Comparison of No-Reference Image Quality Assessment Machine Learning-based Algorithms on Compressed Images

Christophe Charrier, Abdelhakim Saadane, Christine Fernandez-Maloigne

Comparison of No-Reference Image Quality Assessment Machine Learning-based Algorithms on Compressed Images

Christophe Charrier[1] AbdelHakim Saadane[2] Christine Fernandez-Maloigne[3]

[1] Université de Caen-Basse Normandie, ENSICAEN, GREYC UMR CNRS 6072
[2] Université de Nantes, XLIM-SIC UMR CNRS 7252
[3] Université de Poitiers, XLIM-SIC UMR CNRS 7252

ABSTRACT

No-reference image quality metrics are of fundamental interest as they can be embedded in practical applications. The main goal of this paper is to perform a comparative study of seven well known no-reference learning-based image quality algorithms. To test the performance of these algorithms, three public databases are used. As a first step, the trial algorithms are compared when no new learning is performed. The second step investigates how the training set influences the results. The Spearman Rank Ordered Correlation Coefficient (SROCC) is utilized to measure and compare the performance. In addition, an hypothesis test is conducted to evaluate the statistical significance of performance of each tested algorithm.

Keywords: No-reference quality assessment, training, machine learning-based algorithms, compressed images.

1. INTRODUCTION

Lossy image compression techniques such as JPEG2000 allow high compression rates, but only at the cost of some perceived degradations in image quality. The way to evaluate the performance of any compression scheme is a crucial step, and more precisely available ways to measure the quality of compressed images. There is a very rich literature on image quality criteria, generally dedicated to specific applications (optics, detector, compression, restoration, ...).

From several years, a number of researches have been conducted to design robust No-Reference Image Quality Assessment (NR-IQA) algorithms, claiming to have made headway in their respective domains. NR-IQA algorithms generally follow one of three trends. 1) Distortion-specific approaches: they assume a prior knowledge of the degradations involved and employ a specific distortion model to drive an objective algorithm to predict a subjective quality score. These algorithms quantify one or more distortions such as blockiness,[1] blur,[2–4] or ringing[5] and score the image accordingly. 2) Training-based approaches: they train a model to predict the image quality score based on a number of features extracted from the image.[6–10] Algorithms following this approach often use a large number of features without perceptually justifying each of them. 3) Natural scene statistics (NSS) approaches: they rely on the hypothesis that images of the natural world (*i.e.* distortion free images) occupy a small subspace of the space of all possible images and seek to find a distance between the test image and the subspace of natural images.[10–12]

Training-based approaches are as reliable as the features used to train the learning model. For example, DIIVINE[13] is a recent wavelet-based algorithm using a two-stage framework, where the distortion type is predicted first and then, based on this prediction, image quality is estimated. DIIVINE uses a Support Vector Machine (SVM) to classify an image into a distortion class and Support Vector Regression (SVR) to predict quality scores. A large number of features are used for classification and for quality score prediction (88 features) to achieve high performance against human rates. Jung *et al.* in[14] propose a no-reference IQA method that is based on training a neural network. The method proposed can be described in four steps. Initially, a learning set of images is built to be used for the model calibration. The set consists of distorted images subjectively rated by observers

Further author information: (Send correspondence to C.C.)
C.C.: E-mail: christophe.charrier@unicaen.fr
A.S.: E-mail: abdelhakim.saadane@univ-nantes.fr
C.F-M.: E-mail: christine.fernandez@univ-poitiers.fr

as well as by an objective full-reference image quality assessment index. Step two consists of mathematically characterizing the image defects. Next, the neural network is trained, and then applied to a set of test images.

In,[15] an interactive neuro-evolution approach is proposed as a way to learn the properties of a users visual perception. Interactive evolution algorithms are based on evolutionary computational algorithms that invoke probabilistic, heuristic search techniques inspired from principles of heredity, mutability, and natural selection.

In,[6] a no-reference IQA method is described based on object/region detection and training a radial based neural network. The approach can be summarized in two steps; 1) an object/region detection algorithm is applied to an image, and then 2) the spectrum distribution of a detected region is compared with an empirical model to determine a quality score for the detected region. The signal features that are used for object/region detection are chosen to avoid overlap with those features that are used for quality assessment. The features for quality assessment are chosen to take into account recurrent problems in consumer photography such as blurriness (including poor focus and motion blur), under- and over-exposure, and noise (including Gaussian white noise and salt and pepper noise).

