

FREE- REFERENCE IMAGE QUALITY ASSESSMENT FRAMEWORK USING METRICS FUSION AND DIMENSIONALITY REDUCTION

Besma Sadou¹, Atidel Lahoulou², Toufik Bouden¹, Anderson R. Avila³,
Tiago H. Falk³ and Zahid Akhtar⁴

¹Non Destructive Testing Laboratory, University of Jijel, Algeria, ²LAOTI laboratory, University of Jijel, Algeria, ³Institut National de la Recherche Scientifique, University of Québec, Montreal, Canada and ⁴University of Memphis, USA

ABSTRACT

This paper focuses on no-reference image quality assessment(NR-IQA)metrics. In the literature, a wide range of algorithms are proposed to automatically estimate the perceived quality of visual data. However, most of them are not able to effectively quantify the various degradations and artifacts that the image may undergo. Thus, merging of diverse metrics operating in different information domains is hoped to yield better performances, which is the main theme of the proposed work. In particular, the metric proposed in this paper is based on three well-known NR-IQA objective metrics that depend on natural scene statistical attributes from three different domains to extract a vector of image features. Then, Singular Value Decomposition (SVD) based dominant eigenvectors method is used to select the most relevant image quality attributes. These latter are used as input to Relevance Vector Machine (RVM) to derive the overall quality index. Validation experiments are divided into two groups; in the first group, learning process (training and test phases) is applied on one single image quality database whereas in the second group of simulations, training and test phases are separated on two distinct datasets. Obtained results demonstrate that the proposed metric performs very well in terms of correlation, monotonicity and accuracy in both the two scenarios.

KEYWORDS

Image quality assessment, metrics fusion, Singular Value Decomposition (SVD), dominant eigenvectors, dimensionality reduction, Relevance Vector Machine (RVM)

1. INTRODUCTION

These days, images have become very integral part of our daily lives; they have become an essential means of communication. However, during acquisition, compression, transmission and reproduction, the image can suffer from several degradations and artifacts, in this fact, human seeks to evaluate image quality by proposing quality metrics that resemble human judgment which remains the most reliable way to evaluate the quality of images. Depending on the availability of the reference image, objective metrics can be categorized into three categories: full reference, reduced reference or no-reference (also called blind or free-reference) image quality metrics.

In the latter class, we can distinguish two types of metrics: metrics intended for specific degradation, and general-purpose metrics where the type of degradation is not praying in.

In the literature, the most efficient free-reference metrics are based on natural scene statistics (NSS)which assume that all original images are natural and that the distortions disrupt the

naturalness and make images seem unnatural [1]. Several works in this context are proposed in different domains such as Discrete Wavelet Transform (DWT) domain (e.g. BIQI [2] and DIIVINE [3]), the Discrete Cosine Transform (DCT) domain (e.g. BLIINDS [4] and BLIINDS-II [5]) and the spatial domain (e.g. BRISQUE [6] and NIQE [7]). However, the latest generation of metrics exploits the multi-domain information, which simulates well the hierarchical structure of the visual cortex perception [8-9] (e.g. WG-LAB [10] and metrics proposed in [11] and [12]).

In this paper, we develop a new and efficient no-reference image quality analysis NR-IQA metric for grey level images. First, a features vector is extracted using three well known NR-IQA metrics operating in three different domains (i.e. DCT domain, DWT domain and spatial domain) in order to better capture human vision properties. After that, the variable selection process is launched to keep only the most pertinent attributes. This step is performed by using an embedded method namely the dominant eigenvectors after the singular value decomposition (SVD). Finally, the nonlinear regression algorithm of the relevance vector machine (RVM) is applied to generalize prediction of quality scores to out of sample images. The preliminary results were reported in [13]. This paper presents the extended benchmarking experiments on both LIVE (release 2) and CSIQ image quality databases [14, 15] that provide the ground truth data (i.e. the subjective image quality scores published in the form of Difference Mean Opinion Scores-DMOS) as well as the test images from which the features vector is computed.

The remainder of the paper is organized as follows: proposed free-reference image quality framework is presented in section 2. Features extraction and selection processes are detailed in section 3. In section 4, a description of the employed regression module, namely the Relevance Vector Machine (RVM), is given. Simulations, performance benchmarking and validation of the proposed algorithm are provided in section 5 followed by concluding remarks in section 6.

2. PROPOSED FREE- REFERENCE IMAGE QUALITY FRAMEWORK

The perceived quality evaluation framework presented in this paper is shown in figure 1; the process is composed of the following steps:

1. Features extraction: features are extracted from images coming from image databases (LIVE II and CSIQ) using three NR-IQA (BRISQUE, BIQI, BLIINDS-II).
2. Features selection: this step consists of removing superfluous / redundant features and retaining only the relevant ones.
3. Construction of the final model: the selected features from the previous step are used as input to the machine learning algorithm (here Relevance Vector Machine: RVM) to build image quality prediction models. Once the final model is built, it comes the test phase where the test sample is used to validate this model.

