A TWO-STAGE FRAMEWORK FOR BLIND IMAGE QUALITY ASSESSMENT

Anush K. Moorthy and Alan C. Bovik

Dept. Of Electrical and Computer Engineering,
The University of Texas at Austin,
Austin, Texas - 78712.

ABSTRACT

Most present day no-reference/blind image quality assessment (NR IQA) algorithms are distortion specific - i.e., they assume that the distortion affecting the image is known. Here we propose a novel two stage framework for distortion-independent blind image quality assessment based on natural scene statistics (NSS). The proposed framework is modular in that it can be extended beyond the distortion-pool considered here, and each module proposed can be replaced by better-performing ones in the future. We describe a 4-distortion demonstration of the proposed framework and show that it performs competitively with the full-reference peak-signal-to-noise-ratio on the LIVE IQA database. A software release of the proposed index has been made available online: http://live.ece.utexas.edu/research/quality/BIQI_4D_release.zip.

***Index Terms*—** No reference image quality assessment, blind quality assessment, natural scene statistics

1. INTRODUCTION

Objective full-reference image quality assessment (FR IQA) refers to quality assessment of images by an algorithm where, apart from the distorted image, the pristine reference image is made available to the algorithm [1]. The field of FR IQA has seen tremendous activity in the recent past, and some good FR IQA algorithms have been proposed [2]. The 'good-ness' of any algorithm is gauged by measuring the correlation of algorithmic scores with subjective (differential) mean opinion scores (DMOS/MOS) on a large dataset spanning different distortion. Many such datasets have been proposed and a host of FR IQA algorithms have been evaluated for their performance on these datasets [2, 3].

Objective no-reference/blind IQA (NR IQA) refers to quality assessment of images by an algorithm where only the distorted image - i.e., the image whose quality is to be assessed is made available to the algorithm. No information about the reference image is in any way made available during the testing phase of the algorithm - by watermarking or by embedding some information in the transmitted bit-stream for example (this is referred to as reduced reference IQA). Even though the field of NR IQA has not matured as much as that of FR IQA, the recent past has seen some activity in this area [3, 4, 5, 6, 7, 8, 9].

Most present-day NR IQA algorithms assume that the distortion affecting the image is known. For example, there exist algorithms that predict the quality of images compressed using JPEG compression [4] or those that gauge quality of blurred images [9]. NR IQA algorithms that predict quality without knowledge of the distortion affecting the image are scarce [6, 7].

Here, we present a framework that can be employed for NR IQA - the blind image quality index (BIQI). This framework for BIQI; which is based on natural scene statistics (NSS) [10] is modular in the sense that distortions beyond those considered here may be incorporated into the algorithm at a later date. The framework is based on a novel two-stage process, where the type of distortion affecting the image is explicitly assessed in stage 1 and quality assessment is undertaken in stage 2. Distortion identification, based on distorted image statistics (DIS) [11] is undertaken in order to gauge the primary distortion affecting the image (from a pool of possible distortions). Distortion-specific quality assessment (DSQA) - stage two - is then carried out to predict the quality of the image.

The framework proposed here is closest in concept to the approaches in [6, 7] for video quality assessment (VQA); where a combination of techniques are used to measure distortion indicators such as blocking, blur, noise etc., which are then combined using a Minkowski sum. Here, 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. The technique utilized for DIS and DSQA is unique and is competitive with the popular full reference measure - peak signal to noise ratio (PSNR).

In this paper, we describe the framework for BIQIs and in order to demonstrate the effectiveness of the technique, describe a 4-distortion version of the algorithm - BIQI-4D, which is thoroughly evaluated for its performance on the LIVE IQA database [2], where its performance is compared with that of PSNR. Future work will involve an extension of the proposed framework to include other distortions (apart from those considered here), for multiply-distorted images (JPEG2000 compression, followed by packet loss for example), as well as for video quality assessment (VQA).