In,[8] a NR method for assessing JPEG image quality is proposed, using a sequential learning algorithm to grow and prune a radial basis function (GAP-RBF) neural network. The metric is designed to account for the 8×8 blocking artifacts that arise from JPEG encoding. This approach proceeds by extracting features to quantify the encoding distortions. These features are then used in conjunction with the GAP-RBF network model to predict image quality.

Moorthy *et al.* in[10] propose a NR IQA approach based on a two-stage process. The first stage of the approach is essentially a classification stage in which the test image is assigned a probability p_i of belonging to a set of distortion classes $i = 1, \cdots, N$, where N is the number of distortion classes. The second stage then performs an SVM-based quality assessment in a specific distortion class.

Concerning NSS-based algorithms, many of them apply a combination of learning-based approach with features extracted from natural scene statistics . Natural images, like most other natural signals, are highly heterogeneous and variable; yet despite this variability, they contain a large number of statistical regularities.

Visual systems have evolved to exploit these regularities so that the eye can accurately encode retinal images and the brain can correctly interpret them. Thus, characterizing the structure of natural images is critical for understanding visual encoding and decoding in biological vision systems and for applications in image processing and computer vision.

In,[16] Saad *et al.* propose a framework that derives entirely from a simple statistical model of local DCT coefficients. The developed algorithm, namely, BLIINDS-II (BLind Image Integrity Notator using DCT Statistics), greatly improves upon a preliminary algorithm (BLIINDS-I),[17] which uses no statistical modeling and a different set of sample DCT statistics. BLIINDS-I was a successful experiment to determine whether DCT statistics could be used for blind IQA. BLIINDS-II fully unfolds this possibility and provides a leap forward in both performance and in the use of an elegant and general underlying statistical model. A generalized NSS-based model of local DCT coefficients is then derived, and transformed the model parameters into features used for perceptual image quality score prediction. A generalized probabilistic model is obtained for these features, and is used to make probabilistic predictions of visual quality.

In this paper, we present an extensive comparative study of well-known NR-IQA training-based algorithms. Section 2 summarizes the used trial NR-IQA algorithms and the used trial databases. It also specifies how the effectiveness of algorithms is evaluated. As a first step of this study, section 3 compares the trial algorithms when no new learning phase is performed. Section 4 deals with the second step and investigates how the training set influences the results.

2. NO-REFERENCE IQA ALGORITHMS COMPARISON SETUP

2.1 Trial NR-IQA algorithms

The tested NR-IQA algorithms are 1) BIQI,[10] 2) DIIVINE[13] , 3) BLIINDS,[17] 4) BLIINDS-II,[16] 5) BIQ-Anisotropy[11] 6) BRISQUE[18] and 7) NIQE.[19] They are summarized in Table 1. These algorithms, which are

NR-IQA algorithm	comments
BIQI	Machine learning-based approach (SVM).
DIIVINE	Classification (SVM) and Regression (SVR).
BLIINDS	Machine learning-based approach (Probalistic model).
BLIINDS-II	Machine learning-based approach (Probalistic model).
BIQ-Anisotropy	Renyi entropy measure along various orientations.
BRISQUE	Natural scene statistic-based distortion-generic.
NIQE	Space domain natural scene statistic model.

Table 1. NR-IQA trial Algorithms

usually used in comparative studies are selected here mainly because their implementations are publicly available either on the Internet or from the authors. Even if it seems that it is ineffective to analyze only brightness component,[20] all of these algorithms are luminance based.

2.2 Trial databases

To provide comparison of NR-IQA algorithms, three publicly available databases are used: 1) LIVE database,[21] 2) TID2008 database[22] and 3) CSIQ image database.[23] The LIVE database contains 29 original images on which five kinds of distortions have been applied to generate 770 degraded images. The TID2008 database contains 1700 distorted versions of 25 original images (from Kodak Lossless True Color Image Suite) applying 17 distortion types. The CSIQ database consists of 30 original images, each distorted using six types of distortions at four to five different levels.