3. FEATURES EXTRACTION AND SELECTION

In this section, we provide details of feature extraction and selection procedures.

3.1. Image Features Extraction

The image attributes used in this paper come from three learning-based NR-IQA metrics namely BRISQUE, BIQI and BLIINDS-II summarized in table 1 below. The size of the vectors of features is 36, 18 and 24, respectively. The blind metrics where these features come from are described as follows:

3.1.1. BRISQUE [6]

This metric does not require any transformation of the image. It directly extracts NSS features in the spatial domain. For each image, a generalized Gaussian distribution (GGD) is used to estimate the distribution, and then generates the parameters as resulted features. 18 features are extracted using 2 scales, resulting in 36 features used to evaluate the perceptual quality of an image.

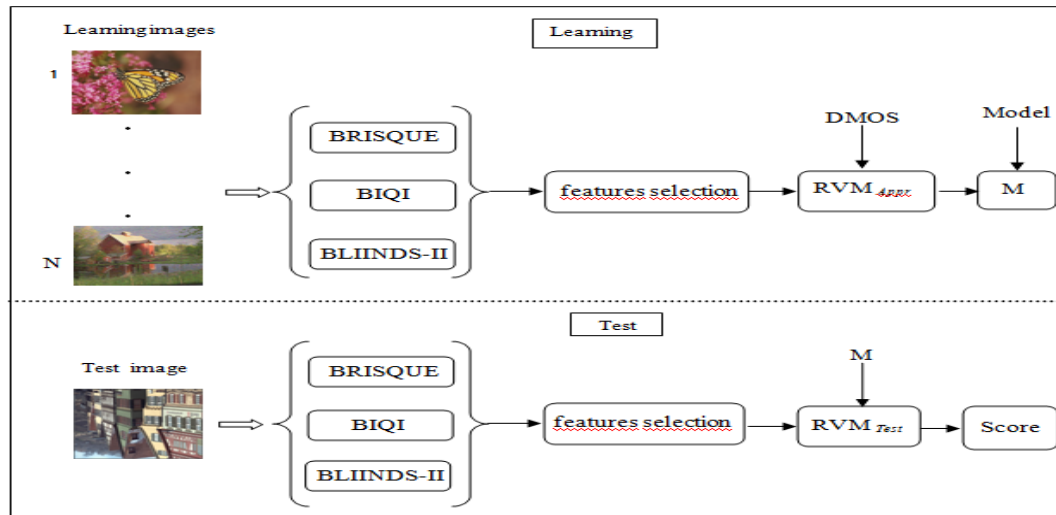


Figure 1. General scheme of the proposed NR-IQA algorithm

3.1.2. BIQI [2]

This algorithm is based on the extraction of NSS in the wavelet domain over three scales and three orientations. Three features are extracted (mean, variance and shape) and used to classify a distorted image into one of N distortions using support vector machine (SVM), then support vector regression (SVR) is used to predict quality score.

3.1.3. BLIINDS-II [5]

Presented by *Saad et al.*, this model works in the DCT domain. A total of 24 features are extracted from the block DCT domain and are affected by changing the type and the level of distortion. These features are then input to the Bayesian inference model to get the perceived quality estimate.

Table 1. NR-IQA metrics considered to investigate the relevance of features for perceptual quality judgment.

NR-IQA algorithm	Domain	Features		Regression/quality estimation
BRISQUE	Spatial domain	36	f_1, \dots, f_{36}	Support VectorRegression (SVR)
BIQI	DWT domain	18	f_{37}, \dots, f_{54}	RelevanceVectorMachine (RVM) + SVR
BLIINDS-II	DCT domain	24	f_{55}, \dots, f_{78}	Probabilistic model + SVR

As a first step, we build in a 78-D vector of original attributes by putting all the extracted features together.

3.2. Feature Selection Technique

All 78 descriptors previously discussed are extracted from LIVE image database release 2 (LIVE II) described in section 5.1. In order to eliminate redundant and irrelevant features and select only useful ones, we used Singular Value Decomposition (SVD), which is one among a large array of techniques used for dimensionality reduction.

SVD decomposes a $M(m \times n)$ matrix into three matrices as:

$$M = USVT \tag{1}$$

where U and V are two orthogonal matrices of $(m \times p)$ and $(n \times p)$ dimensions, respectively. S is a $(p \times p)$ diagonal matrix.

p is called the rank of matrix M .

The diagonal positive entries of matrix S are called singular values of M . These values are arranged in descending order of their magnitude.

