A Two-Step Framework for Constructing Blind Image Quality Indices

Anush Krishna Moorthy and Alan Conrad Bovik, *Fellow, IEEE*

***Abstract*—Present day no-reference/no-reference image quality assessment (NR IQA) algorithms usually assume that the distortion affecting the image is known. This is a limiting assumption for practical applications, since in a majority of cases the distortions in the image are unknown. We propose a new two-step framework for no-reference image quality assessment based on natural scene statistics (NSS). Once trained, the framework does not require any knowledge of the distorting process and the framework is modular in that it can be extended to any number of distortions. We describe the framework for blind image quality assessment and a version of this framework—the blind image quality index (BIQI) is evaluated on the LIVE image quality assessment database. A software release of BIQI has been made available online: http://live.ece.utexas.edu/research/quality/BIQI_release.zip.**

I. Introduction

OBJECTIVE blind/no-reference (NR) image quality assessment (IQA) refers to algorithms that seek to predict the quality of distorted images without any knowledge of pristine reference images and that correlate well with human perception of quality. Recently, the field of NR IQA has seen a significant rise in activity [1]–[4]; however there is considerable room for improvement. This is largely due to the fact that NR IQA is an extremely difficult problem to solve. In fact, only recently has the field of the 'easier' full-reference (FR) IQA matured to produce algorithms that correlate well with human perception of quality [5].

Present day NR IQA algorithms generally assume that the distortion affecting the image is known. For example, there exist NR IQA algorithms that seek to assess the quality of JPEG/JPEG2000 compressed images [1], [2] or blurred images [3]. Here, we propose a new two-step general-purpose framework for designing no-reference image quality indices based on natural scene statistic (NSS) models of images. The two steps are image distortion classification based on a measure of how the NSS are modified, followed by quality assessment, using an algorithm specific to the decided distortion. Once trained, an algorithm of the proposed framework does not require further foreknowledge of the distortion affecting the images to be assessed. The framework is modular in that it can be extended to any number of distortions.

The works closest in concept to ours are those proposed in [6], [7] for video quality assessment (VQA). In both these cases, a combination of techniques were used to measure blockiness, blur, corner outliers and noise were combined into a quality score by a Minkowski sum. The authors used a set of pre-fixed thresholds as well as parameters obtained using a training set in order to produce these scores, which were then tested on a set of videos with a combination of the considered distortions. Our work differs from that in [6] and [7] in several ways. We do not explicitly seek to characterize the structure of blockiness and other distortions using local filters, but instead utilize concepts from NSS to produce an easily extensible approach to other distortions. We utilize a unique 2-stage strategy which identifies the likeliest distortion in the image and then quantifies this distortion using a NSS-based approach. Finally, we test our algorithm on the publicly available LIVE IQA dataset [5].

Thus, we present a new and unique framework for no-reference image quality index design. The aim is to produce algorithms that are truly 'no-reference'—no information about the distortion affecting the image is contained within the distorted image[1] or otherwise made available to the algorithm.[2] In order to create such a no-reference algorithm, we utilize techniques from NSS [8].

It has been demonstrated that subband responses of natural scenes tend to follow a non-Gaussian (heavy-tailed) distribution, which can be parametrized [8]. Given that this is true, a pertinent question is whether there exists such a general statistical description for distortions. Even though researchers have observed that distortions affect scene statistics [2], our aim is to assess whether such changes in NSS are *systematic* and *parameterizable*. Here, we show that not only do distortions affect NSS, they are also systematic and parameterizable.

A description of distorted image statistics (DIS) when obtained, can be used as a *distortion-specific signature* for classifying an image into a particular distortion category. Once such classification is achieved, it is as if the algorithm is aware of the distortion. The algorithm can then deploy a distortion-specific IQA algorithm. Although this may sound simple, there exist many subtleties that we explore in this paper. For example, our description here assumes that the distortion classes are disjoint—an assumption that is not always true. Here, we propose a novel scheme to handle such subtleties. Further, although there exist many distortion-specific IQA algorithms which we can use after the initial classification, we develop a general framework for IQA based on DIS. Indeed, this approach to QA does not measure distortion-specific indicators such as blocking, but provides a modular strategy that adapts itself to the distortion in question.

Manuscript received January 08, 2010; revised February 08, 2010. First published March 18, 2010; current version published April 02, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Patrizio Campisi

The authors are with the Laboratory for Image and Video Engineering, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: anushmoorthy@mail.utexas.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2010.2043888

[1]Some authors define such IQA as NR as well. However, using supplemental information regarding the source—by watermarking for example—we regard as reduced-reference QA.