2.3 Statistical Significance and hypothesis testing

The Spearman Rank Order Correlation Coefficient (SROCC) is computed between the subjective values and the predicted scores obtained from trial NR-IQA algorithms.

In addition, to ascertain which differences between NR-IQA schemes performance are statistically significant, we applied an hypothesis test using the residuals between the DMOS values and the ratings provided by the trial IQA algorithms. This test is based on the t-test that determines whether two population means are equal or not. This test yields us to take a statistically-based conclusion of superiority (or not) of an NR-IQA algorithm.

3. ORIGINAL SETUP NR-IQA ALGORITHMS COMPARISON

As a first step, all trial algorithms have been compared without performing any new learning phase.

To be able to compare obtained results, only degradations that are common within the three image databases are considered. This yields us to judge the influence of many factors such as the training set, the images used to design databases, levels of degradation, etc.

Table 2 shows obtained SROCC for LIVE database and NR-IQA trial algorithms. Tables 3 and 4 give similar results for TID2008 database and CSIQ image database respectively. From these results, three main observations can be made. The first one is that the performance of each metric are dependent on the used database. This is particularly true for BRISQUE when applied to Gaussian noise. It shows an SROCC that varies from 0.252 (CSIQ database) to 0.986 (LIVE database). Another example illustrating the variation in performance is that of BLIINDS that displays for JPEG degradation an SROCC of 0.055, 0.264 and 0.839 for TID2008, CSIQ and LIVE respectively. This first observation suggests that particular attention should be paid to the selection of databases. The three used databases, as stated in,[20] are probably not well suited to the NR metrics evaluation. The second observation deals with the performance comparison of each metric for the three used databases. Except for BIQ-An, a higher SROCC is always observed for the LIVE database. This behavior was somewhat expected since all metrics have been trained on LIVE subsets. The third and last observation deals with the general behavior of each metric. In this case, one can notice that three schemes perform best for all used databases: DIIVINE, BLIINDS-II and BRISQUE. For the three common distortions (JP2K, JPEG, and Gaussian blur), BLIINDS-II outperforms both DIIVINE and BRISQUE.

LIVE subset	BIQI	DIIVINE	BLIINDS	BLIINDS-II	BIQ-An	NIQE	BRISQUE
JP2K	0.802	0.913	0.912	0.943	0.173	0.906	0.938
JPEG	0.879	0.910	0.839	0.954	0.086	0.847	0.923
White Noise	0.958	0.984	0.974	0.980	0.686	0.975	0.986
Gaussian Blur	0.821	0.921	0.957	0.941	0.595	0.945	0.978
Fast fading	0.730	0.863	0.750	0.933	0.541	0.882	0.929
Entire db	0.824	0.916	0.799	0.912	0.323	0.877	0.941

Table 2. SROCC values between actual subjective scores and predicted ones for the LIVE database without specific training.

TID2008 subset	BIQI	DIIVINE	BLIINDS	BLIINDS-II	BIQ-An	NIQE	BRISQUE
JP2K	0.694	0.853	0.821	0.901	0.672	0.897	0.866
JPEG	0.844	0.631	0.055	0.878	0.078	0.864	0.799
Add. Gaussian Noise	0.738	0.809	0.334	0.663	0.055	0.780	0.821
Gaussian blur	0.747	0.824	0.663	0.839	0.570	0.817	0.799
JPEG trans. errors	0.283	0.265	0.018	0.165	0.070	0.123	0.289
Cumulative subsets	0.552	0.566	0.403	0.584	0.293	0.559	0.619

Table 3. SROCC values between actual subjective scores and predicted ones for the TID2008 database without specific training.

Table 5 gives obtained results when a One-sided t-test is used to provide statistical significance of NR-IQA/DMOS on LIVE database. Each entry in this table is coded using six symbols. The position of each symbol corresponds to one subset of the LIVE database: JP2K, JPEG, White noise, Gaussian blur, Fast fading, all data. Each symbol gives the result of the hypothesis test on the subset. If the symbol equals '1', the NR-IQA on the row is statistically better than the NR-IQA on the column ('0' means worse, '-' is used when NR-IQAs are indistinguishables).