For feature selection, we used the same algorithm as the column select problem [16]. This algorithm can be summarized in the following steps:

- i. Input the matrix where rows are images and columns are features.
- ii. Compute the centralized data.
- iii. Apply SVD to get the main components.
- iv. Get the dimensions having most of the variation (select the dominant eigenvectors, e.g. representing the 95% of the data).
- v. Compute leverage scores using the dominant eigenvectors of the principal components (.i.e. the norm of the eigenvector's coefficients).
- vi. Sort the leverage scores in descending order.
- vii. Get the indices of the vectors with the largest leverage scores.

In this paper, we select the features which have a leverage score greater than or equal to 0.4. Figure 2 shows the resulting selected features with their leverage scores. We can note that the most significant features come from BLIINDS-II quality metric.

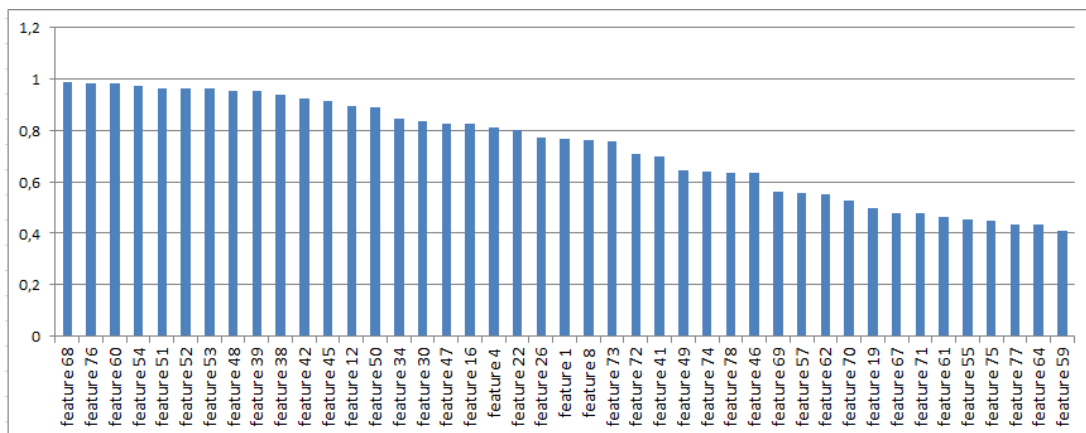


Figure 2. The most significant features with their leverage scores

4. PREDICTION MODEL

In this work, Relevance Vector Machine (RVM) [17] is employed as prediction model instead of support vector machine (SVM) [18, 19], which is the most common. This choice is made based on the benefits the RVMs offer over the SVMs, mainly probabilistic predictions and automatic estimation of the hyper-parameters.

For a given set of samples $\{x_i, t_i\}_{i=1}^N$ where x_i is the input variable vector, t_i is the target value, N is the length of training data. The RVM regression expression is:

$$t(x) = \sum_{i=1}^N w_i K(x, x_i) + w_0 + \varepsilon_n \quad (2)$$

where N is the number of data points, $w=[w_1, \dots, w_N]$ is weight, w_0 is bias, $K(x, x_i)$ is kernel function and $\varepsilon_n = N(0, \sigma^2)$ is error term with zero mean Gaussian process and variance σ^2 . Usually, the Gaussian Kernel is preferred, and its formula is:

$$K(x, x_i) = \exp[-(x - x_i)^T (x - x_i) / 2S^2] \quad (3)$$

where S^2 is the width. Assuming that the samples $\{x_i, t_i\}_{i=1}^N$ are independently generated. The likelihood of all data set can be written as follows:

$$P(t | w, \sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left\{-\frac{1}{2\sigma^2} \|t - \Phi w\|^2\right\} \quad (4)$$

Here, Φ is a design matrix having the size $N \times (N + 1)$ with:

$$\Phi(x_i) = [1, K(x_i, x_1), K(x_i, x_2), \dots, K(x_i, x_N)]^T \quad (5)$$

The highest probability estimation of w and σ^2 of equation (4) may suffer from serious over-fitting. To solve this issue, Tipping [17] imposed an explicit zero-mean Gaussian prior probability distribution for the weights, w , with diagonal covariance of α as follows:

$$P(w | \alpha) = \prod_{i=0}^N N(w_i | 0, \alpha_i^{-1}) \quad (6)$$

where α is a vector of $N+1$ named hyper parameters.