[2]Note that the proposed algorithm is limited by the set of distortions it is trained on.

The first contribution of this work is the development of a novel two-stage modular framework for distorted image statistics based no-reference quality assessment. The modularity of the proposed framework implies that this approach is truly extensible in that any addition of distortion categories beyond those discussed here may be easily accomplished [7]. The second contribution is an innovative strategy to assess quality when the distortion is known. A combination of these two contributions results in a no-reference image quality index. In this paper, we describe a a specific algorithm of our proposed framework—the blind image quality index (BIQI)—and test its performance by the construction of an example 5-distortion algorithm on the LIVE image database [2]. We demonstrate that our no-reference algorithm performs well in terms of correlation with human perception, and that it is competitive with classical *full-reference* IQA algorithms. We also provide a software release of the BIQI version described here at [9].

II. A Framework for No-Reference Image Quality Indices

The framework for creating a no-reference image quality index proceeds as follows. Given a distorted image, the algorithm first estimates the presence of a set of distortions in the image. In our demonstration, this set consists of JPEG, JPEG2000 (JP2K), white noise (WN), Gaussian Blur (Blur) and Fast fading (FF). These distortions are those from the LIVE IQA database [5]. The amount or probability of each distortion in the image is gauged and denoted as p_i, $\{i = 1, \ldots, 5\}$. This first stage is essentially a classification stage. The second stage evaluates the quality of the image along each of these distortions. Let q_i, $\{i = 1, \ldots, 5\}$ represent the quality scores from each of the five quality assessment algorithms (corresponding to the five distortions). The quality of the image is then expressed as as a probability-weighted summation: $BIQI = \sum_{i=1}^{5} p_i \cdot q_i$. Note that by defining the index in this fashion, we are inherently introducing modularity in the system. Even though we use certain algorithms for IQA and a specific classifier, an implementation of our index is not limited by the actual IQA algorithms or the classifier.

Having described the framework, we now describe how distortions affect image statistics and how these distorted statistics can be used to classify images into distortion categories considered here. Using these statistics, we also develop a quality assessment algorithm that can be used to evaluate the quality of a distorted image given knowledge of the distortion. These two stages, when combined as described above, form an implementation of the above framework.

III. A Demonstration of the No-Reference Image Quality Index

A. Distorted Image Statistics

An image is subjected to a wavelet transform over three scales and three orientations using the Daubechies 9/7 wavelet basis [10]. The subband coefficients so obtained are parametrized using a generalized Gaussian distribution (GGD). The GGD is $f_X(x; \mu, \sigma^2, \gamma) = ae^{-[b|x-\mu|]^\gamma}$ $x \in \Re$; where, μ, σ^2 and γ are the mean, variance and shape-parameter of the distribution and $a = \beta\gamma/2\Gamma(1/\gamma)$, $b = (1/\sigma)\sqrt{\Gamma(3/\gamma)/\Gamma(1/\gamma)}$; $\Gamma(\cdot)$ is the gamma function $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$ $x > 0$.

The shape parameter γ controls the 'shape' of the distribution. For example, $\gamma = 2$ yields a Gaussian distribution and $\gamma = 1$ yields a Laplacian distribution. The parameters of the distribution (μ, σ^2 and γ) are estimated using the method proposed in [11]. Since wavelet bases act as band-pass filters, the responses are zero-mean and we are left with 2 parameters (σ^2 and γ) for each subband. An 18-D vector $\vec{f_i}$ $(3 \text{ scales} \times 3 \text{ orientations} \times 2 \text{ parameters})$ is thus formed from these estimated parameters and is the representative feature vector for that image. This feature vector characterizes the distortion that the image is subjected to. Given a training and test set of distorted images, we use a classifier with $\vec{f_i}$ as the feature vector to classify the images into five different distortion categories, based on the distortion type—JPEG, JPEG2000, WN, Blur, and FF. The classifier used is a support vector machine (SVM) [12]. SVMs are popular as classifiers since they perform well in high-dimensional spaces, avoid over-fitting and have good generalization capabilities [12]. In our demonstration, a multiclass SVM with a radial-basis function (RBF) kernel is used to classify a given image into one of five distortion categories.

We note that by classifying the images into these categories, we do not mean to insinuate that the classes are disjoint. In fact, some of these classes overlap to some extent. For example, JPEG images exhibit some amount of blur apart from blocking. Hence, even though we report classification results (see below) our goal is not an absolute classification, but an indication of the amount of each distortion present in the image. This amount of distortion is extracted from the probability estimates provided by the SVM. A greater value indicates a higher proportion of that distortion in the image. There are two parameters to be set in the SVM—(C, γ) [12]; these are set using 5-fold cross validation on the training set of images.

