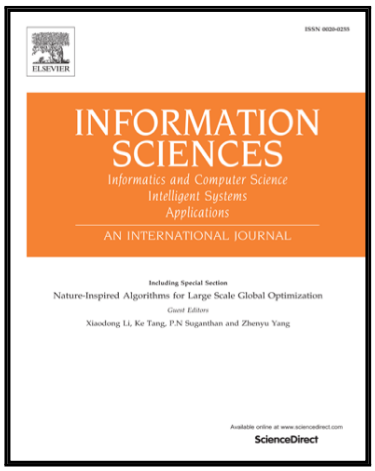

PATCH-IQ: A patch based learning framework for blind image quality assessment

Redzuan Abdul Manap , Ling Shao , Alejandro F. Frangi

PII: S0020-0255(17)30917-9
DOI: 10.1016/j.ins.2017.08.080
Reference: INS 13085

To appear in: *Information Sciences*

Received date: 28 March 2017
Revised date: 24 August 2017
Accepted date: 25 August 2017

Please cite this article as: Redzuan Abdul Manap , Ling Shao , Alejandro F. Frangi , PATCH-IQ: A patch based learning framework for blind image quality assessment, *Information Sciences* (2017), doi: 10.1016/j.ins.2017.08.080

PATCH-IQ: A PATCH BASED LEARNING FRAMEWORK FOR BLIND IMAGE QUALITY ASSESSMENT

Redzuan Abdul Manap[1, 3], Ling Shao[2] and Alejandro F. Frangi[1]

[1]Centre for Computational Imaging and Simulation Technologies in Biomedicine (CISTIB), Department of Electronic and Electrical Engineering, University of Sheffield, Sheffield, S1 3JD, UK

[2]School of Computing Sciences, University of East Anglia, Norwich, NR4 7TJ, UK

[3]Department of Electronic and Computer Engineering, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, Durian Tunggal, Melaka, 76100, Malaysia

Corresponding Author: Ling Shao. Email: ling.shao@ieee.org. Phone: +44 (0)1603592603

ABSTRACT

Most well-known blind image quality assessment (BIQA) models usually follow a two-stage framework whereby various types of features are first extracted and used as an input to a regressor. The regression algorithm is used to model human perceptual measures based on a training set of distorted images. However, this approach requires an intensive training phase to optimise the regression parameters. In this paper, we overcome this limitation by proposing an alternative BIQA model that predicts image quality using nearest neighbour methods which have virtually zero training cost. The model, termed PATCH based blind Image Quality assessment (PATCH-IQ), has a learning framework that operates at the patch level. This enables PATCH-IQ to provide not only a global image quality estimation but also a local image quality estimation. Based on the assumption that the perceived quality of a distorted image will be best predicted by features drawn from images with the same distortion class, PATCH-IQ also introduces a distortion identification stage in its framework. This enables PATCH-IQ to identify the distortion affecting the image, a property that can be useful for further local processing stages. PATCH-IQ is evaluated on the standard IQA databases, and the provided scores are highly correlated to human perception of image quality. It also delivers competitive prediction accuracy and computational performance in relationship to other state-of-the-art BIQA models.

KEYWORDS

Image quality assessment, blind image quality assessment, interest point detection, spatial domain features, nearest neighbour classification and regression.

1. INTRODUCTION

Image quality assessment (IQA) aims at quantifying the quality of natural images with objective quality metrics. For multimedia applications which the end user is a human consumer, IQA anticipates through such metrics the image quality as perceived by human observers. IQA metrics based on human perception are often considered as the gold standard for perceptual assessment of image quality. These subjective metrics are commonly obtained by conducting image quality experiments where participating human observers rate the quality of images presented to them. The ratings are then averaged across all observers yielding a mean opinion score (MOS) or differential mean opinion score (DMOS). The MOS/DMOS constitutes a subjective metric of perceived image quality. However, these subjective metrics must involve human observers makes them expensive, time-consuming, and unfeasible for deployment in most real world applications. An IQA model that can automatically provide objective image quality measurement consistent with human perceptual measures (MOS/DMOS) is preferred.

Objective IQA models can generally be classified into three categories [16]: full-reference IQA (FR-IQA), reduced-reference IQA (RR-IQA) and blind IQA (BIQA). FR-IQA models evaluate the quality of a natural image by comparing the entire information difference between the image and its reference image. A reference image refers to a similar image considered distortion-free and of perfect quality. The simplest FR-IQA metrics to be used are mean squared error (MSE) and peak signal-to-noise ratio (PSNR). However, they do not correlate well with human perceptual measures [30]. Several improved FR-IQA models were then proposed based on various mechanism such as human visual system (HVS) [2], image structure [9],[32],[40], or image statistics [5]. RR-IQA models do not require full information of the reference image. A set of parameters relevant to visual perception of image quality are first selected from the reference image before being used with

the test image to estimate its quality. Well-known RR-IQA models include RRED [27], RR-SSIM [23], and OSVP [33]. High correlation with human perceptual measures are obtained by these FR-IQA and RR-IQA models. But full or partial reference image information may not be available in some applications. In such cases, a BIQA model that needs no reference information is more favourable.

BIQA models can be further classified into two main categories [18]: distortion-specific (DS) models and general-purpose models. DS BIQA models use specific distortion model to estimate quality based on an assumption that the distortion affecting the image is known beforehand [6],[15],[39]. However, these models can only be employed in specific application domains wherein the specific degradation is meaningful but cannot be used in a more general setting without substantial redesign. On the other side, no prior knowledge of the distortion affecting the image is required in general-purpose BIQA models. Instead, image quality is derived solely assuming the image is degraded by the same distortion mechanism that affects a database of image exemplars. Such image exemplars can be obtained from standard IQA databases such as LIVE [25], CSIQ [13] and TID2008 [22]. Using such exemplars and their provided MOS / DMOS values, the models are then trained to predict the MOS / DMOS of the image.

Some readers may notice the terms ‘quality’ and ‘distortion’ being used. They are difficult to specify unambiguously as there is no universal or formal definition to them. For this paper, we use the terms as follows: ‘Image quality’ is the integrated perception of the overall degree of excellence of an image [7] while ‘image distortion’ refers to any degradation to the appearance of an image that occurs during the image’s acquisition, communication or processing systems [31]. The perception of image quality and image distortion differ depending on the application. This work focuses on an image communication system. In this context, the quality of an image is associated to how well the image is acquired, processed and communicated over the transmission network. In a typical image communication system, the image usually undergoes the processes of acquisition, compression and transmission before being presented to the end users. These processes may introduce many distortions to the image. For examples, the acquisition step may introduce blurring and/or noise artefacts, the compression step may generate JPEG compression artefacts while the transmission step may introduce noise and/or packet loss artefacts. This work only deals with distortions that are normally encountered in a typical image communication system.

The remainder of the paper is organised as follows. In Section 2, we review previous approaches in developing BIQA models and explain the motivation behind our proposed model. Section 3 then describes the framework of our model. In Section 4, experimental setup and results are presented followed by later analysis. The paper is then concluded in Section 5.

2. PREVIOUS BIQA MODELS AND MOTIVATION

The majority of previous general-purpose BIQA models focus on extracting features that carry discriminative information about image quality. Most models employ handcrafted features designed based on the natural scene statistics (NSS) approach. NSS based models assume that certain statistical properties of natural images will be changed with the presence of distortions and the image perceptual quality can be inferred by appropriately quantifying the changes. The models can be differentiated by the features used. For instance, BIQI [20] and DIIVINE [21] employ features derived from wavelet transforms while BLIINDS-II [24] uses DCT features. Other models, such as BRISQUE [19], GMLOG [34] and DESIQUE [41] utilise features obtained in the spatial domain while CBIQ [35] uses Gabor features. Features based on the natural colour statistics (NCS) are also explored in [29]. Recently, NFEQM [38] and NFERM [8] derive their features based on free energy principle.

Meanwhile, several BIQA models use features learned directly from raw image pixels. The first work using this approach was proposed by CORNIA [36]. Its promising performance leads to other variations such as supervised CORNIA [37] and CNN [12]. The extracted features are then used as an input to regression algorithms to learn the mapping between the features’ space and the image quality score space. Kernel-based learning methods are used usually. Support vector machine (SVM) and support vector regression (SVR) with linear/radial basis function are frequently used to this effect.

Features employed by these models are generally invariant to distortion. High prediction performances correlated with human perceptual measures are reported by these models when tested on various types of distortions in standard IQA databases. Our model, however, tackles the BIQA task from an alternative angle.