In this way, using Baye's rule, the posterior over all unknowns could be calculated given the defined non informative prior distribution:

$$P(w, \alpha, \sigma^2 | t) = \frac{P(t | w, \alpha, \sigma^2) \cdot P(w, \alpha, \sigma^2)}{\int P(t | w, \alpha, \sigma^2) P(w, \alpha, \sigma^2) dw d\alpha d\sigma^2} \quad (7)$$

Full analytical solution of this integral (7) is obdurate. Thus, decomposition of the posterior according to:

$$P(w, \alpha, \sigma^2 | t) = P(w | t, \alpha, \sigma^2) P(\alpha, \sigma^2 | t) \quad (8)$$

is called upon to ease the solution [17]. The posterior distribution over the weights is calculated using Bayes rule and is given by:

$$P(w | t, \alpha, \sigma^2) = \frac{P(t | w, \sigma^2) P(w | \alpha)}{P(t | \alpha, \sigma^2)} \quad (9)$$

The resulting posterior distribution over the weights is the multivariate Gaussian distribution:

$$P(w | t, \alpha, \sigma^2) = N(\mu, \Sigma) \quad (10)$$

where the mean and the covariance are respectively expressed by:

$$\Sigma = (\sigma^{-2} \Phi^T \Phi + A)^{-1} \quad (11)$$

$$\mu = \sigma^{-2} \Sigma \Phi^T t \quad (12)$$

with diagonal $A = \text{diag}(\alpha_0, \dots, \alpha_N)$.

For uniform hyper priors over α and σ^2 , one requires only to maximize the term $(t/\alpha, \sigma^2)$:

$$P(t/\alpha, \sigma^2) = \int P(t/w, \sigma^2) P(w, \alpha) dw \quad (13)$$

$$P(t/\alpha, \sigma^2) = \left[(2\pi)^{-N} / \sqrt{|\sigma^2 + \Phi A^{-1} \Phi^T|} \right] \times \exp \left\{ -\frac{1}{2} t^T (\sigma^2 + \Phi A^{-1} \Phi^T)^{-1} t \right\} \quad (14)$$

By simply forcing the derivatives of Equation (14) to zero, we can get the re-estimation formulas for α and σ^2 respectively as follow:

$$\alpha_i^{new} = \frac{1 - \alpha_i \Sigma_{ii}}{\mu_i^2} \quad (15)$$

$$(\sigma^2)^{new} = \frac{\|t - \Phi \mu\|^2}{N - \sum_i (1 - \alpha_i \Sigma_{ii})} \quad (16)$$

5. EXPERIMENTAL SETUP

5.1. Datasets

Our simulations are carried out over two datasets: LIVE (release 2) and CSIQ image visual quality assessment databases.

5.1.1. LIVE II image quality database

LIVE image database was developed at the Laboratory for Image and Video Engineering in collaboration with the Center for Perceptual Systems at the University of Texas at Austin, USA. The first release was made available online in 2003 while release 2 on which we conducted the present study was published in 2005 [14].

LIVE II was built upon 29 high resolution color images. These images were subjected to five distortion types: JPEG, JPEG2000, white noise in the RGB components, Gaussian blur, and transmission errors in the JPEG2000 bit stream using a fast-fading Rayleigh channel model. This resulted in 982 test images for which subjective testing was performed in seven sessions. According to the adjectival categorical judgment method, during the psychometric experiments, each observer was shown the images randomly one at a time (single stimulus) and was asked to assign each image with an adjective that indicates his/her perception of the quality of images. The quality adjectives (bad, poor, fair, good, excellent) are then converted to the corresponding discrete numerical values from 1 to 5 respectively.

5.1.2. CSIQ image quality database

The Categorical Subjective Image Quality (CSIQ) database was developed at the Image Coding Analysis Laboratory at Oklahoma State University, USA [15]. It was derived from 30 color high resolution square images that were distorted using six different image processing algorithms including JPEG and JPEG2000 compression, Gaussian blurring, Additive Gaussian white noise, Global contrast decrements, and Additive Gaussian pink noise. Thus, 900 distorted images have been generated out of which the subjective ratings of only 866 test images are provided.

Visual dissimilarity measurements were performed based on a linear displacement strategy. This consists of presenting simultaneously all the test versions of an image across a monitor array. Observers are then asked to place these images so that the horizontal distance between two test images reflects the perceived dissimilarity between them.

The two datasets descriptions are summarized in Table 2 below.

5.1.3. Ground Truth Data

Raw subjective scores that correspond to the perceptual quality judgment results obtained after the psycho-visual experiments are usually converted to Mean Opinion Scores (MOS), with subject reliability of 95% confidence interval, as recommended by the Video Quality Experts Group (VQEG) Phase I FR-TV [20]. Higher value of MOS corresponds to higher visual quality of the image.

In both the two image quality databases described above, the original undistorted images are subjectively assessed as well. Raw scores were converted to Difference Mean Opinion Scores (DMOS) that represent the difference between the MOS obtained for the reference image and its test version. A low DMOS means little degradation whereas an important value corresponds to severe distortions in the image.

Table 2. Summary of LIVE II and CSIQ databases descriptions.