Tables 6 and 7 show similar results for TID2008 and CSIQ databases. For table 6, the position of the symbol represents the tested subsets: JP2K, JPEG, Additive Gaussian noise, Gaussian blur, JPEG transmission error, all data. Considering table 7, the position of the symbol represents the tested subsets: Additive pink Gaussian noise, JP2K, JPEG, Gaussian noise, Gaussian blur, Global Contrast Decrements, all data.

These results confirm that, without any training, performance of NR-IQA algorithms remain highly dependent on the used databases even if BRISQUE seems to be the best performing in each case.

4. INFLUENCE OF THE TRAINING SET

As a second step of this study, we investigate how the selection of the training set influences the results. To perform such investigation, we proceed as follows : the training set is generated using different subsets randomly extracted from all databases. The remaining subsets are used to construct the test set. All trial NR-IQA algorithms are trained on this new training set. Finally, performances of the used NR metrics are estimated on the new test set.

CSIQ subset	BIQI	DIIVINE	BLIINDS	BLIINDS-II	BIQ-An	NIQE	BRISQUE
JP2k	0.708	0.830	0.575	0.895	0.460	0.906	0.866
JPEG	0.867	0.799	0.264	0.901	0.012	0.883	0.903
Gaussian Noise	0.324	0.176	0.293	0.379	0.091	0.299	0.252
Add. Gaussian Pink Noise	0.879	0.866	0.555	0.801	0.303	0.810	0.925
Gaussian Blur	0.771	0.871	0.774	0.891	0.739	0.892	0.903
Global Contrast Decrement	0.585	0.396	0.078	0.012	0.767	0.232	0.029
Cumulative subsets	0.619	0.596	0.170	0.577	0.286	0.628	0.566

Table 4. SROCC values between actual subjective scores and predicted ones for the CSIQ database without specific training.

	BIQI	DIIVINE	BLIINDS	BLIINDS-II	BIQ-An	NIQE	BRISQUE
BIQI	------	00-000	01-0-1	00-000	111111	0--00-	000000
DIIVINE	11-111	------	-1-011	00-00-	111111	-1-0-1	00-000
BLIINDS	10-1-0	-0-100	------	00--00	111111	----11	00-000
BLIINDS 2	11-111	11-11-	11--11	------	111111	11--10	11-0-0
BIQ-An	000000	000000	000000	000000	------	000000	000000
NIQE	1--11-	-0-1-0	----00	00--01	111111	------	00-000
BRISQUE	111111	11-111	11-111	00-1-1	111111	11-111	------

Table 5. Statistical significance matrix of NR-IQA/DMOS on LIVE database. Each entry in the table is a codeword consisting of 6 symbols. The position of the symbol represents the tested subsets: JP2K, JPEG, White noise, Gaussian blur, Fast fading, all data. Each symbol gives the result of the hypothesis test on the subset: '1' means that the algorithm for the row is statistically better that the algorithm for the column, '0' means it is worse, and '-' means it is indistinguishable.

	BIQI	DIIVINE	BLIINDS	BLIINDS-II	BIQ-An	NIQE	BRISQUE
BIQI	------	01----	-11--1	0----0	-11--1	0-----	0----0
DIIVINE	10----	------	-111--	-0---0	1111-1	-0----	-0---0
BLIINDS	-00--0	-000--	------	0000-0	1----1	0000-0	000--0
BLIINDS 2	1----1	-1---1	1111-1	------	1111-1	-----1	--0---
BIQ-An	-00--0	0000-0	0----0	0000-0	------	0000-0	0000-0
NIQE	1-----	-1----	1111-1	-----0	1111-1	------	-----0
BRISQUE	1----1	-1---1	111--1	--1---	1111-1	-----1	------

Table 6. Statistical significance matrix of NR-IQA/DMOS on TID database subsets. Each entry in the table is a codeword consisting of 6 symbols. The position of the symbol represents the tested subsets: JP2K, JPEG, Additive Gaussian noise, Gaussian blur, JPEG transmission error, all data. Each symbol gives the result of the hypothesis test on the subset: '1' means that the algorithm for the row is statistically better that the algorithm for the column, '0' means it is worse, and '-' means it is indistinguishable.