Rather than designing new quality predictive features, we concentrate on designing a learning framework that needs no explicit training phase, which is often required by previous BIQA models. The proposed model, dubbed PATCH based Image Quality assessment (PATCH-IQ), is based on a five-stage framework. Given a natural image, PATCH-IQ first samples local patches at the locations of the image's interest points. At the second stage, it then extracts spatial domain BIQA features from those patches. A distortion identification process is next performed at the third stage based on assumption that the perceived quality of a distorted image will be best predicted by features drawn from images with the same distortion class. PATCH-IQ uses a nearest neighbour classifier to perform such a task. The patches correspond to the identified distortion class are then utilised in the fourth stage to predict local image quality. This is done via a k-nearest neighbour regression that associates the local image quality with the DMOS of the annotated patches constrained to the identified distortion class. Finally, an overall image quality score is derived by pooling the local scores of all patches in the image.

PATCH-IQ brings these key properties to IQA. First, PATCH-IQ predicts an image quality directly from a set of annotated patches using a nearest neighbour technique. This is motivated by the fact that the cost of learning for the technique is virtually zero where its training process only involves storing the feature vectors and the labels of the patches. While previous BIQA models require re-training when images with new distortion types are introduced, these images (viz its patches) types can be simply added to the set alleviating the need of explicit training phase. Second, most of the previous models accumulate statistical features over the entire image. Therefore, they can only provide a global estimate of image quality. By first operating at patch level, PATCH-IQ can provide local image quality prediction and a global one. Third, PATCH-IQ can also identify the distortion affecting the image, a property that is not available in most of the previous BIQA models. The last two properties can be useful for further local processing stages such as filtering, restoration or enhancement.

The proposed model is motivated by promising results achieved by our previous BIQA model NPNO [17]. However, there are two substantial differences. First, PATCH-IQ employs an interest points-based patch sampling strategy as opposed to the previous non-overlapping sampling strategy. Second, PATCH-IQ also utilises different spatial domain features to perform distortion identification and BIQA tasks. In addition, the previous work was tested only on IQA databases with single distortion with limited results and analysis. Further experiments and analysis including tests on multiple distorted images are included in this paper to further demonstrate the capability of the proposed model.

3. FRAMEWORK FOR DISTORTION IDENTIFICATION AND BLIND IMAGE QUALITY ASSESSMENT

Figure 1 illustrates the proposed framework for PATCH-IQ. There are 5 main stages: 1) Patch extraction; 2) Feature extraction; 3) Distortion identification; 4) Local (patch) quality estimation; and 5) Global (image) quality estimation. These five stages are described as follows.

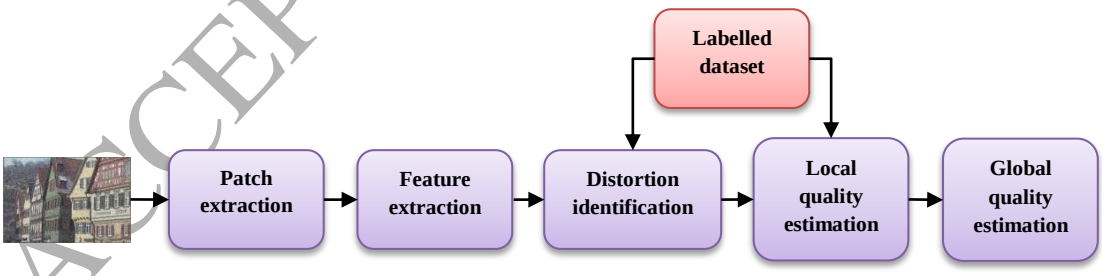

Figure 1: Proposed PATCH-IQ framework

3.1 Patch Extraction

A patch sampling strategy based on interest points of an image is considered. Interest points of an image generally refer to points detected in the image to simplify further processing in a vision system. They are normally at regions of interest, the regions within an image with high information content [28]. The main application of interest points in computer vision and image processing field is to find points / regions in the image domain likely to represent objects. Therefore, they are often employed in processing tasks such as object recognition and image matching. PATCH-IQ tries to extend interest points' application to a BIQA task. It has

been shown that most of the time human focus on object-like regions, i.e. the regions around interest points when looking at an image [11]. In that respect, PATCH-IQ assumes that any distortion applied to those regions will carry greater impact on how human perceived image quality than the distortion in any other image regions such as background. By first finding the location of interest points in an image, patches that contain more relevant information on perceptual image quality can be identified and selected. PATCH-IQ achieves this by using an interest point detector.

A wide variety of interest point detectors exist in the literature. In this work, SIFT [14] is chosen due to its simplicity and good performance. SIFT takes an image and transforms it into a large collection of local feature vectors containing descriptors that are useful to identify objects in an image. There are 4 stages involved in SIFT: 1) Scale-space extrema detection; 2) Keypoint localisation; 3) Orientation assignment; and 4) Keypoint descriptor. The first two stages aim at identifying the locations of stable keypoints at which image features / descriptors will be extracted. The third stage assigns consistent orientation to these keypoints based on local image properties while the last stage uses local gradient information to create the descriptors. The resulting SIFT descriptors may not be useful in estimating image quality. PATCH-IQ, however, does not require the use of SIFT descriptors. Instead, it only utilises the first two stages of SIFT to help find the locations at which patches will be extracted. Based on the above assumption that the regions surrounding the keypoints contain greater information on image quality, PATCH-IQ samples patches of size using the provided keypoints' coordinates as centres. One may argue that an image affected by distortion can give lots of false keypoints as edges lose sharpness. These false keypoints obviously are not useful for object recognition or detection purposes. For quality assessment, these keypoints are still useful since, usually, the whole image is distorted. The extracted image patches still carry information on image quality. An example of this process is shown in Figure 2. Note that PATCH-IQ only extracts patches at the identified keypoint locations. If there is no keypoint detected at any particular image area, no patch is extracted at that area.

Figure 2: Patch extraction using interest point sampling strategy

3.2 Feature Extraction

Two main factors affect the choice of features. First, it is crucial to employ features with low computational requirements since they are to be extracted at patch level. PATCH-IQ utilises spatial domain features to avoid expensive computation normally encountered by image transform-based features. Second, the selected features should carry information not only on perceptual quality but on the distortion in the image. The same features as implemented in the BRISQUE model [19] are adopted by PATCH-IQ. Specifically PATCH-IQ utilises the empirical distributions of locally normalised luminance coefficients and pairwise products of these coefficients to design 18 statistical features for both tasks. Given an image patch , its locally normalised luminance is first obtained by computing local mean subtraction and divisive normalisation at each location ( ) :

$$\hat{}(\quad) \quad \frac{(\quad) \quad (\quad)}{(\quad)}, \tag{1}$$

where the local mean field ( ) is defined as:

$$(\quad) \quad \sum \quad \sum \quad (\quad), \tag{2}$$

and the local variance field ( ) is given by:

$$(\quad) \quad \sqrt{\sum \quad \sum \quad (\quad (\quad) \quad (\quad))} \, . \tag{3}$$

In these equations, and are spatial indices with and being the patch height and width respectively. Here, is a constant to prevent the denominator in Equation (1) from falling to zero

while is a Gaussian weighting function sampled with 3 standard deviations and rescaled to unit sum and is the function window size. The empirical distribution of these coefficients is then modelled by a generalised Gaussian distribution (GGD) with zero mean as in [20]:

$$(\quad) \quad [\ (\ | \quad |) \ ] , \tag{4}$$

with

$$(\quad) , \tag{5}$$

$$(\quad) \sqrt{(\quad) \ (\quad)} , \tag{6}$$

and

$$( ) \quad \int \quad . \tag{7}$$

In Equation (4), and are the mean, the variance and the shape parameter of the distribution respectively whereas ( ) is the gamma function. The estimated parameters: and are then chosen as the first two features.

Next, the empirical distributions of pairwise products of neighbouring luminance coefficients are computed on four orientations: horizontal, vertical, main-diagonal, and secondary-diagonal. Instead of GGD, the distributions are modelled by an asymmetric generalised Gaussian distribution (AGGD). The AGGD with zero mode is defined as [19]:

$$(\quad) \quad \frac{}{(\quad)(\quad)} \quad [\ (\quad) \ ] \quad , \tag{8}$$

and

$$(\quad) \quad \frac{}{(\quad)(\quad)} \quad [\ (\quad) \ ] \quad , \tag{9}$$

where

$$\sqrt{(\quad) \ (\quad)} \text{ and } \sqrt{(\quad) \ (\quad)} . \tag{10}$$

In these equations, and are the shape parameter, the left variance and the right variance of the distribution respectively. The three parameters and the mean of the best AGGD fit are then selected at each orientation to obtain another 16 features. In agreement with BRISQUE implementation, PATCH-IQ extracts these 18 features over two scales. A total of 36 features are used by PATCH-IQ to perform both distortion identification and quality estimation. Table 1 summarises the extracted features.