Database		LIVE II	CSIQ
Image format		24-bits/pixel RGB color (768x512)	24-bits/pixel RGB color (512x512)
Number of reference images		29	30
Distortions	JPEG2000	227	150
	JPEG	233	150
	Gaussian Blur	174	150
	Noise	174	150
	Others	174	266
Total number of images		982	866
observers		20-29	35
Scores' scale of DMOS		0 : for undistorted images 1 .. 100 : for distorted images	0 .. 9

5.2. Performance Evaluation Protocol

The performance of our metric is benchmarked using three criteria: Pearson Correlation Coefficient (PCC), Spearman Rank Order Correlation Coefficient (SROCC) and Root Mean Square Error (RMSE) all computed between subjective and objective scores. The first criterion gives estimation about the prediction linear correlation, the second measures the prediction monotonicity and the third provides estimation about accuracy of the quality metric. The lower values of RMSE the better accuracy is, contrary to PCC and SROCC where the highest values indicate the best results.

Before computing PCC, a nonlinear mapping between subjective DMOS and objective scores is carried out using the logistic function with five parameters [21]. The expression of the quality score which is the predicted DMOS is given by:

$$DMOS_p = \beta_1 \text{logistic}(\beta_2, DMOS - \beta_3) + \beta_4 DMOS + \beta_5 \quad (17)$$

where $DMOS$ and $DMOS_p$ are the predicted scores before and after regression, respectively. β_1 to β_5 are the regression parameters estimated using *fminsearch* function in Matlab's optimization Toolbox. The logistic function is given by equation (18) below:

$$\text{logistic}(\tau, DMOS) = \frac{1}{2} - \frac{1}{1 + \exp(\tau DMOS)}$$

Images of the datasets are randomly split into two non-overlapping sets; 80% for training and the remaining 20% for test phase. This random splitting is repeated 100 times in order to ensure the robustness of our metric because of the limited size of the samples (see Table 2). At the end, we calculate the average of the obtained performance criteria.

5.3. Simulation Results

5.3.1. Preliminary results

In this sub-section, we present the preliminary simulation results reported in [13]. Tables 3, 4 and 5 show PCC, SROCC and RMSE mean values between subjective and objective scores on each of the five distortion subsets and the entire database (noted by ALL). These values are compared to six state-of-the-art general-purpose NR-IQA metrics (BIQI, BLIINDS, DIIVINE, BLIINDS-II, BRISQUE and NIQE).

Table 3.PCCof different methods on LIVE II database.

	JPEG2k	JPEG	WN	G-Blur	FF	ALL
BIQI	0.750	0.630	0.968	0.800	0.722	0.740
BLIINDS	0.807	0.597	0.914	0.870	0.743	0.680
DIIVINE	0.922	0.921	0.988	0.923	0.888	0.917
BLIINDS-II	0.963	0.979	<u>0.985</u>	0.948	0.944	0.923
BRISQUE	0.923	<u>0.974</u>	0.985	0.951	0.903	0.942
NIQE	0.937	0.956	0.977	<u>0.953</u>	<u>0.913</u>	0.915
Proposed	<u>0.962</u>	0.945	0.981	0.957	0.911	0.953

Table 4.SROCCof different methods on LIVE II database.

	JPEG2k	JPEG	WN	G-Blur	FF	ALL
BIQI	0.736	0.591	0.958	0.778	0.700	0.726
BLIINDS	0.805	0.552	0.890	0.834	0.678	0.663
DIIVINE	0.913	0.910	0.984	0.921	0.863	0.916
BLIINDS-II	0.951	<u>0.942</u>	0.978	0.944	0.927	0.920
BRISQUE	0.914	0.965	0.979	0.951	0.877	<u>0.940</u>
NIQE	0.917	0.938	0.966	0.934	0.859	0.914
Proposed	<u>0.949</u>	0.924	<u>0.982</u>	<u>0.946</u>	<u>0.884</u>	0.955

Table5.RMSE of different methods on LIVE II database.

	JPEG2k	JPEG	WN	G-Blur	FF	ALL
BIQI	16.540	24.580	6.930	11.100	19.480	18.360
BLIINDS	14.780	25.320	11.270	9.080	18.620	20.010
DIIVINE	<u>9.660</u>	<u>12.250</u>	<u>4.310</u>	<u>7.070</u>	<u>12.930</u>	<u>10.900</u>
Proposed	5.478	6.535	1.975	5.320	7.395	6.8364

Numerical results show that the proposed general-purpose no-reference image quality assessment metric achieves good performances in terms of correlation (table 3), monotonicity (table 4) and accuracy (table5). The first position best results are mentioned in bold whereas the second positions best results are the underlined values. We can notice that the proposed metric gives the first or the second- best performance for all the subsets except that of the encoded images via JPEG algorithm.

Furthermore, the scatter plot of our method for test set with median SROCC is provided in figure 3; where the horizontal axis corresponds to the objective scores and the vertical axis corresponds to subjective scores. Every dot in the plot represents an image in the database. It can be seen that most of the dots are clustered around the red line that represent ideal linear correlation line "Proposed=DMOS", this means that the proposed metric achieves good correlation with human scores.