	BIQI	DIIVINE	BLIINDS	BLIINDS-II	BIQ-An	NIQE	BRISQUE
BIQI	--------	--0-01-	11---1	100-01-	111-1-1	100-01-	000-010
DIIVINE	--1-10-	-------	111--01	100--0-	111-101	10--00-	00--000
BLIINDS	00---00	000--10	-------	000---0	1-1-10-	000-0-0	000-0-0
BLIINDS 2	011-10-	011--1-	111---1	-------	111-101	----0--	0---0-0
BIQ-An	000-0-0	000-010	0-0-01-	000-010	-------	000-010	000-010
NIQE	011-10-	01--11-	111-1-1	----1--	111-101	-------	0-----0
BRISQUE	111-101	11--111	111-1-1	1---1-1	111-101	1-----1	-------

Table 7. Statistical significance matrix of NR-IQA/DMOS on CSIQ database subsets. Each entry in the table is a codeword consisting of 7 symbols. The position of the symbol represents the tested subsets: Additive pink Gaussian noise, JP2K, JPEG, Gaussian noise, Gaussian blur, Global Contrast Decrements, all data. Each symbol gives the result of the hypothesis test on the subset: '1' means that the algorithm for the row is statistically better that the algorithm for the column, '0' means it is worse, and '-' means it is indistinguishable.

TID2008 subset	BIQI	DIIVINE	BLIINDS	BLIINDS-II	BIQ-An	NIQE	BRISQUE
Additive noise in color components is more intensive than additive noise in the luminance component	0.46	0.366	0.441	0.778	0.117	0.742	0.495
Spatially correlated noise	0.589	0.722	0.245	0.446	0.311	0.758	0.584
Masked noise	0.897	0.872	0.688	0.866	0.175	0.854	0.623
High frequency noise	0.787	0.864	0.583	0.586	0.016	0.687	0.582
Impulse noise	0.406	0.188	0.039	0.719	0.110	0.812	0.723
Quantization noise	0.747	0.824	0.663	0.839	0.570	0.817	0.799
Image denoising	0.306	0.759	0.509	0.760	0.481	0.605	0.570
JPEG2000 transmission errors	0.367	0.021	0.211	0.606	0.262	0.493	0.260
Non eccentricity pattern noise	0.010	0.032	0.033	0.142	0.067	0.016	0.163
Local blockwise distortions of different intensity	0.022	0.060	0.150	0.457	0.124	0.183	0.175
Mean shift (intensity shift)	0.024	0.050	0.355	0.057	0.384	0.138	0.091
Cumulative subsets	0.282	0.145	0.072	0.342	0.037	0.134	0.225

Table 8. SROCC values between actual subjective scores and predicted ones for the remaining subsets the TID2008 database.

4.1 Performance without training

Before investigating the influence of the training set on the performance of machine learning-based algorithms, all trial NR-IQA algorithms were tested on the remaining subsets of TID 2008 (remaining degradations) that have not been employed in previous section. As for previous section, no new training phase has been performed. Obtained results are given in Table 8. As expected, the performance of all metrics drops significantly. However, except BIQ-An, one can notice that SROCC values remain somewhat interesting for subsets (masked noise, high frequency noise, quantization noise...) similar to the learnt ones (remember that all metrics have been trained on LIVE subsets). On these non learned subsets, BLIINDS-II seems to be the best.

4.2 Training set design

To determine whether the design of the learning base can influence the performance of machine learning-based NR-IQA algorithms, we selected degradations that do not overlap all databases. In addition, this will help us to have an idea on the generalization capability of tested algorithms.

Table 9 presents existing degradations in the three trial databases.

The subsets that are candidate to design the training set are those who are only present in the TID database but not in the two others (LIVE and CSIQ). From Table 9, one observes that eleven subsets are candidate. All associated images to those eleven subsets will define the set from which the training set is designed. A leave-one-out cross validation approach is applied to construct the training base. This means that each subset is included once within the test set (including both LIVE and CSIQ databases) while the remaining sets form the final training set.

Once the training phase is completed, all trial algorithms are tested on both LIVE and CSIQ databases.