Given that an image belongs to a particular distortion category, we reuse the computed feature vector—$\vec{f_i}$—to compute the quality of an image. In order to produce a quality index, we utilize support vector regression (SVR) [12]. The ν-SVM is utilized to perform such a regression [13]. Specifically, for each distortion that we consider, a ν-SVM is trained using quality scores from the training set to learn the mapping from the feature space to subjective quality. When presented with a distorted image from the category that the SVM is trained for, this regression provides a score representative of the quality of that image. Note that this technique is generic, and can be used across all distortion types, if the distortions are known. Coupled with the above stage of distortion categorization, given any distorted image, the algorithm will produce a quality score; which, as we shall show soon, correlates well with human perception. Finally, each ν-SVM constructed for QA requires a set of parameters (C, γ, ν) to be determined. As before, for C, γ, a 5-fold cross-validation on the training set is used to select these parameters. ν is the default value (0.5). Changing the value of ν does not seem to affect results much.

In our trials we found that this approach for IQA based on DIS works quite well for images corrupted by white noise and blur and to some extent for JPEG2000 and FF. However, the performance for JPEG compression is less impressive. Since the framework for no-reference image quality assessment is independent of the specific algorithm that is used in each of the quality assessment modules, we can replace the JPEG module with an existing (off-the-shelf) algorithm that performs better than our approach [1].