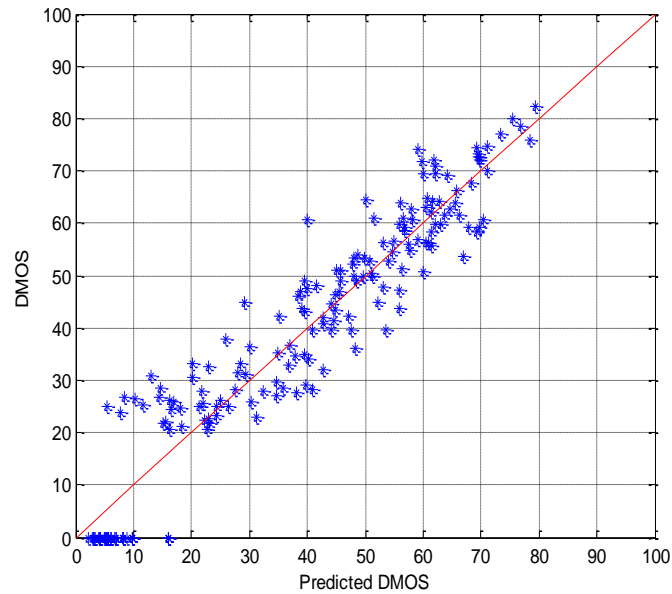


Figure 3. The scatter plots of the predicted perceived quality vs. the DMOS

5.3.2. Extended results

In order to confirm that the performance of our metric does not depend on the learning datasets and still achieves competitive predictive performances, additional experiments have been carried out. The training and testing phases of the learning process have been performed alternately on two distinct image quality datasets, namely LIVE (release 2) and CSIQ introduced in section 5.1. The common points between the two databases are: (1) the four common types of degradations are JPEG, JPEG2000, blur, and noise(2) the subjective scores are DMOS values. Nevertheless, the scales of DMOS are different as can be noticed from Table 2. In this case, we map the subjective scores of the training database into the scale of the subjective scores of the test database.

Since preliminary results show that the proposed metric clearly outperforms BLIINDS, DIVINE and NIQE algorithms, we compare extended version of experiments to only the competing no-reference image quality measures namely BIQI, BLIINDS II and BRISQUE. The final comparison results are shown in sub-sections below.

5.3.2.1. Training on LIVE II / Testing on CSIQ

At this stage, the training process is performed on the LIVE II dataset while the tests are made over the CSIQ dataset. Tables 6 to 8 show respectively the mean values of correlation, monotonicity and accuracy descriptors between subjective and objective scores on each of the four common distortion subsets (JPEG2000, JPEG, noise and Gaussian blur) as well as the entire database (noted by ALL). These values are compared to the competing NR-IQA metrics (BIQI, BLIINDS-II and BRISQUE).

Table 6. PCCof different methods trained on LIVE II database and tested on CSIQ.

	JPEG2k	JPEG	Noise	G-Blur	ALL
BIQI	0.665	0.502	0.553	0.867	0.623
BLIINDS-II	0.921	<u>0.914</u>	<u>0.780</u>	0.924	0.885
BRISQUE	<u>0.908</u>	0.935	0.918	<u>0.919</u>	<u>0.883</u>
Proposed	<u>0.914</u>	<u>0.882</u>	<u>0.877</u>	<u>0.921</u>	<u>0.881</u>

Table 7.SROCCof different methods trained on LIVE II database and tested on CSIQ.

	JPEG2k	JPEG	Noise	G-Blur	ALL
BIQI	0.652	0.482	0.547	0.761	0.587
BLIINDS-II	0.893	<u>0.856</u>	<u>0.743</u>	0.892	0.855
BRISQUE	<u>0.886</u>	0.894	0.907	<u>0.888</u>	<u>0.853</u>
Proposed	<u>0.884</u>	<u>0.834</u>	<u>0.860</u>	<u>0.890</u>	<u>0.851</u>

Table 8.RMSE of different methods trained on LIVE II database and tested on CSIQ.

	JPEG2k	JPEG	Noise	G-Blur	ALL
BIQI	0.236	0.264	0.139	0.142	0.220
BLIINDS-II	0.122	<u>0.124</u>	<u>0.105</u>	0.109	0.131
BRISQUE	<u>0.132</u>	0.108	0.066	<u>0.112</u>	<u>0.132</u>
Proposed	<u>0.127</u>	<u>0.143</u>	<u>0.080</u>	<u>0.111</u>	<u>0.133</u>

5.3.2.2. Training on CSIQ / Testing on LIVE II

Here, the training process is performed on the CSIQ dataset while the tests are made over the LIVE II dataset. Tables 9 to 11 show respectively the mean values of PCC, SROCC and RMSE between desired and automatically calculated scores on each of the four common distortion subsets (JPEG2000, JPEG, noise and Gaussian blur) as well as the entire database (noted by ALL). These values are compared to the competing NR-IQA metrics (BIQI, BLIINDS-II and BRISQUE).