4.3 Results

Tables 10 and 11 present obtained results respectively for LIVE and CSIQ databases. Several observations can be made. First, One can observe a drop of SROCC values of all metrics when these values are compared to those in Tables 2 and 4. Knowing that for tables 2 and 4 all metrics have been trained on LIVE subsets, the observed drop in performance is then mainly induced by the unlearnt degradations of the test set. If one observes the

Degradation	**Databases**		
	LIVE	TID	CSIQ
Addtive Gaussian Noise	X	X	X
Additive noise in color components is more intensive than additive noise in the luminance component		X	
Spatially correlated noise		X	
Masked noise		X	
High frequency noise		X	
Impulse noise		X	
Quantization noise		X	
Gaussian blur	X	X	X
Image denoising		X	
JPEG compression	X	X	X
JPEG 2000 compression	X	X	X
JPEG transmission errors	X	X	
JPEG2000 transmission errors		X	
Non eccentricity pattern noise		X	
Local block-wise distortions of different intensity		X	
Mean shift (intensity shift)		X	
Contrast change		X	X

Table 9. Common degradation on the three trial databases.

LIVE subset	BIQI	DIIVINE	BLIINDS	BLIINDS-II	BIQ-An	NIQE	BRISQUE
JP2K	0.502	0.604	0.543	0.977	0.104	0.630	0.756
JPEG	0.655	0.756	0.439	0.877	0.207	0.754	0.732
White Noise	0.651	0.876	0.675	0.861	0.192	0.891	0.851
Gaussian Blur	0.672	0.725	0.493	0.810	0.204	0.698	0.901
Fast fading	0.261	0.460	0.376	0.612	0.015	0.680	0.715
Entire db	0.461	0.682	0.698	0.812	0.265	0.807	0.798

Table 10. SROCC values between actual subjective scores and predicted ones for the LIVE database after training.

CSIQ subset	BIQI	DIIVINE	BLIINDS	BLIINDS-II	BIQ-An	NIQE	BRISQUE
JP2K	0.515	0.670	0.461	0.681	0.230	0.670	0.601
JPEG	0.691	0.703	0.192	0.617	0.014	0.599	0.710
Gaussian Noise	0.561	0.354	0.198	0.501	0.073	0.289	0.561
Add. Gaussian Pink Noise	0.710	0.789	0.456	0.798	0.267	0.698	0.810
Gaussian Blur	0.651	0.771	0.709	0.771	0.561	0.535	0.789
Global Contrast Decrement	0.234	0.126	0.165	0.189	0.235	0.432	0.381
Cumulative subsets	0.598	0.551	0.117	0.542	0.167	0.589	0.475

Table 11. SROCC values between actual subjective scores and predicted ones for the CSIQ database after training.

SROCC values of the entire database, the drop is on average higher for LIVE database. Secondly, for a given degradation, the drop in performance varies from one metric to another. The relative drop can be quite small (7.8% for LIVE database, Gaussian blur degradation and BRISQUE metric) or very important (140% for LIVE database, JPEG degradation and BIQ-An degradation). Finally, for a given metric, the drop in performance varies from one database to another. This variation is induced by the different image contents of the specific databases. Hence, a minimum variation reflects the metric robustness against complexity of images. Based on such analysis, NIQE and BRISQUE followed by BLINDS-II seem to be the most robust metrics.

5. CONCLUSION

In this paper, we perform a comparative study of seven well known no-reference learning-based image quality algorithms. To test the performance of these algorithms, three public databases are used: 1) LIVE, 2) TID and 3) CSIQ databases that are publicly available. The SROCC is utilized to measure and compare the performance.

As a first step, the trial algorithms are compared when no new learning phase is performed. Results are analyzed considering the same degradations for all subsets. One observes that the performance of each metric are dependent on the used database, and that except for BIQ-An, a higher SROCC is always observed for the LIVE database. This behavior was somewhat expected since all metrics have been trained on LIVE subsets.

The second step investigates how the training set influences the results. Before performing a new learning page, we analyze quality scores obtained using all trial NR-IQA algorithms on no learnt distortions without proceeding with a new learning process. As expected, the performance of all metrics drops significantly. However, except BIQ-An, one can notice that SROCC values remain somewhat interesting for subsets similar to the learnt ones. In a second way, a new learning phase is investigated to determine how it impacts the performance of IQA schemes. From the three considered databases, only exclusive degradations with one database are used. We identify eleven candidate subsets, all extracted from TID. A leave-one-out approach is used to generate the learning process. The test set is designed from both LIVE and TID database.