Table 1: List of extracted features

Feature ID	Scale	Orientation	Feature Description
1-2	1	-	Shape parameter and variance of GGD model of normalised luminance coefficients
3-6		Horizontal	Shape parameter, mean, left variance and right variance of AGGD model of pairwise products
7-10		Vertical	
11-14		Main-diagonal	
15-18		Secondary-diagonal	
19-20	2	-	Shape parameter and variance of GGD model of normalised luminance coefficients
21-24		Horizontal	Shape parameter, mean, left variance and right variance of AGGD model of pairwise products
25-28		Vertical	
29-32		Main-diagonal	
33-36		Secondary-diagonal	

3.3 Labelled Dataset Construction

Since PATCH-IQ employs a nearest-neighbour technique, a labelled dataset consisting of BIQA features extracted from patch exemplars must be constructed. Most of BIQA models employ the 80:20 train-test ratios to train their regression models. PATCH-IQ follows the same partition setting to build the dataset, i.e. patches from 80% of the randomly selected reference images from a standard IQA database and their distorted versions are used to extract the features for the dataset. Given an image, patches of size are first sampled at the interest point locations. BIQA features, as in section 3.2, are then extracted on those patches. However, instead of using all the extracted features, only features from patches are utilised in each image. This is done to ensure all images contribute the same number of features and to reduce the computational demands of the framework. The selected features are then combined over all images to form the dataset. Denote the total of labelled images by , the size of feature matrix for the dataset is:

$$[\quad] . \tag{11}$$

PATCH-IQ assigns the dataset patches with two labels. The first label is the distortion class. Each patch is labelled according to the distortion type in its source image. The second label is the subjective score. Each patch is assigned with its source images' subjective score, provided in the chosen IQA database. Assigning the score in this way is acceptable as the distortion levels across the database images are uniform. An example of a dataset built from the distorted versions of one reference image is shown in Figure 3. There is no need to fix a *priori*, the number of distortion classes in the dataset. If the images from new distortion classes are provided, they can be added directly to the dataset.

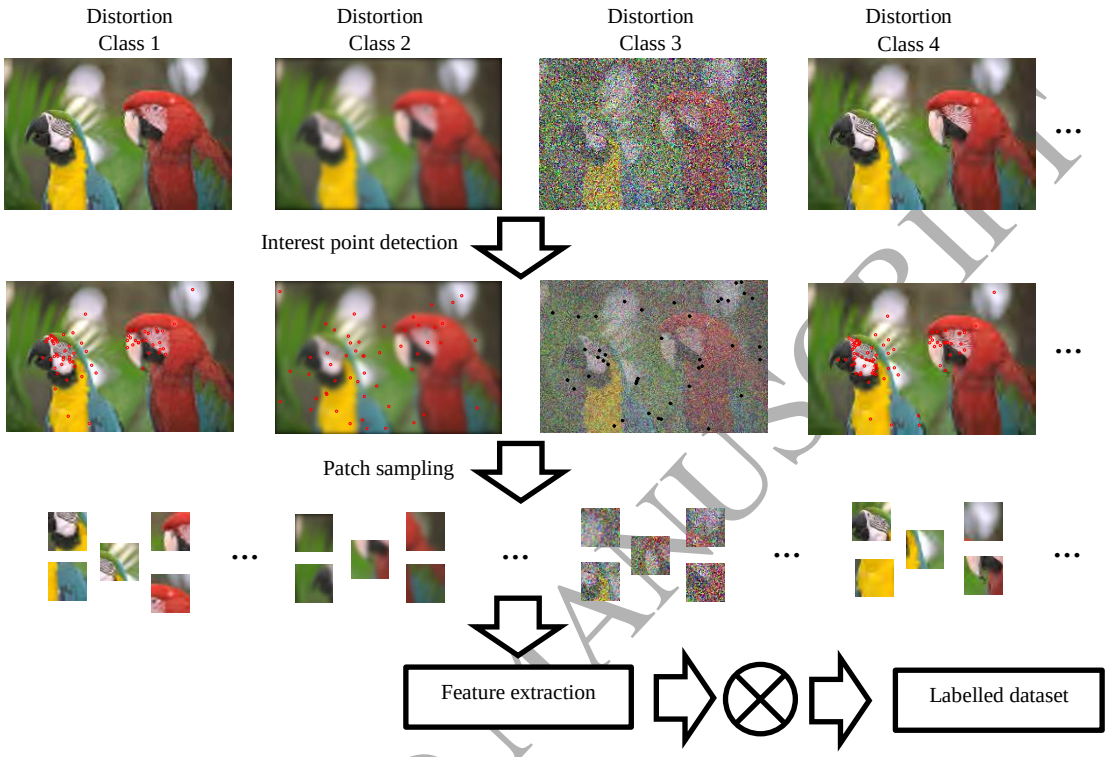

Figure 3: Example of labelled dataset construction

3.4 Distortion Identification

Our intuition in predicting image quality is that the quality of a distorted image will be best predicted using the images from the same distortion class. Therefore, we introduced a distortion identification stage prior to the quality estimation stage. Note that the BRISQUE features utilised in this work were mainly developed to evaluate the quality of an image. However, we observed that the features could also be utilised for distortion identification. To show that the utilised features capture image distortion, a 2-D scatter plot between the shape and the variance parameters of the GGD model of the normalised luminance coefficients is generated. Figure 4 shows the results for the undistorted reference images and their corresponding distorted versions from the LIVE IQA database [25]. The database consists of 5 distortion types: JPEG2000 compression (JP2K), JPEG compression (JPEG), additive white noise (WN), Gaussian blur (GB), and simulated fast fading channel (FF). It is easy to visualise from Figure 4 that images containing different types of distortions are well separated in the GGD parameter space showing the suitability of using these two features to perform distortion classification.

Meanwhile, in Figure 5, a 3-D scatter plot of the shape, the right variance and the left variance parameters of the AGGD model of the horizontally paired products is plotted using the same set of images. Again, it shows that different distortions occupy different regions of the parameter space. This justifies the use of these AGGD parameters as the features for distortion classification purposes. Similar patterns could be observed for other features extracted from different orientations and scales. We show in sub-section 4.7 how the classification performance varies when different groups of features are employed.

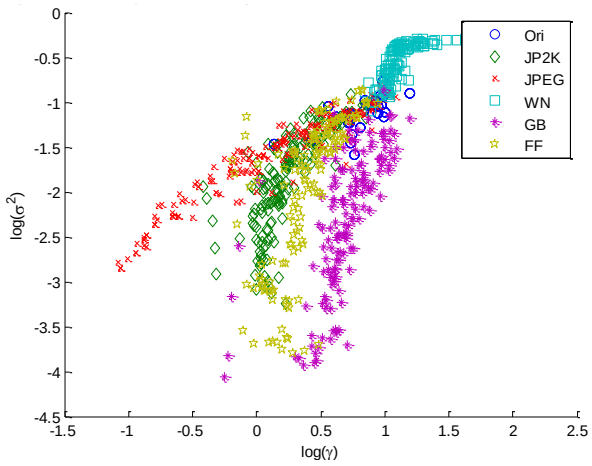

Figure 4: 2-D scatter plot between the shape and the scale parameters of the GGD model of the normalised luminance coefficients for the LIVE IQA database images.

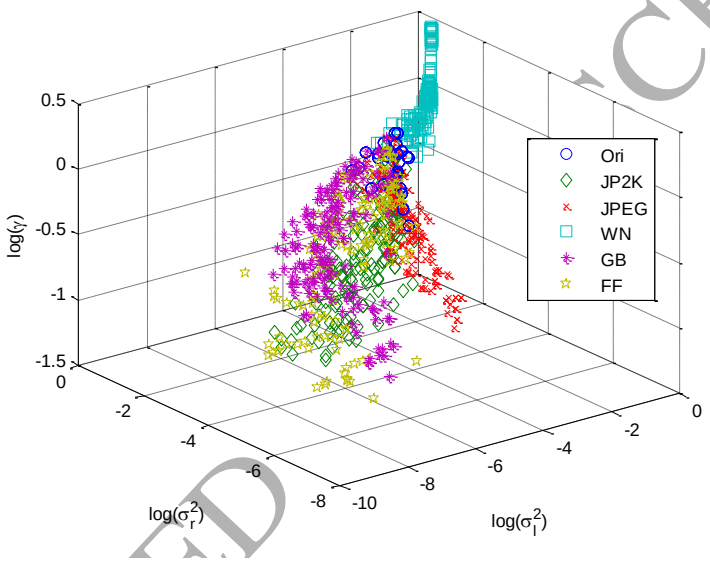

Figure 5: 3-D scatter plot of the shape, the left variance and the right variance parameters of the AGGD model of the pairwise product in horizontal orientation for the LIVE IQA database images

Given a test image , PATCH-IQ extracts the images' BIQA features using the same procedure in Sections 3.1 and 3.2. To further reduce computational time, only features from patches are chosen to form the image's feature matrix . PATCH-IQ then identifies the distortion type associated with the image by employing a nearest neighbour based classifier. In a nearest neighbour classification case, it has been shown that the optimal distance measurement is image-to-class (I2C) distance rather than the usually used image-to-image (I2I) distance. A popular I2C based classifier, the Naïve Bayes nearest neighbour (NBNN) [1], is utilised here. PATCH-IQ computes the distance between and the feature matrix from each of the distortion classes in the dataset . The predicted distortion class ˆ for the image is then represented by the class with the minimum I2C distance value:

$$\hat{\quad} \quad \| \quad (\ ) \| \ , \tag{12}$$

where ( ) is the NN-descriptor of in the distortion class .