Table 9. PCCof different methods trained on CSIQ database and tested on LIVE II.

	JPEG2k	JPEG	Noise	G-Blur	ALL
BIQI	0.585	0.451	0.711	0.746	0.402
BLIINDS-II	0.910	<u>0.881</u>	<u>0.879</u>	<u>0.925</u>	0.848
BRISQUE	<u>0.846</u>	0.905	0.904	0.931	<u>0.835</u>
Proposed	<u>0.816</u>	<u>0.876</u>	<u>0.857</u>	<u>0.884</u>	<u>0.803</u>

Table 10.SROCCof different methods trained on CSIQ database and tested on LIVE II.

	JPEG2k	JPEG	Noise	G-Blur	ALL
BIQI	0.579	0.431	0.359	0.570	0.379
BLIINDS-II	0.907	<u>0.843</u>	<u>0.855</u>	<u>0.909</u>	0.845
BRISQUE	<u>0.842</u>	0.873	0.918	0.932	<u>0.831</u>
Proposed	<u>0.810</u>	<u>0.854</u>	<u>0.788</u>	<u>0.865</u>	<u>0.796</u>

Table 11. RMSE of different methods trained on CSIQ database and tested on LIVE II.

	JPEG2k	JPEG	Noise	G-Blur	ALL
BIQI	0.197	0.214	0.168	0.157	0.221
BLIINDS-II	0.100	<u>0.113</u>	<u>0.114</u>	<u>0.089</u>	0.128
BRISQUE	<u>0.129</u>	0.102	0.102	0.085	<u>0.133</u>
Proposed	<u>0.140</u>	<u>0.115</u>	<u>0.123</u>	<u>0.110</u>	<u>0.144</u>

Algorithm performances are computed with respect to linear correlation (tables 6 and 9), monotonicity (tables 7 and 10) and accuracy (tables 8 and 11). High values of PCC and SROCC and low values of RMSE indicate the best results. In tables 6 to 11, first position best performances are mentioned in bold whereas the second and three positions best results are the underlined values. From results presented above, we can note that our no-reference image quality assessment metric achieves competitive predictive performances with BLIINDS-II and BRISQUE metrics when the training and the test phases of the learning process are applied alternately on CSIQ and LIVE-release2 databases.

Extended validation results corroborate well with the preliminary simulations tests. It has also been proven that the predictive performances of the proposed algorithm do not depend on the learning datasets.

6. CONCLUSION

The two main ideas of this work is that the most successful NR-IQA metrics are based on NSS features and that the multi-domain information simulate well the hierarchical structure of the visual cortex perception. For these reasons, the features used to build the present NR-IQA metric are collected from three NR-IQA methods based on NSS features operating in three different domains (spatial, DWT and DCT). Only pertinent features are input to the relevance vector machine algorithm to predict the objectives score. The step of dimensionality reduction is achieved using Singular Value Decomposition (SVD) based dominant eigenvectors. Numerical experiments show that the proposed metric is tightly competitive with BLIINDS II and BRISQUE methods. It also outperforms BLIINDS, BIQI, DIVINE and NIQE algorithms as it is independent from the datasets on which the learning process is applied.

REFERENCES

- [1] D. Zhang, Y. Ding, N. Zheng, "Nature scene statistics approach based on ICA for no-reference image quality assessment", Proceedings of International Workshop on Information and Electronics Engineering (IWIEE), 29 (2012), 3589- 3593.
- [2] A. K. Moorthy, A. C. Bovik, A two-step framework for constructing blind image quality indices[J], IEEE Signal Process. Lett., 17 (2010), 513-516.
- [3] L. Zhang, L. Zhang, A.C. Bovik, A Feature-Enriched Completely Blind Image Quality Evaluator, IEEE Transactions on Image Processing, 24(8) (2015), 2579- 2591.
- [4] M.A. Saad, A.C. Bovik, C. Charrier, A DCT statistics-based blind image quality index, Signal Process. Lett. 17 (2010) 583–586.
- [5] M. A. Saad, A. C. Bovik, C. Charrier, Blind image quality assessment: A natural scene statistics approach in the DCT domain, IEEE Trans. Image Process., 21 (2012), 3339-3352.

