Professor of Biomedical Image Computing at the University of Sheffield. He is also the Director of the Centre for Computational Imaging and Simulation Technologies in Biomedicine (CISTIB). His main research interests are in medical image computing, medical imaging and image-based computational physiology. He is a Fellow of the IEEE, SPIE member and SIAM member.
a.frangi@sheffield.ac.uk