3.5 Local Quality Estimation

The next stage is to estimate the quality of the image patches. PATCH-IQ works based on the intuition that the quality of a patch would be best predicted by patches of the same distortion type. Therefore, it performs quality estimation utilising only the labelled patches within the distortion class identified in the previous stage. PATCH-IQ also assumes that patches with similar features are perceived to have the same quality. Here, better

quality prediction can be achieved by utilising a set of labelled patches that are similar to the test patch in feature space. PATCH-IQ performs this through a *k*-NN regression algorithm.

For each test patch , the Euclidean distances between the patch and the selected labelled patches from the dataset is first calculated in the feature space. The labelled patches are then rearranged in ascending order according to the computed distances. The first labelled patches are then utilised to estimate the patch quality. However, instead of using common inverse distance weighting scheme, the patch quality is estimated through a simple linear regression:

( ) , (13)

where are the optimised weights for the patch feature vector .The weights can be calculated as:

( ) , (14)

where is the feature matrix of the selected labelled patches and represents their corresponding DMOS scores.

3.6 Global Quality Estimation

The final stage of the framework is basically a pooling stage. The patches' scores are pooled to yield the global quality score for the image. Instead of typical average or max pooling, PATCH-IQ employs an inverse weighting rule to pool all the patches' scores. Each local score is assigned a weight based on their minimum Euclidean distance computed in the previous local quality estimation stage. The image-level quality score for the image is given as:

$$\frac{\Sigma}{\Sigma}, \quad (15)$$

where $\frac{\Sigma}{}$. (16)

4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Evaluation Protocol

Databases: Performance evaluation of BIQA models is usually conducted using human-rated image databases. Each image in these databases is assigned with a DMOS / MOS value, which can measure correlation between a quality score predicted by a BIQA model and a quality score given by human. There are many subjective image databases available within the IQA research area. Three of the widely used databases: LIVE [25], CSIQ [13] and LIVEMD [10] are employed in this work. The LIVE database contains 982 images of which 779 images are distorted. These distorted images are generated when 29 reference images are subjected to 5 types of distortions at 5 to 6 degradation levels. The 5 distortions in the database are: JP2K, JPEG, WN, GB, and FF. Each image is provided with a DMOS value in the range between -5 and 130. Meanwhile, the CSIQ database consists of 30 reference images. Each reference image is distorted with 6 types of distortions at 4 to 5 degradation levels, yielding 866 distorted images. DMOS values assigned to these images are in the range between 0 and 1. All distorted images in the LIVE and CSIQ databases are subjected to a single type of distortion. The LIVEMD database also provides examples of multiple distorted images. 15 reference images are first blurred at 4 levels. The images are then subjected to one of two different distortions at 4 levels: JPEG or WN. A total of 225 single / multiple distorted images are generated for each of the two cases, GBJPEG and GBWN. For all three databases, a lower DMOS value indicates a higher quality image.

Framework parameters: The parameters are empirically determined. The number of patches for each labelled image and the number of test image patches are set at 30 and 100 respectively while the patch size is 256. For the feature extraction stage, the local window size is 3 and constant is 1 as in the BRISQUE model. Meanwhile, the number of NN patches for linear regression in the local quality estimation stage is set at 1000.

Performance metrics: Three metrics are commonly used to evaluate the performance of any BIQA models. They are: the linear correlation coefficient (LCC), the Spearman rank order correlation coefficient (SROCC), and the root mean square (RMSE). The LCC is used to measure a model's prediction accuracy and the SROCC is used to measure the prediction monotonicity of the model. The final metric RMSE can also evaluate the prediction accuracy of the model. A value closer to 1 (or -1) for both LCC and SROCC and a value closer to 0 for RMSE indicate higher correlation with human subjective score.

Benchmarked models: We compared PATCH-IQ with three FR-IQA models: PSNR, SSIM and FSIM and four state-of-the-art BIQA models: BIQI, BRISQUE, GMLOG and CORNIA, whose source codes are publicly available. Our previous BIQA model, NPNO is also included in the comparison. Since the four BIQA models require training, the databases are first divided into two subsets: a training set and a test set. The training set is generated from 80% randomly selected reference images and their associated distorted images while the remaining 20% reference images and their distorted versions are used for testing. There is no overlap between the two sets. The same training set is used to construct the labelled dataset required in PATCH-IQ and NPNO. LIBSVM [3], [4] is used to perform regression for the training-based BIQA models where their parameters are determined through cross validation in accordance to their papers.

4.2 Evaluation on Single Databases

Two experiments were performed to ascertain the overall performance and the distortion-specific (DS) performance. In the overall performance experiment, the train-test run is conducted across all distorted images regardless of their classes. This is to evaluate how well the model performs across all distortion types. In the DS performance experiment, the experiment is only carried out on images in a single distortion class. This is to evaluate how well it performs for one particular distortion. The median results across 1000 runs for both experiments are tabulated in Tables 2 and 3, respectively. For simplicity, only the SROCC results are shown for the DS performance experiment. Similar patterns can be observed for the LCC and RMSE results. Note that for the CSIQ database, we include results only from four distortions also present in the LIVE database: JP2K, JPEG, WN and GB. The top BIQA models are in bold.

Table 2: Median values across 1000 runs of the overall performance experiment

IQA model	LIVE			CSIQ		
	LCC	SROCC	RMSE	LCC	SROCC	RMSE
PSNR	0.8821	0.8829	12.8983	0.8562	0.9292	0.1444
SSIM	0.9464	0.9486	8.8035	0.9347	0.9362	0.0990
FSIM	0.9612	0.9639	7.5461	0.9675	0.9629	0.0710
BIQI	0.8486	0.8427	15.4068	0.8089	0.7491	0.1867
BRISQUE	0.9431	0.9421	9.3953	0.9304	0.9101	0.1073
GMLOG	0.9505	0.9503	8.8290	0.9394	0.9249	0.0997
CORNIA	0.9394	0.9416	9.9204	0.9110	0.8873	0.1254
NPNO	0.9525	**0.9540**	8.6407	0.9535	0.9384	0.0876
PATCH-IQ	***0.9562***	***0.9540***	***8.1490***	***0.9586***	***0.9430***	***0.0813***

Table 3: Median SROCC values across 1000 runs of the DS performance experiment

IQA model	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	0.8303	0.9062	0.9328	0.8656	0.6885	0.7635	0.9102	0.5397	0.7826
BRISQUE	0.9164	0.9640	0.9791	0.9446	0.8872	0.8977	0.9212	0.9207	0.9186
GMLOG	0.9268	0.9630	0.9831	0.9288	0.9012	0.9161	0.9364	0.9408	0.9083
CORNIA	0.9205	0.9359	0.9608	0.9519	**0.9052**	0.8942	0.8820	0.7862	0.9041
NPNO	**0.9497**	0.9725	0.9853	0.9448	0.8745	**0.9395**	0.9314	0.9591	0.9230
PATCHIQ	*0.9331*	***0.9732***	***0.9867***	***0.9697***	*0.8821*	*0.9326*	***0.9533***	***0.9654***	***0.9430***

In the overall performance experiment, PATCH-IQ has the best values for all three performance metrics among the competing BIQA models when tested in the LIVE database. Similar results are obtained for the CSIQ database. For the DS performance experiment, PATCH-IQ has the highest SROCC value among the competing BIQA models for images distorted by JPEG compression artefacts, WN or GB. It also gives the second best

performance in JP2K cases while giving comparable performance in FF cases. Compared to FR-IQA models, PATCH-IQ also achieves better overall performance compared to PSNR and SSIM while approaching FSIM. In individual distortions, it outperforms PSNR and yields competitive performance to SSIM and FSIM. It also outperforms both models for WN images. PATCH-IQ's performance is promising given it requires no reference image as its input as opposed to the FR-IQA models.