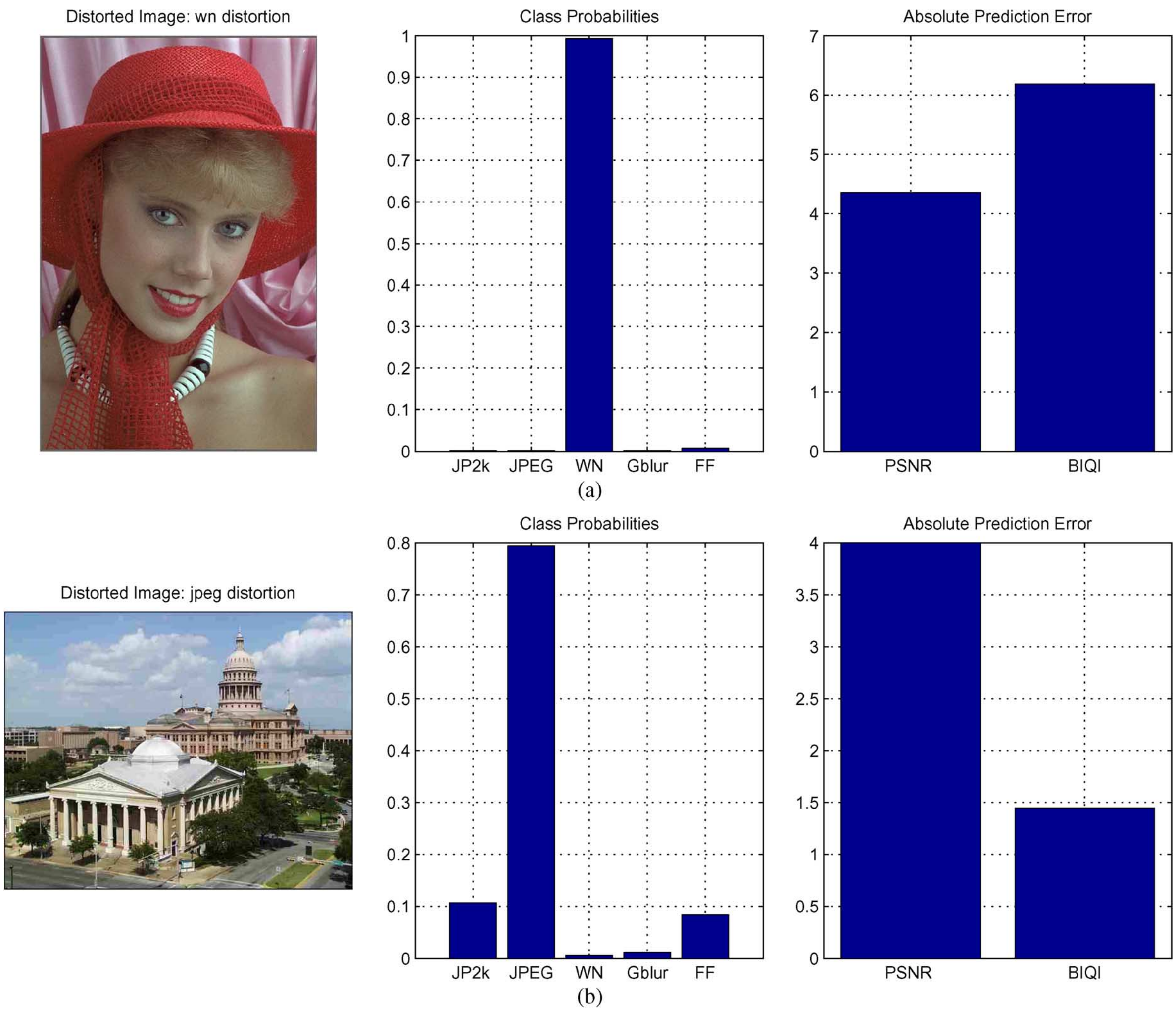

Fig. 1. BIQI operation. (a) and (b) Classification probabilities for the images (correct class as image title) on the left and the absolute prediction error between BIQI and DMOS. The error between PSNR and DMOS is included for comparison.

Hence, our demonstration of the proposed framework, labeled BIQI, consists of the classifier as described above, followed by JPEG2000, noise, blur and FF modules based on SVR, and a JPEG module based on the algorithm in [1]. Fig. 1 shows an example of how BIQI operates for two different distortions. A software version is available in [9].

IV. Results

In order to demonstrate performance of the proposed algorithm, we use the LIVE image quality assessment database [5]. The LIVE IQA database consists of 29 reference images and five distortion types—JPEG, JPEG2000 (JP2K), Blur, Noise and FastFading (FF). A total of 808 distorted images along with the associated differential mean opinion score (DMOS) is available. DMOS is representative of the perceived quality of the image and was computed from subjective scores of over 29 subjects in a large scale study, details of which may be found in [5].

We randomly select 15 reference images—and their associated distorted versions—for testing, and 14 (different) reference images—and their associated distorted versions—for training. This way, we ensure that the training and test set are disjoint. The sets do not share content, and because of the design of the dataset, they do not share specific distortion severities either. In this way, our algorithm is independent of content and specific distortion severity. Therefore, our demonstration of the algorithm only learns the distortion space as a whole, instead of specific distortion levels. The training set is used to train the classifier as well as the quality assessment algorithms as described in the previous section. The algorithms are then tested on the test set. Such a train-test combination is conducted 1000 times with random permutations of the LIVE database, in order to ensure that the algorithm is robust across all contents and distortion severities.

Over 1000 such train-test sets, the median classification accuracy was 81.5161% $(\text{mean} = 81.5161\%, \text{std. dev.} = 3.1708\%)$. As we have mentioned, the absolute classification accuracy is not of import; we are interested in the probability estimates, and finally in the quality estimates.

Table I shows the median value of Spearman's rank ordered correlation coefficient (SROCC) over 1000 trials for our version of BIQI. It is obvious that BIQI performs well in terms of correlation with human perception; and is competitive with the *full-reference* PSNR across distortion types and across the database. This is a remarkable result, as it shows that algorithms

TABLE I
MEDIAN SPEARMAN'S RANK ORDERED CORRELATION COEFFICIENT (SROCC) BETWEEN ALGORITHM AND DMOS

	JP2k	JPEG	WN	Blur	FF	All
PSNR	0.8558	0.8762	0.9388	0.7292	0.8592	0.8535
BIQI	0.7995	0.8914	0.9510	0.8463	0.7067	0.8195

TABLE II
MEDIAN LINEAR CORRELATION COEFFICIENT (LCC) BETWEEN ALGORITHM AND DMOS

	JP2k	JPEG	WN	Blur	FF	All
PSNR	0.8616	0.8849	0.9249	0.7476	0.8589	0.8481
BIQI	0.8086	0.9011	0.9538	0.8293	0.7328	0.8205

TABLE III
MEDIAN ROOT-MEAN-SQUARED-ERROR (RMSE) BETWEEN ALGORITHM AND DMOS

	JP2k	JPEG	WN	Blur	FF	All
PSNR	12.7725	14.8307	10.6394	12.2175	14.5182	14.4555
BIQI	14.8427	13.7552	8.4094	10.2347	19.2911	15.6223

that operate without any reference information can offer performance competitive with the predominant IQA algorithm over the last five decades.