2. A FRAMEWORK FOR BLIND IMAGE QUALITY INDICES

The framework for blind image quality indices (BIQIs) proceeds as follows. First the distorted image is transformed into the wavelet domain using Daubechies 9/7 wavelet basis [12] over three orientations and three scales. The wavelet transform is used since it coarsely mimics the scale-space-orientation decomposition hypothesized to occur in the human visual system (HVS) in area V1 of the primary visual cortex [13].

Research in the field of natural scene statistics (NSS) has discovered that coefficients of each subband from such a transform are distributed in a Laplacian fashion (heavy-tails, high concentration around origin) [10]. In [11], we demonstrated that each distortion affects these subband statistics in a characteristic way, and this characteristic can be parametrized. Such a parametrization of subband statistics is obtained by fitting these statistics using a generalized gaussian distribution (GGD). The GGD is:

$f_X(x;\mu,\sigma^2,\gamma) = ae^{-[b|x-\mu|]^\gamma} \quad x \in \Re$; where, μ, σ^2 and γ are the mean, variance and shape-parameter of the distribution and $a = \frac{\beta\gamma}{2\Gamma(1/\gamma)}, b = \frac{1}{\sigma}\sqrt{\frac{\Gamma(3/\gamma)}{\Gamma(1/\gamma)}}$; $\Gamma(\cdot)$ is the gamma function: $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt \quad x > 0$.

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. As in [11], the parameters of the GGD (μ, σ^2 and γ) are estimated using the method proposed in [14]. Since wavelet basis are band-pass in nature, the responses are zero-mean and hence for each subband we estimate a 2 parameters (σ^2 and γ), which are stacked across subbands to form an 18 dimensional feature vector (3 scales $\times$ 3 orientations $\times$ 2 parameters). This feature vector is sufficient to identify the primary distortion affecting the image with high accuracy [11]. In our implementation, such identification is performed using a support vector machine (SVM) [15]; however the framework is not limited by the form of the classifier used.

Our goal here is not distortion identification, but NR IQA. Further, realize that distortions affecting the image may not be independent of each other - for example, in a JPEG compressed image, apart from blocking, blur may exist as well. Hence, instead of performing an absolute classification, we perform a probabilistic one, where the classifier returns probability estimates of the kind of distortion present in the image being analyzed. In the demonstration to follow, this set of distortions includes JPEG and JPEG2000 (JP2K) compression, Gaussian blur (GB) and White Noise (WN).

Once a probability distribution over the distortion set is obtained - let us call this p_i, {i = 1, 2, 3, 4} - the same 18-dimensional feature vector obtained from the subbands is utilized in conjunction with support vector regression (SVR) [15] in order to map the statistics onto a quality score. Such a mapping is distortion specific - i.e., there exist 4 different SVRs (trained on the 4 distortions that we consider here) which map the feature onto a quality score. The statistics of each distorted image (immaterial of the primary distortion that we identify) is subjected such a mapping onto a quality score. Let us denote these quality scores as q_i, {i = 1, 2, 3, 4}. The final quality of the image is then computed as a probability-weighted sum: $BIQI = \sum_{i=1}^{4} p_i \cdot q_i$.

Such a definition allows for an easy extension to a greater number of distortions, beyond those considered here. Further, for cases such as JPEG compression, where the distorted image exhibits primarily blocking along with *some* blur, such a *linear* summation suffices. For multiply distorted images, however, a simple linear summation may not suffice. Indeed, as we shall see later, for the multiple distortion of JPEG2000 compression followed by packet loss, such a linear model performs poorly. Future work will involve an in-depth analysis of how such distortions interact with each other and their overall effect on perceived quality. Also note that since the final score is defined as a linear sum, the performances of the QA modules are not independent of each other. As we shall soon see, the use of a better QA algorithm will lead to an overall improvement in performance of BIQI.

Having described the overall framework of BIQI, we now demonstrate a 4-distortion version of BIQI, which we label: BIQI-4D.