4.3 Evaluation on Multiple Distortion Database

To further investigate the effectiveness of the proposed framework, all the competing BIQA models are tested on the LIVEMD database. The database is more challenging as it also contains images that underwent multiple distortions. A similar experimental procedure is implemented as in the single distortion database. The results are presented in Table 4. The first five columns show the results from the DS performance experiment while the last column represents the results from the overall performance experiment. The top two models are in bold. The results suggest that PATCH-IQ generally has good prediction performance in the overall performance experiment where it consistently produces the top LCC, SROCC and RMSE values. In the DS performance experiment, it again performs the best in WN and GB cases while comes second in JPEG cases. For multiple distortions cases, PATCH-IQ is among the top two BIQA models for images distorted by GB and WN. For GBJPEG images, it has the second best SROCC value and gives comparable LCC and RMSE values.

Table 4: Median values across 1000 iterations on the LIVEMD database

LCC						
	GBJPEG	GBWN	GB	JPEG	WN	ALL
BIQI	0.7417	0.1291	0.8629	0.1005	0.5434	0.3312
BRISQUE	**0.8311**	0.8359	**0.8932**	0.6287	**0.9353**	**0.9188**
GMLOG	0.8118	0.7803	0.7712	0.6743	0.8451	0.8693
CORNIA	0.8250	**0.8658**	0.8537	0.5301	0.8026	0.9133
NPNO	**0.8458**	0.7141	0.8917	**0.7619**	0.8144	0.8749
PATCH-IQ	*0.8205*	***0.8551***	***0.8948***	***0.7264***	***0.9475***	***0.9311***
SROCC						
	GBJPEG	GBWN	GB	JPEG	WN	ALL
BIQI	0.7515	0.0617	0.8585	0.0833	0.5505	0.3570
BRISQUE	0.8172	0.8327	**0.8834**	0.6667	**0.8833**	0.9003
GMLOG	0.8107	0.7619	0.7755	0.6667	0.8000	0.8451
CORNIA	0.8089	**0.8551**	0.8349	0.4833	0.7667	**0.9017**
NPNO	**0.8382**	0.6628	0.8797	**0.7667**	0.8000	0.8459
PATCH-IQ	***0.8239***	***0.8640***	***0.8842***	***0.7167***	***0.8833***	***0.9106***
RMSE						
	GBJPEG	GBWN	GB	JPEG	WN	ALL
BIQI	8.8770	44.3462	9.4397	10.3556	12.7309	25.8004
BRISQUE	7.9994	8.4818	**8.7194**	7.2798	**6.3376**	**8.4282**
GMLOG	8.3557	9.9733	12.4358	7.4688	9.7714	10.2195
CORNIA	**7.8099**	**8.0264**	10.1168	8.1784	9.3534	8.6813
NPNO	**7.9912**	10.3441	9.2250	**5.2061**	13.0281	9.9040
PATCH-IQ	*8.2604*	***8.4084***	***8.5511***	***5.7410***	***5.8369***	***8.1425***

4.4 Statistical Significance and Hypothesis Testing

The differences in median correlations between the competing BIQA models may not be statistically significant. Therefore, a hypothesis test to evaluate the statistical significance difference between each model is conducted. As the SROCC and LCC values follow right-skewed unimodal distributions, the Wilcoxon rank-sum test is employed avoiding the normality assumption required by a typical t-test [26]. The Wilcoxon rank-sum test measures the equivalence of the median values of two independent samples. The test is performed on the SROCC values obtained from the 1000 runs of experiments at a significance level of 0.01. The null hypothesis is that the SROCC values of the two BIQA models are drawn from the populations with equal median while the alternative hypothesis is that the median of one model is greater than the other.

The results are shown in Table 5. A score of '1' implies there is a statistically significant difference between both models and the model in row has a larger median than the model in column. A score of '-1' also implies there is a statistically significant difference between the models, but the model in column now has a larger median than the model in row. A score of '0' indicates the null hypothesis cannot be rejected; there is no statistically significant difference between both row and column models. On the LIVE database, PATCH-IQ has

no statistically significant difference to NPNO but differs to the rest. However, on the CSIQ and the LIVEMD databases, PATCH-IQ is different to all models.

Table 5: Results of the Wilcoxon rank-sum test using the SROCC values of competing BIQA models

LIVE						
	BIQI	BRISQUE	GMLOG	CORNIA	NPNO	PATCH-IQ
BIQI	0	-1	-1	-1	-1	-1
BRISQUE	1	0	-1	1	-1	-1
GMLOG	1	1	0	1	-1	-1
CORNIA	1	-1	-1	0	-1	-1
NPNO	1	1	1	1	0	0
PATCH-IQ	*1*	*1*	*1*	*1*	*0*	*0*
CSIQ						
	BIQI	BRISQUE	GMLOG	CORNIA	NPNO	PATCH-IQ
BIQI	0	-1	-1	-1	-1	-1
BRISQUE	1	0	-1	1	-1	-1
GMLOG	1	1	0	1	-1	-1
CORNIA	1	-1	-1	0	-1	-1
NPNO	1	1	1	1	0	-1
PATCH-IQ	*1*	*1*	*1*	*1*	*1*	*0*
LIVEMD						
	BIQI	BRISQUE	GMLOG	CORNIA	NPNO	PATCH-IQ
BIQI	0	-1	-1	-1	-1	-1
BRISQUE	1	0	1	0	1	-1
GMLOG	1	-1	0	-1	0	-1
CORNIA	1	0	1	0	1	-1
NPNO	1	-1	0	-1	0	-1
PATCH-IQ	*1*	*1*	*1*	*1*	*1*	*0*

4.5 Influence of Framework Parameter on Its Performance

To investigate the effect of varying the number of images in the labelled dataset, all three databases are partitioned under three train-test ratios: 80:20, 50:50 and 30:70. The remaining five competing BIQA models are also evaluated under the same settings. The number of selected patches for each labelled image and test image and the patch size are fixed as before. The SROCC results for the overall performance experiment are shown in Table 6. As expected, the performances of all tested BIQA models decrease as the number of samples reduces. On the LIVE database, PATCH-IQ has the second best SROCC values after NPNO at 50% and 30% training ratios. However, when tested on the CSIQ and the LIVEMD databases, PATCH-IQ produces the best SROCC values at all three training ratios. This shows it has better generalisation ability than NPNO. Compared to the remaining four BIQA models, it is also interesting to note that PATCH-IQ still gives better SROCC values at 50% ratio than the other models' results at 80% ratio on the CSIQ and LIVEMD while slightly lags behind GMLOG on the LIVE database. This proves that the PATCH-IQ framework works well when the number of training samples is small.

Table 6: SROCC comparison for different training (labelled) samples ratio

Database	LIVE			CSIQ			LIVEMD		
Ratio	80%	50%	30%	80%	50%	30%	80%	50%	30%
BIQI	0.8427	0.8346	0.8147	0.7491	0.7369	0.7182	0.3570	0.3421	0.3218
BRISQUE	0.9421	0.9274	0.9033	0.9101	0.8951	0.8722	0.9003	0.8921	0.8828
GMLOG	0.9503	0.9402	0.9251	0.9249	0.9091	0.8870	0.8451	0.8121	0.7760
CORNIA	0.9416	0.9374	0.9290	0.8873	0.8812	0.8734	0.9017	0.8984	0.8933
NPNO	**0.9540**	**0.9510**	**0.9452**	0.9384	0.9295	0.9143	0.8459	0.8427	0.8331
PATCHIQ	***0.9540***	*0.9471*	*0.9346*	***0.9430***	***0.9320***	***0.9151***	***0.9106***	***0.9056***	***0.8950***

The results of varying the number of patches in each labelled image on the LIVE database at 80% training ratio are shown in Table 7 and Figure 6, respectively. A higher number of utilised patches will lead to higher SROCC and LCC values. However, it will lead to longer computation time for the identification of the distortion. Here, PATCH-IQ chooses the lowest number of patches that outperforms the state-of-the-art BIQA models while has acceptable processing time.