- [6] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, *IEEE Trans. Image Process.* 21 (2012), 4695 - 4708.
- [7] A. Mittal, R. Soundararajan, A. C. Bovik, Making a completely blind image quality analyzer, *IEEE Signal Process. Lett.*, 20 (2013), 209-212.
- [8] N. Kruger, P. Janssen, S. Kalkan, M. Lappe, A. Leonardis, J. Piater, A. Rodriguez-Sanchez, L. Wiskott, "Deep hierarchies in the primate visual cortex: What can we learn for computer vision?", *IEEE Trans. Pattern Anal. Mach. Intell.*, 35 (2013), 1847–1871.
- [9] D. J. Felleman, D. C. Van Essen, "Distributed hierarchical processing in the primate cerebral cortex," "Distributed hierarchical processing in the primate cerebral cortex,"
- [10] B. Sadou, A. Lahoulou, T. Bouden, A New No-reference Color Image Quality Assessment Metric in Wavelet and Gradient Domains, 6th International Conference on Control Engineering and Information Technologies, Istanbul, Turkey, 25-27 October (2018), 954-959.
- [11] Q. Wu, H. Li, F. Meng, K. N. Ngan, S. Zhu, No reference image quality assessment metric via multi-domain structural information and piecewise regression. *J. Vis. Commun. Image R.*, 32(2015), 205–216.
- [12] X. Shang, X. Zhao, Y. Ding, Image quality assessment based on joint quality-aware representation construction in multiple domains, *Journal of Engineering* 4 (2018), 1-12.
- [13] B. Sadou, A.Lahoulou, T.Bouden, A.R. Avila, T.H. Falk, Z. Akhtar, "Blind Image Quality Assessment Using Singular Value Decomposition Based Dominant Eigenvectors for Feature Selection", 5th Int. Conf. on Signal and Image Processing (SIPRO'19), Toronto, Canada, pp. 233-242, 2019.
- [14] H. R. Sheikh, Z. Wang, L. Cormack, A. C. Bovik, LIVE Image Quality Assessment Database Release 2 (2005), <http://live.ece.utexas.edu/research/quality>
- [15] E. Larson, D. M. Chandler, Categorical image quality assessment (CSIQ) database. <http://vision.okstate.edu/?loc=csiq>
- [16] M. W. Mahoney, P. Drineas, "CUR matrix decompositions for improved data analysis," in *Proc. the National Academy of Sciences*, February 2009.
- [17] M.E. Tipping. The relevance vector machines. In *Advances in Neural Information Processing Systems 12*, Solla SA, Leen TK, Muller K-R (eds). MIT Press: Cambridge, MA (2000), 652-658.
- [18] D. Basak, S. Pal, D.C. Patranabis, Support vector regression, *Neural Information Processing – Letters and Reviews*, 11 (2007).
- [19] B. SchÖlkopf, A.J. Smola, *Learning with Kernels*. MIT press, Cambridge, (2002).
- [20] Final VQEG report on the validation of objective quality metrics for video quality assessment: http://www.its.bldrdoc.gov/vqeg/projects/frtv_phaseI/
- [21] H. R. Sheikh, M. F. Sabir, A. C. Bovik, A statistical evaluation of recent full reference image quality assessment algorithms, *IEEE Trans. Image Process.*, 15 (2006), 3440–3451.

AUTHORS

BesmaSadou Is currently a PhD student in the department of Electronics at university of Jijel (Algeria). She also works as full-time teacher of mathematics at the middle school. Her research interests are focused on reduced and no-reference image quality assessment.

AtidelLahoulou Is Doctor in Signals and Images from Sorbonne Paris Cité (France) since 2012. She earned her Habilitation Universitaire in 2017 and is currently associate professor in the department of computer science at university of Jijel (Algeria). She is also head of research team at LAOTI laboratory (Jijel). Her research interests include multimedia quality evaluation and enhancement, biometrics, machine learning and cyber security.

ToufikBouden Received his PhD degree in automatics and signal processing from Electronics Institute of Annaba University (Algeria) in 2007. He was the head of Non Destructive Testing Laboratory from 2009 until 2017. Since 2015, he is full professor in the department of Automatics, University of Jijel, Algeria. His areas of research are signal and image processing, non-destructive testing and materials characterization, biometrics, transmission security and watermarking, chaos, fractional system analysis, synthesis and control.

Anderson R. Avila Received his B.Sc. in Computer Science from Federal University of Sao Carlos, Brazil, in 2004 and his M.Sc in Information Engineering from Federal University of ABC in 2014. In October 2013, Anderson worked as a short-term visiting researcher at INRS, where he now pursues his Ph.D degree on the topic of speaker and emotion recognition. His research interests include pattern recognition and multimodal signal processing applied to biometrics.

Tiago H. Falk Is an Associate Professor at INRS-EMT, University of Quebec and Director of the Multimedia Signal Analysis and Enhancement (MuSAE) Lab. His research interests are in multimedia quality measurement and enhancement, with a particular focus on human-inspired technologies.

Zahid Akhtar Is a research assistant professor at the University of Memphis (USA). Prior to joining the University of Memphis, he was a postdoctoral fellow at INRS-EMT-University of Quebec (Canada), University of Udine (Italy), Bahcesehir University (Turkey), and University of Cagliari (Italy), respectively. Dr. Akhtar received a PhD in electronic and computer engineering from the University of Cagliari (Italy). His research interests are biometrics, affect recognition, multimedia quality assessment, and cybersecurity.