3. BLIND IMAGE QUALITY INDEX: A DEMONSTRATION

In this demonstrations we consider a pool of four distortions - JPEG compression, JPEG2000 (JP2K) compression, white noise (WN) and Gaussian blur (GB). These distortions were considered here for multiple reasons including their current practical need. Notice that each distortion is primary and is essentially independent of the other distortions. Even though such independence is not of significance for the distortion classification stage; as we mentioned above, the linearity assumption will hold only when these classes do not overlap to significant degree. Further, secondary distortions (i.e., multiply distorted image) will not necessarily adhere to the linear summation. Thus, even though the LIVE IQA dataset that we use as a test-bed consists of another category of distortion - fast fading (FF), which is JPEG2000 compression followed by a lossy channel - we do not consider FF in the primary pool of distortions. However, in order to demonstrate that the linear model is ineffective for secondary distortions such as FF, we report the results of a 5-distortion version of BIQI (BIQI-5D), even though our primary contribution here is BIQI-4D.

As we mentioned, the first stage of probabilistic distortion classification is achieved using a support vector machine (SVM) [15]. SVMs are popular as classifiers since they perform well in high-dimensional spaces, avoid over-fitting and have good generalization capabilities [15]. In our demonstration, a multi-class SVM is used to classify a given image into one of 4 distortion categories. As we have shown before, such classification can be achieved with high accuracy [11].

The distortion-specific quality assessment modules are then implemented using support vector regression (SVR). The ν-SVM is utilized to perform such a regression [16]. Specifically, for each distortion that we consider, a ν-SVM is trained using quality scores from the training set (see below) to learn the mapping from the feature space to subjective quality. When presented with a distorted image, this regression provides a score representative of the quality of that image, assuming the presence of the trained distortion. Note that this technique is generic, and can be used across all distortion types, if the distortions are known. A combination of these two stages of classification and QA leads to a set of probability estimates and quality scores which are combined linearly as described before.

One may ponder over the suitability of using an SVR to map the feature vector onto a quality score. Indeed, even though the computed feature has been shown to perform well for classification (since it seems to capture the distortion-specific signatures) [11], it is unclear if this vector is sufficient for QA. As we shall see, this is a pertinent question. We have found that for global distortions such as GB and WN, the computed feature is sufficient for QA; however, for distortions that manifest locally (eg., JPEG, JPEG2K), these features, in their current form, are insufficient for QA (even though they seem to classify images in these categories with relatively high accuracy). Owing to the modularity of the proposed framework, we can easily replace the poorly performing QA modules with ones that perform better and here, as a demonstration, we replace the JPEG SVR-based QA module with a better-performing one [4]. As we shall see such a replacement will improve overall BIQI performance. Although the feature vector does not perform very well for JPEG/JPEG2000 compression in its current form, it does not imply that this feature is useless for QA for these distortions. Future work will involve understanding how to utilize these features, using HVS-based properties such as spatial masking [13], and/or properties of localization [17].

4. RESULTS

The proposed algorithm is evaluated on the LIVE IQA dataset [2]. This dataset consists of 29 reference and 779 distorted images along with the differential mean opinion score (DMOS) for each distorted image which was obtained from a large scale human study [2].

Table 1. Median Spearman's rank ordered correlation coefficient (SROCC) between algorithm and DMOS.

	JP2k	JPEG	WN	Blur	All
PSNR	0.8581	0.8782	0.9392	0.7348	0.8567
BIQI-PURE	0.7699	0.6423	0.9513	0.8005	0.7292
BIQI-4D	0.8077	0.9120	0.9543	0.8375	0.8665

Table 2. Median linear correlation coefficient (LCC) between algorithm and DMOS.

	JP2k	JPEG	WN	Blur	All
PSNR	0.8640	0.8867	0.9255	0.7515	0.8503
BIQI-PURE	0.7775	0.6539	0.9501	0.7675	0.7222
BIQI-4D	0.8128	0.9226	0.9649	0.8232	0.8722

DMOS is representative of the perceived quality of the image, and a higher correlation of algorithmic scores with DMOS indicates better performance. The measures of performance used here are the Spearman's rank ordered correlation coefficient (SROCC), the linear (Pearson's) correlation coefficient (LCC) and root mean squared error (RMSE) between algorithm scores and DMOS. LCC and RMSE are computed after passing the algorithmic scores through a logistic function as described in [2].