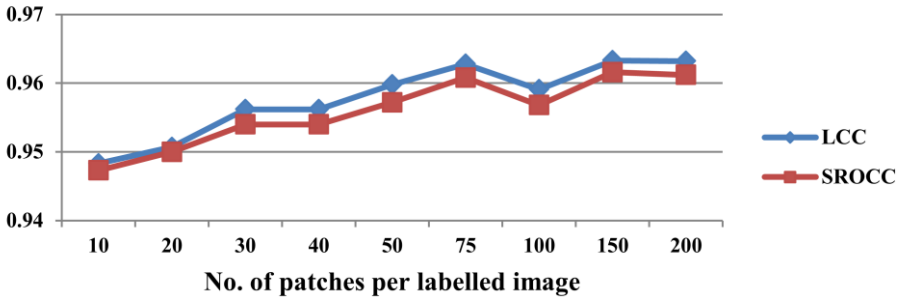

Figure 6: LCC and SROCC comparison for different number of patches in a labelled image on LIVE database

Table 7: LCC and SROCC comparison for different number of patches in a labelled image

Patch No	10	20	30	40	50	75	100	150	200
LCC	0.9483	0.9507	*0.9562*	0.9562	0.9598	0.9628	0.9591	0.9633	0.9632
SROCC	0.9473	0.9500	*0.9540*	0.9540	0.9572	0.9608	0.9568	0.9616	0.9612

The effect of the number of the nearest neighbour patches used for linear regression on the model performance is investigated. All other parameters are again fixed at the initial values. The performance variation of PATCH-IQ when tested on the LIVE database is shown in Table 8. Based on the results, there is a small variation on the obtained values, indicating that the effect of the number of labelled patches is not significant. The number that provides the optimum performance is empirically chosen. Here, the optimum performance is achieved when the number is set at 1000.

Table 8: Performance variations for different numbers of NN patches used in regression

Patch No.	5	10	50	100	500	1000	2000	3000	ALL
LCC	0.9456	0.9493	0.9280	0.9502	0.9530	***0.9562***	0.9529	0.9496	0.9450
SROCC	0.9420	0.9458	0.9336	0.9486	0.9516	***0.9540***	0.9514	0.9487	0.9441
RMSE	9.1137	8.8202	10.3455	8.6313	8.4352	***8.1490***	8.4401	8.7318	9.0979

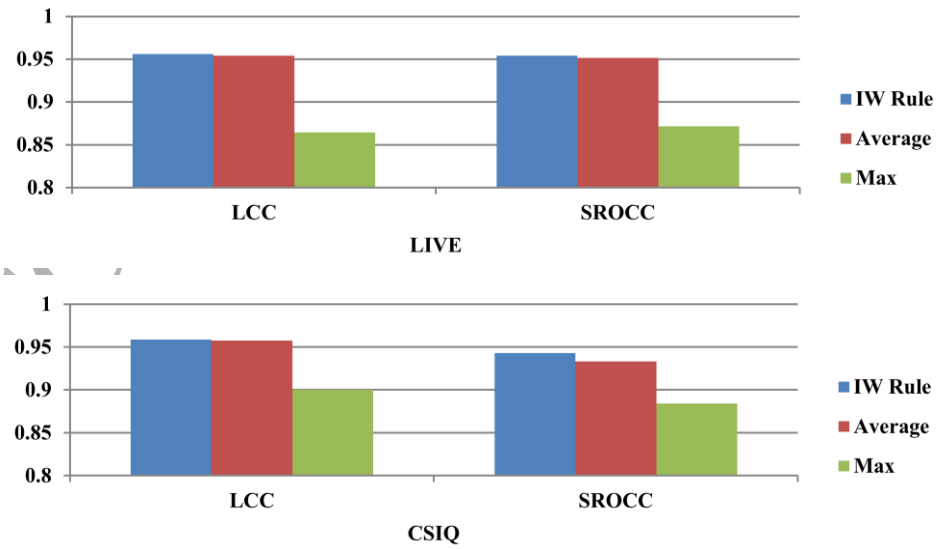

Figure 7: LCC and SROCC comparison for different pooling methods on LIVE and CSIQ databases

The quality prediction performance of PATCH-IQ also depends on how the scores from test patches are pooled. In Section 3.6, PATCH-IQ pools all the patches' scores by assigning weight to each score according to an inverse weighting rule. Two other pooling methods: average pooling and max pooling are tested and the

results are shown in Table 9 and Figure 7. Among the three pooling methods, we can observe that the inverse-weighted based pooling method based consistently produces the highest SROCC, LCC and RMSE values. It provides slight improvement to average pooling while better than max pooling.

Table 9: Performance comparison for different pooling methods

Database	LIVE			CSIQ		
Metrics	LCC	SROCC	RMSE	LCC	SROCC	RMSE
IW Rule	0.9562	0.9540	8.1490	0.9586	0.9430	0.0813
Average	0.9540	0.9514	8.4809	0.9574	0.9333	0.0850
Max	0.8646	0.8717	19.4673	0.9001	0.8842	0.1790

4.6 Distortion Identification Accuracy

Another useful property of PATCH-IQ is its ability to identify the distortion affecting the image. To show that the chosen NBNN classifier is capable to provide good classification performance, the median classification accuracy over 1000 runs of experiments on all three databases is reported. The results are tabulated in Table 10. The chosen classifier consistently achieves good performance across many distortions with the minimum accuracy value of 80%. Since the classifier uses the extracted spatial domain features as its input descriptors, the results indicate that the features are not only suitable for quality estimation but also suitable for distortion identification purposes.

Table 10: Median classification accuracy across 1000 iterations

LIVE	JP2K	JPEG	WN	GB	FF	ALL
Accuracy	88.57	97.22	100	96.67	80	91.98
CSIQ	JP2K	JPEG	WN	GB	FF	ALL
Accuracy	90	86.67	93.33	90	-	89.17
LIVEMD	GBJEG	GBWN	GB	JPEG	WN	ALL
Accuracy	100	99.98	99.99	93.76	91.97	98.56

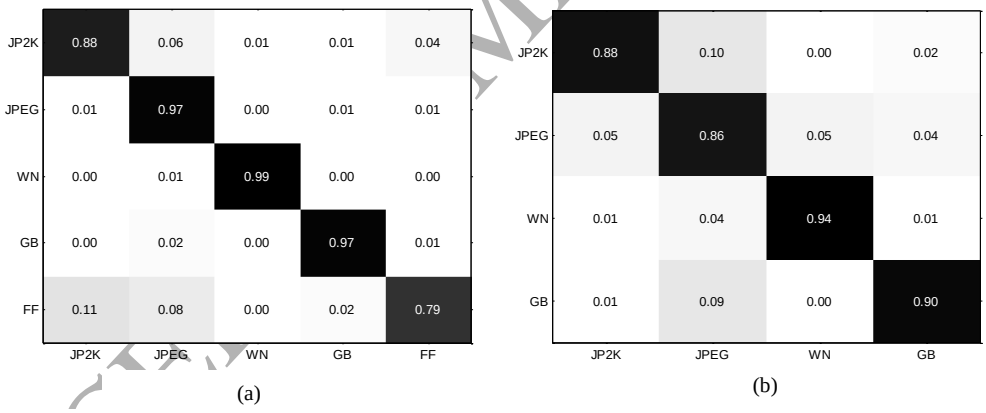

Figure 8: Mean confusion matrix across 1000 runs of experiments for distortion classification: (a) LIVE and (b) CSIQ

To allow visualisation of the classification performance, Figure 8 plots the confusion matrix for each distortion classes in both the LIVE and the CSIQ databases. We can use the confusion matrix to see if PATCH-IQ is confusing two distortion classes. Each column of the matrix represents the instances in the predicted distortion class while each row represents the instances in the actual distortion class. Each row adds up to 1 and the values represent the mean percentage for the 1000 runs of experiments. Higher value indicates greater confusion. In the LIVE database, we can see that WN, GB and JPEG are generally well classified by PATCH-IQ and not confused with other distortion. JP2K and FF are most confused with each other whereby about 11% of FF images are misclassified as JP2K images and about 4% of JP2K images are predicted as FF images. This is because FF images in the database are essentially JP2K compressed images followed by packet-loss errors [25]. In the CSIQ database, good classification performance is achieved by PATCH-IQ with less than 6% of the WN images are misclassified. JPEG is the most confused distortion with 10% of the images are misclassified as JP2K or WN images while another 4% are wrongly predicted as GB images.

4.7 Feature Analysis

To visualise the relationship between the extracted features and the human perception of image quality, the SROCC values between each feature and DMOS values for the LIVE database images are plotted in Figure 9. We can see that each feature capture quality information differently and they vary depending on distortion. The proposed features correlate well with the human perception of quality for images affected by noise. Most features are also useful for quality prediction of blurred images or images distorted by JP2K compression artefacts. In each distortion case, we can observe that the variance parameters of both the GGD model and the AGGD model have better correlation with subjective scores compared to the shape parameters of the models. Meanwhile, among all the proposed features, the mean parameters of the AGGD models capture quality information the least. Another observation we can made is the same features extracted in different orientations generally have similar correlation values.