We utilize a probability weighted summation to compute the final BIQI score, hence, the performance of each of the QA modules are not independent of each other. In fact, the use of an algorithm that correlates well with human perception in one category will probably help improve the overall performance of BIQI. This is a unique advantage, since overall better performance may be obtained as we find or create NR algorithms that correlate better with human perception.

We also measure the linear correlation coefficient (LCC) and root mean squared error (RMSE) between BIQI and DMOS. Before computing LCC or RMSE, BIQI is transformed using a logistic function fit as prescribed in [5]. Table II lists the (median of the) LCC values of BIQI and PSNR and Table III lists the (median of the) RMSE values (smaller is better) for each distortion type and across all distortions. Again, the performance of BIQI is competitive with that of full-reference PSNR.

The reader will notice that even though BIQI performs competitively with (and in some cases beats) PSNR across all of these measures, there is some room for improvement in the JP2K and FF case. This is essentially because our JP2K measure is not much better than PSNR. Just as we replaced the JPEG module, one can replace the JP2K module with a better performing one; for example the one in [2] to improve performance. Even though we report results for the FF case, realize that FF consists of JP2K compression followed by packet loss. We believe that the current approach may not be ideal for such multiple distortions. Future work will involve devising a theory to predict the quality of such multiply distorted images that fits into our current framework. Future work will involve improving the SVR measure of quality for JPEG and JPEG2000 compression.

V. CONCLUSION

In this letter, we described a framework for constructing an objective no-reference/no-reference (NR) image quality assessment (IQA) measure. The framework is unique, since it assesses the quality of an image completely blind—i.e., without any knowledge of the source distortion. This is achieved by using distorted image statistics (DIS)—an extension of natural scene statistics for distorted images. In this paper, we discussed DIS, and demonstrated that each distortion has a unique signature which can be characterized by the use of DIS and used this signature to classify images into distortion categories. We also described how distortion-aware IQA may be undertaken using DIS. We then combined distortion-classification with distortion-aware IQA to produce a demonstration of the blind image quality index (BIQI) which is of value on its own. BIQI was tested on the LIVE image database and was shown to perform well in terms of correlation with human perception. Indeed, BIQI, an NR measure, performed competitively with (and in many cases beat) PSNR, an FR measure, across all distortions and in overall performance. A software release of BIQI has been made available at [9] to further research in the are of NR IQA.

The architecture of proposed framework is modular, and although we have used only a few techniques for classification and QA, one can replace any module with a better-performing one. Such a replacement can be expected to lead to better overall performance. Future research will involve improving current BIQI performance, extending BIQI to a larger set of distortions including multiply distorted images, extending the approach to include explicitly computed perceptually relevant features and developing BIQI for video.

REFERENCES

[1] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *Proc. IEEE ICIP*, 2002, vol. 1, pp. 477–480.
[2] H. R. Sheikh, A. C. Bovik, and L. K. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1918–1927, 2005.
[3] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: Application to JPEG2000," *Signal Process.: Image Commun.*, vol. 19, no. 2, pp. 163–172, 2004.
[4] J. Caviedes and F. Oberti, "A new sharpness metric based on local kurtosis, edge and energy information," *Signal Process.: Image Commun.*, vol. 19, no. 2, pp. 147–161, 2004.
[5] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.
[6] M. Farias and S. Mitra, "No-reference video quality metric based on artifact measurements," in *Proc. IEEE ICIP*, 2005, vol. 3, pp. 141–144.
[7] J. Caviedes and J. Jung, "No-reference metric for a video quality control loop," in *5th World Multiconf. on Systemics, Cybernetics and Informatics*, 2001.
[8] A. Srivastava, A. Lee, E. Simoncelli, and S. Zhu, "On advances in statistical modeling of natural images," *J. Math. Imag. Vis.*, vol. 18, no. 1, pp. 17–33, 2003.
[9] A. K. Moorthy and A. C. Bovik, Buqi-4d Software Release 2009 [Online]. Available: http://live.ece.utexas.edu/research/quality/biqi_release.zip
[10] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.
[11] K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized Gaussiandistributions in subband decompositions of video," *IEEE Tran. Circuits Syst. Video Technol.*, vol. 5, no. 1, pp. 52–56, Feb. 1995.
[12] V. Vapnik, *The Nature of Statistical Learning Theory*. Berlin, Germany: Springer Verlag, 2000.
[13] B. Schölkopf, A. Smola, R. Williamson, and P. Bartlett, "New support vector algorithms," *Neural Comput.*, vol. 12, no. 5, pp. 1207–1245, 2000.