From obtained results, one observes that 1) for a given degradation, the drop in performance varies from one metric to another, and 2) for a given metric, the drop in performance varies from one database to another. This variation is induced by the different image contents which depending on the database type.

REFERENCES

1. R. Muijs and I. Kirenko, "A no-reference blocking artifact measure for adaptive video processing," in *European Signal Processing Conference (Eusipco)*, 2005.
2. Z. ParvezSazzad, Y. Kawayoke, and Y. Horita, "No-reference image quality assessment for jpeg2000 based on spatial features," *Signal Processing: Image Communication* **23(4)**(4), pp. 257—268, 2008.
3. N. Ramin, *Vers une métrique sans reference de la qualité spatiale d'un signal video dans un contexte multimedia*. PhD thesis, Université de Nantes, 2009.
4. R. Barland and A. Saadane, "Blind quality metric using a perceptual map for jpeg2000 compressed images," in *International Conference on Image Processing (ICIP)*, 2006.
5. L. Hantao, N. Klomp, and I. Heynderickx, "A no-reference metric for perceived ringing artifacts in images," *IEEE Transactions on Circuits and Systems for Video Technology* **20**(4), pp. 529–539, 2010.
6. H. Luo, "A training-based no-reference image quality assessment algorithm," in *International Conference on Image Processing (ICIP)*, pp. 2973—2976, 2004.
7. Y. R. Tsoy, V. G. Spitsyn, and A. V. Chernyavsky, "No-reference image quality assessment through interactive neuroevolution," in *International Conference on Computer Graphics and Vision*, pp. 23–27, 2007.
8. R. V. Babu, S. Suresh, and A. Perkis, "No-reference JPEG image quality assessment using GAP-RBF," *Signal Processing* **87 (6)**(6), pp. 1493—1503, 2007.
9. A. K. Moorthy and A. C. Bovik, "Visual importance pooling for image quality assessment," *IEEE J. Selected Topics in Signal Process., Special Issue on Visual Media Quality Assessment* **3**(2), pp. 193—201, 2009.
10. A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indice," *IEEE Signal processing letters,* **17**(5), pp. 513—516, 2010.

11. S. Gabarda and G. Cristobal, "Blind image quality assessment through anisotroy," *JOSA* **24**(12), pp. B42–B51, 2007.
12. T. Brandao and M. P. Queluz, "No-reference image quality assessment based on dct-domain statistics," *Signal Processing* **88**(4), pp. 822–833, 2008.
13. A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From scene statistics to perceptual quality," *IEEE Transactions Image Processing* **20**(12), pp. 3350–3364, 2011.
14. M. Jung, D. Léger, and M. Gazalet, "Univariant assessment of the quality of images," *J. Elect. Imaging* **11**(3), pp. 354–364, 2002.
15. Y. R. Tsoy, V. G. Spitsyn, and A. V. Chernyavsky, "No-reference image quality assessment through interactive neuroevolution," in *International Conference on Computer Graphics and Vision*, pp. 23–27, 2002.
16. M. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the dct domain," *IEEE Transactions on Image Processing* **21**(8), pp. 3339–3352, 2012.
17. M. Saad, A. C. Bovik, and C. Charrier, "A DCT statistics-based blind image quality index," *IEEE Signal Processing Letters* **17**(2), pp. 583–586, 2010.
18. A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing* **21**(12), pp. 4695–4708, 2012.
19. A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Process. Lett.* **20**(3), pp. 209–212, 2013.
20. N. Ponomarenko, O. Eremeev, V. Lukin, and K. Egiazarian, "Statistical evaluation of no-reference image visual quality metrics," in *EUVIP10*, pp. 50–54, 2010.
21. Laboratory for Image & Video Engineering, University of Texas (Austin), "LIVE Image Quality Assessment Database," *http://live.ece.utexas.edu/research/Quality* , 2002.
22. N. Ponomarenko, M. Carli, V. Lukin, K. E. ans J. Astola, and F. Battisti, "Color image database for evaluation of image quality metrics," in *International Workshop on Multimedia Signal Processing*, pp. 403–408, (Australia), Oct. 2008.
23. E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging* **19**(1), p. 011006, 2010.