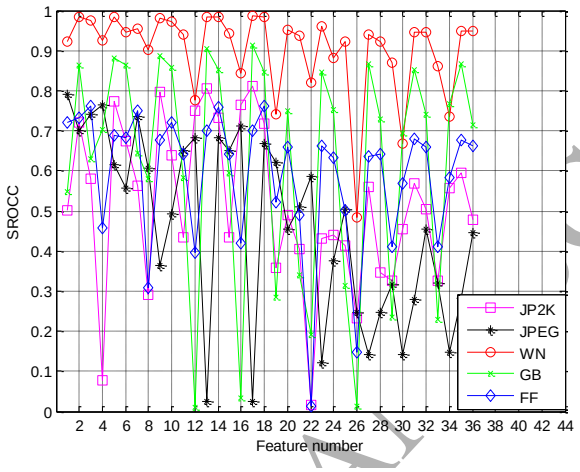

Figure 9: Correlation of the extracted features with the DMOS for different distorted images in the LIVE database

To evaluate the contributions of different features on both the distortion classification and the quality prediction performances, five groups of feature are tested on the framework; 1) All features (denoted as PATCH-IQ), 2) The GGD model-based features (denoted as PATCH-IQ2), 3) The AGGD model-based features (denoted as PATCH-IQ3), 4) All features except the mean parameter of the AGGD models (PATCH-IQ4) and 5) The variance parameters of both the GGD model and the AGGD model (PATCH-IQ5). PATCH-IQ2 should study the contribution of features derived directly from the locally normalised luminance coefficients whereas PATCH-IQ3 is to evaluate the effects of features derived from the pairwise products of these coefficients. Meanwhile, features for PATCH-IQ4 and PATCH-IQ5 are selected based on the previous discussions.

The median classification accuracy values over 1000 runs of experiments for the five PATCH-IQ versions tested on the LIVE database are tabulated in Table 11. We can see that the best classification results for both the overall and the DS experiments are achieved when all 36 features are utilised. PATCH-IQ3 has better classification accuracy than PATCH-IQ2 showing that the AGGD model-based features contribute more to a distortion identification task than the GGD model-based features. We can also observe that removing the mean parameters of the AGGD models as in PATCH-IQ4 has little effect to the classification performance. This indicates that the mean parameters of the AGGD models have small contributions to such a task. The classification accuracy also drops when only variance parameters are utilised as in PATCH-IQ5.

Table 11: Median classification accuracy values for different group of features on the LIVE database

	PATCH-IQ	PATCH-IQ2	PATCH-IQ3	PATCH-IQ4	PATCH-IQ5
JP2K	88.57	82.35	88.57	88.24	79.42
JPEG	97.22	88.57	97.22	96.92	94.29
WN	100	96.67	100	100	100
GB	96.67	96.67	96.67	96.67	93.33
FF	80	66.67	79.42	80	66.67
ALL	91.98	85.80	91.93	91.82	85.80

Table 12 shows the median SROCC values over 1000 trials obtained by the same five PATCH-IQ versions when tested on the LIVE database. Few similar observations can be made here. First, the best quality prediction performances for both experiments are produced when PATCH-IQ utilised all the proposed features. Second, PATCH-IQ3 has better correlation values in most distortion cases than PATCH-IQ2. This indicates that the AGGD model-based features have better correlation to human perceptual measures than the GGD model-based features. Third, PATCH-IQ4 achieves similar prediction performances to PATCH-IQ for images affected by noise and compression artefacts while only suffers a slight degradation in performance for GB and FF images. This shows that the mean parameters of the AGGD models contribute little to a quality prediction task. Meanwhile, PATCH-IQ5 also achieves close prediction performance to PATCH-IQ in both experiments. This suggests that, while the variance parameters of both the GGD model and the AGGD models may not be suitable features for a distortion classification task, they are useful features for image quality prediction.

Table 12: Median SROCC values for different group of features on the LIVE database

	PATCH-IQ	PATCH-IQ2	PATCH-IQ3	PATCH-IQ4	PATCH-IQ5
JP2K	0.9331	0.9090	0.9241	0.9331	0.9187
JPEG	0.9733	0.9591	0.9733	0.9733	0.9720
WN	0.9867	0.9671	0.9867	0.9867	0.9867
GB	0.9697	0.9406	0.9666	0.9671	0.9693
FF	0.8821	0.8665	0.8661	0.8729	0.8732
ALL	0.9540	0.9319	0.9481	0.9534	0.9465

4.8 Computational Complexity

Having a fast computation speed is always desirable especially for applications that require online quality assessment like adaptive coding in video streaming. PATCH-IQ's processing time is analysed in this sub-section. There are three major stages that consume most of the processing time: (1) patch and feature extraction; (2) distortion identification; and (3) local quality estimation. Processing times are based on un-optimised MATLAB R2011b code on an 8GB RAM computer with an Intel i5 3.20 GHz processor.

Extracting IQA features is the most time consuming part of the model framework as features are extracted at the patch level rather than at the image level. A higher number of patches will lead to longer extraction time. Additional computation time is also required for interest points' detection. In addition, the choice of statistical features to be utilised also plays important roles in keeping acceptable processing time. On average, utilising spatial domain features described in sub-section 3.2 and setting the number of test patches as in sub-section 4.1, PATCH-IQ requires 1.03 seconds to extract the features in a typical 512 768 image. Processing time of the distortion identification stage is determined by the I2C distance computation. It depends on the size of the labelled dataset. The dataset size is determined by the number of labelled images and the number of patches within those images. A larger dataset will require longer time to compute the I2C distance between the test patches and their nearest neighbour labelled patches. However, a larger database will lead to better prediction performance. Therefore, there is a clear trade-off between the prediction performance and the I2C distance computation time. Choosing an appropriate dataset size is essential to ensure fast computations while achieving competitive prediction performance. At 80% ratio, PATCH-IQ requires an additional 0.05 second to perform distortion identification.

Finally, the local quality estimation processing time is directly related to the number of NN patches selected for linear regression. A higher number of patches will lead to longer quality estimation time. Setting the parameters as in sub-section 4.1, an extra 0.08 second performs quality estimation for all test patches. We do not consider the time to construct the labelled dataset as it is assumed that it is already available prior to the testing stage. The average run-time comparison between PATCH-IQ and the competing BIQA models is shown in Table 13. BIQI is the fastest but has the worst performance among all the compared models. PATCH-IQ is slower than others except CORNIA. However, given its superior performance, PATCH-IQ can be a better option for IQA applications when real time computation is not a key requirement.

Table 13: Average run-time

BIQA model	BIQI	BRISQUE	GMLOG	CORNIA	NPNO	*PATCH-IQ*
Run-time (s)	0.08	0.18	0.10	2.43	0.19	*1.16*

5. CONCLUSION AND FUTURE WORK

In this paper, a simple but effective BIQA model that estimates image quality without the presence of a reference image is presented. The model, PATCH-IQ, is based on a five-stage framework that operates in a spatial domain. In contrast to many previous BIQA models, PATCH-IQ predicts the quality of an image directly from a set of annotated patches using nearest neighbour methods. The approach alleviates the need of any prior training phase. PATCH-IQ can also estimate image quality locally and identify the distortion affecting the image, two useful properties that are not available in most of current BIQA models. The model is tested extensively on three subject-rated image databases. The experimental results demonstrated that the image quality estimates of PATCH-IQ are highly correlated with human perceptual measures of image quality across various kinds of image distortions. PATCH-IQ also has greater performance to all competing BIQA models in quality prediction accuracy and robustness. PATCH-IQ also generalises well across different databases including the one with multiple distorted images.

Despite these promising results, several steps could be taken to improve PATCH-IQ. PATCH-IQ relies on a labelled dataset and has only being tested on distortions in the standard IQA databases. Introducing new types of distortions will increase the dataset size, leading to higher memory and processing time requirements. Here, the use of parallel computing or less computational expensive feature extraction methods could be explored to speed-up the process. We could also integrate incremental learning techniques in the dataset construction to help dealing with an increasing number of new distortion classes. In addition, obtaining accurate image distortion class is essential to provide PATCH-IQ with better regression inputs for quality estimation stages. While PATCH-IQ uses spatial domain features and a NBNN classifier to perform the classification, different features and other nearest neighbour classifiers could also be tested to obtain better classification accuracy.

ACKNOWLEDGEMENT

This work is supported by the Malaysian Government and Universiti Teknikal Malaysia Melaka (UTeM) through SLAI Sponsorship Awards.