Since we use an SVM and an SVR for distortion classification and QA, these machines need to be trained in order to learn the classes and the feature-to-quality mapping. For this purpose, we randomly divide the LIVE dataset into 15 images (and their associated distorted versions) for training and 14 images (and their associated distorted versions) for testing . We train the SVM and SVR on the training set and then compute performance of the test set. The LIVE dataset is then randomly permuted into another such 15 train-14 test split and a performance evaluation is again undertaken. This process is repeated over 1000 such train-test combinations, and the median of the obtained correlations are reported here. By performing an analysis in this fashion, 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 as well. 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.

Even though the classification accuracy is not of import here, we report the values for completeness. Over such 1000 train-test combinations, median classification accuracy was 81.5161% (mean = 81.5161%, std. dev.= 3.1708%).

The results for QA are seen in Tables 1 - 3. In these tables, BIQI-PURE refers to the version of BIQI where the QA modules are all based on the SVR mapping from feature space to DMOS. It is clear that such a QA module performs poorly, especially for JPEG compression. BIQI-4D refers to that version of BIQI where the JPEG QA module from BIQI-PURE is replaced by the one proposed in [4]. As can be seen, the overall performance of BIQI improves when a better performing QA algorithm is utilized. The tables also list the (median) performance of PSNR - a *full-reference* algorithm. Notice how BIQI-4D is competitive with the *full-reference* PSNR. We believe that the final goal for any NR IQA algorithm is to perform competitively with present day FR IQA algorithms, and hence we report PSNR performance here - even though such a comparison is not entirely fair.

The reader will notice that even though BIQI-4D performs well,

Table 3. Median root mean square error (RMSE) between algorithm and DMOS.

	JP2k	JPEG	WN	Blur	All
PSNR	12.67	14.71	10.54	12.14	14.21
BIQI-PURE	15.82	24.08	8.74	11.74	18.73
BIQI-4D	14.68	12.29	7.34	10.45	13.24

there is still some room for improvement in the JPEG2000 case. There are many solutions to this, including a replacement of the JPEG2K QA module with a better performing one (for example the one in [18]). As we have mentioned before, future work will involve an exploration of how the present SVR based QA performance (BIQI-PURE) may be improved.

In order to get a feel for how BIQI-4D works, in Fig 1, we show a distorted image from each distortion considered here, the probabilistic output of the SVM-based classifier and the absolute difference between the predicted DMOS score and the actual DMOS. We also plot this absolute difference between PSNR score and DMOS for a comparison. The correct class of distortion forms the title of the image.

Finally, as we mentioned in the previous section, FF distortion from the LIVE dataset is a form of secondary/multiply distorted image and does not fit into our linear combination framework. Further, it is unclear if the computed feature vector is sufficient for FF QA in its current form. In order to demonstrate that the proposed framework needs to be specifically extended for multiply distorted images, we report results from a 5-distortion version of BIQI which includes JPEG, JP2k, GB, WN and FF. The results are reported as the median SROCC across 1000 trials of train-test pairs as above. Note that the JPEG QA module in BIQI-5D is the one from [4] as in BIQI-4D. Overall SROCC of PSNR across these 1000 runs is 0.8535 while that for BIQI-5D is 0.8195. For the FF distortion, PSNR has SROCC of 0.8592 while that for BIQI-5D is 0.7067.