REFERENCES

[1] O. Boiman, E. Shectman, and M. Irani, In defense of nearest neighbor based image classification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1-8. DOI: http://dx.doi.org/10.1109/cvpr.2008.4587598

[2] D. M. Chandler and S. S. Hemami, VSNR: A wavelet-based visual signal-to-noise ratio for natural images, IEEE Trans. Image Process. 16 (9) (2007) 2284-2298. DOI: http://dx.doi.org/10.1109/TIP.2007.901820

[3] C. –C. Chang and C. –J. Lin, LIBSVM: A library for support vector machines [Online]. Available: https://www.csie.ntu.edu.tw/~cjlin/libsvm

[4] C. –C. Chang and C. –J. Lin, LIBSVM, ACM Trans. Intell. Syst. and Technol. 2 (3) (2011) 1-27. DOI: http://dx.doi.org/10.1145/1961189.1961199

[5] C. Charrier, O. l'ezoray, and G. Lebrun, Machine learning to design full-reference image quality assessment algorithm, Signal Process. Image Commun. 27 (3) (2012) 209-219. DOI: http://dx.doi.org/10.1016/j.image.2012.01.002

[6] E. Cohen and Y. Yitzhaky, No-reference assessment of blur and noise impacts on image quality, Signal, Image and Video Process. 4 (3) (2010) 289-302. DOI: http://dx.doi.org/10.1007/s11760-009-0117-4

[7] P. Engeldrum, A theory of image quality: The image quality circle, J. Imaging Sci. and Technol. 48 (5) (2004) 446-456.

[8] K. Gu, G. Zhai, X. Yang, and W. Zhang, Using free energy principle for blind image quality assessment, IEEE Trans. Multimed. 17 (1) (2015) 50-63. DOI: http://dx.doi.org/10.1109/TMM.2014.2373812

[9] R. Hong, J. Pan, S. Hao, M. Wang, F. Xue, and X. Wu, Image quality assessment based on matching pursuit, Inf. Sci. 273 (2014) 196-211. DOI: http://dx.doi.org/10.1016/j.ins.2014.03.009

[10] D. Jayaraman, A. Mittal, A. K. Moorthy, and A. C. Bovik, Objective quality assessment of multiply distorted images, in: Proceedings of the IEEE Asilomar Conference on Signals, Systems and Computers, 2012, pp. 1693-1697. DOI: http://dx.doi.org/10.1109/ACSSC.2012.6489321

[11] T. Judd, K. Ehinger, F. Durand, and A. Torralba, Learning to predict where humans look, in: Proceedings of the IEEE Conference on Computer Vision, 2009, pp. 2106-2113. DOI: http://dx.doi.org/10.1109/iccv.2009.5459462

[12] L. Kang, P. Ye, Y. Li, and D. Doermann, Convolutional neural networks for no-reference image quality assessment, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1733-1740. DOI: http://dx.doi.org/10.1109/CVPR.2014.224

[13] E. C. Larson and D. M. Chandler, Most apparent distortion: Full-reference image quality assessment and the role of strategy, J. Electron. Imaging 19 (1) (2010) 1-21. DOI: http://dx.doi.org/10.1117/1.3267105

[14] D. G. Lowe, Distinctive image features from scale-invariant keypoints, J. Comput. Vis. 60 (2) (2004) 91-110. DOI: http://dx.doi.org/10.1023/b:visi.0000029664.99615.94

[15] Q. Lu, W. Zhou and H. Li, A no-reference image sharpness metric based on structural information using sparse representation, Inf. Sci. 369 (2016) 334-346. DOI: http://dx.doi.org/10.1016/j.ins.2016.06.042

[16] R. A. Manap and L. Shao, Non-distortion-specific no-reference image quality assessment: a survey, Inf. Sci. 301 (2015) 141-160. DOI: http://dx.doi.org/10.1016/j.ins.2014.12.055

[17] R. A. Manap, , L. Shao and A. F. Frangi, A non-parametric framework for no-reference image quality assessment, in: Proceedings of the IEEE Global Conference on Signal and Information Processing, 2015, pp. 562-566. DOI: http://dx.doi.org/10.1109/globalsip.2015.7418258

[18] R. A. Manap, L. Shao, and A. F.Frangi, Non-parametric quality assessment of natural images, IEEE Multimed. 23 (4) (2016) 22-30. DOI: http://dx.doi.org/10.1109/mmul.2016.2

[19] A. Mittal, A. K. Moorthy, and A. C. Bovik, No-reference image quality assessment in the spatial domain, IEEE Trans. Image Process. 21 (12) (2012) 4695-4708. DOI: http://dx.doi.org/10.1109/TIP.2012.2214050

[20] A. K. Moorthy and A. C. Bovik, A two-step framework for constructing blind image quality indices, IEEE Signal Process. Lett. 17 (5) (2010) 513-516. DOI: http://dx.doi.org/10.1109/LSP.2010.2043888

[21] A. K. Moorthy and A. C. Bovik, Blind image quality assessment: from natural scene statistics to perceptual quality, IEEE Trans. Image Process. 20 (12) (2011) 3350-3364. DOI: http://dx.doi.org/10.1109/TIP.2011.2147325

[22] 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, J. Adv. Mod. Radioelectron. 10 (4) (2009) 30-45.

[23] A. Rehman and Z. Wang, Reduced-reference image quality assessment by structural similarity estimation, IEEE Trans. Image Process. 21 (8) (2012) 3378-3389. DOI: http://dx.doi.org/10.1109/TIP.2012.2197011

[24] 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. 21 (8) (2012) 3339-3352. DOI: http://dx.doi.org/10.1109/TIP.2012. 2191563

[25] H. R. Sheikh, Z. Wang, L. Cormark, and A. C. Bovik, LIVE Image Quality Assessment Database Release 2 [Online]. Available: http://live.ece.utexas.edu/research/quality

[26] D. Sheskin, Handbook of parametric and nonparametric statistical procedures, London, UK: Chapman & Hall, 2004.

[27] R. Soundararajan and A. C. Bovik, Rred indices: reduced reference entropic differencing for image quality assessment, IEEE Trans. Image Process., 21 (2) (2012) 517-526. DOI: http://dx.doi.org/10.1109/TIP.2011.2166082

[28] T. Tuytelaars, Dense interest points, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 2281-2288. DOI: http://dx.doi.org/10.1109/CVPR.2010.5539911

[29] Q. Wang, J. Chu, L. Xu and Q. Chen, A new blind image quality framework based on natural color statistic, J. Neurocomput. 172 (3) (2016) 1798-1810. DOI: http://dx.doi.org/10.1016/j.neucom.2015.09.057

[30] Z. Wang and A. C. Bovik, Mean squared error: Love it or leave it? A new look at signal fidelity measures, IEEE Signal Process. Mag. 26 (1) (2009) 98-117. DOI: http://dx.doi.org/10.1109/MSP.2008.930649

[31] Z. Wang and A. C. Bovik, Modern image quality assessment, Synthesis Lectures on Image, Video and Multimedia Processing, Morgan and Claypool Publishers, 2006.

[32] 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. 13 (4) (2004) 600-612. DOI: http://dx.doi.org/10.1109/TIP.2003. 819861

[33] J. Wu, W. Lin, G. Shi, L. Li and Y. Fang, Orientation selectivity based visual pattern for reduced reference image quality assessment, Inf. Sci. 351 (2016) 18-29. DOI: http://dx.doi.org/10.1016/j.ins.2016.02.043

[34] 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. 23 (11) (2014) 4850-4862. DOI: http://dx.doi.org/10.1109/TIP.2014.2355716

[35] P. Ye and D. Doermann, No-reference image quality assessment using visual codebooks, IEEE Trans. Image Process. 21 (7) (2012) 3129-3138. DOI: http://dx.doi.org/10.1109/TIP.2012.2190086

[36] P. Ye, J. Kumar, L. Kang, and D. Doermann, Unsupervised feature learning framework for no-reference image quality assessment, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 1098-1105. DOI: http://dx.doi.org/10.1109/CVPR.2012. 6247789

[37] P.Ye, J. Kumar, L. Kang, and D. Doermann, Real-time no-reference image quality assessment based on filter learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 987-994. DOI: http://dx.doi.org/10.1109/CVPR.2013.132

[38] G. Zhai, X. Wu, X. Yang, W. Lin, and W. Zhang, A psychovisual quality metric in free-energy principle, IEEE Trans. Image Process. 21 (1) (2012) 41-52. DOI: http://dx.doi.org/10.1109/TIP.2011.2161092

[39] J. Zhang and T. M. Le, A new no-reference quality metric for JPEG2000 images, IEEE Trans. Consumer Electron. 56 (2) (2010) 743-750. DOI: http://dx.doi.org/10.1109/TCE.2010.5505996

[40] L. Zhang, X. Mou, and D. Zhang, FSIM: A feature similarity index for image quality assessment, IEEE Trans. Image Process. 20 (8) (2011) 2378-2386. DOI: http://dx.doi.org/10.1109/TIP.2011.2109730

[41] Y. Zhang and D. M. Chandler, An algorithm for no-reference image quality assessment based on log-derivative statistics of natural scenes, Proc. SPIE-IS&T Electron. Imaging 8653 (2013) 1-10. DOI: http://dx.doi.org/10.1117/12.2001342