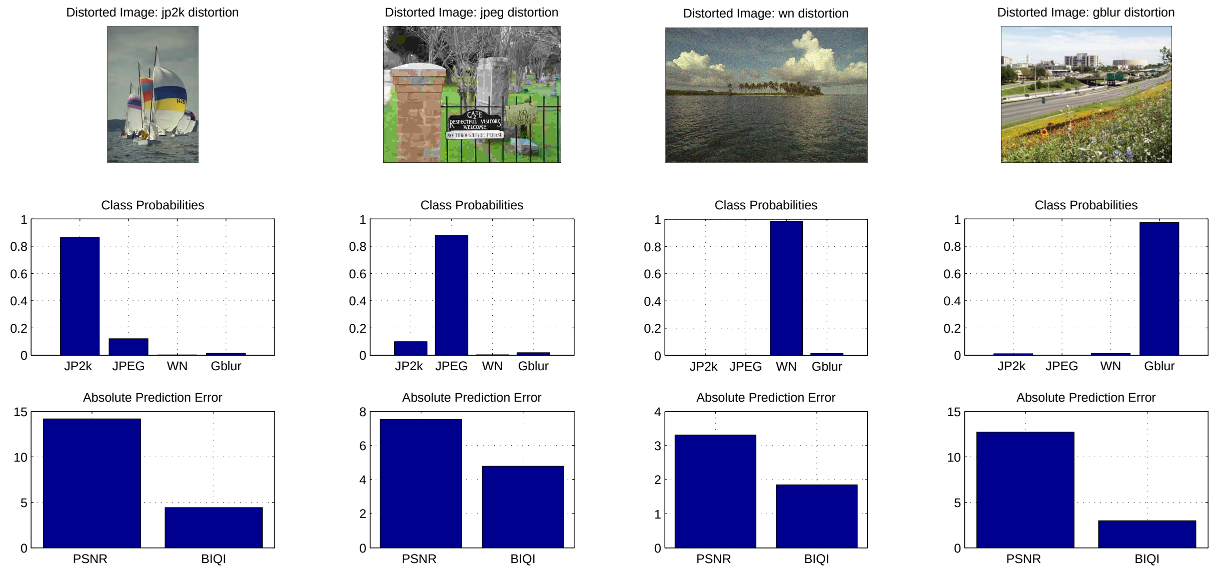
5. CONCLUSION

We described a novel framework for blind/no reference quality indices (BIQIs) based on natural scene statistics (NSS). The proposed framework consists of two stages - probabilistic distorted classification followed by distortion-specific quality assessment. The combination of the two stages leads to a quality index. Here, we demonstrated a 4-distortion version of BIQI called BIQI-4D and tested its performance on the LIVE IQA database and compared its performance to the *full-reference* peak signal to noise ration (PSNR). BIQI-4D was shown to perform competitively with PSNR. Future work will involve extending this framework to more singular distortions, tackling multiply distorted images and video quality assessment as well. A software release of the proposed index has been made available online: http://live.ece.utexas.edu/research/quality/BIQI_4D_release.zip

6. REFERENCES

[1] Z. Wang and A. C. Bovik, *Modern image quality assessment*, vol. 2 of *Synthesis Lectures on Image, Video, and Multimedia Processing*, Morgan & Claypool Publishers, 2006.

[2] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algo-



(a) (b) (c) (d)

Fig. 1. Examples of BIQI-4D. (a), (b), (c) & (d) show the classification probabilities for the images (correct class as image title) on the top and the absolute prediction error between BIQI and DMOS. The error between PSNR and DMOS is included for comparison.

rithms," *IEEE Tran. Image Proc.*, vol. 15, no. 11, pp. 3440–3451, 2006.

[3] Z.M.P. Sazzad, Y. Kawayoke, and Y. Horita, "Spatial features based no reference image quality assessment for JPEG2000," *IEEE Int'l Conf. Image Proc.*, vol. 3, pp. 517–520, 2007.

[4] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," *IEEE Int'l Conf. Image Proc.*, vol. 1, pp. 477–480, 2002.

[5] H. R. Sheikh, A. C. Bovik, and L. K. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Tran. on Image Proc.*, vol. 14, no. 11, pp. 1918–1927, 2005.

[6] M.C.Q. Farias and S.K. Mitra, "No-reference video quality metric based on artifact measurements," *IEEE Int. Conf. on Image Proc.*, vol. 3, pp. 141–144, 2005.

[7] J. Caviedes and Jung, "No-reference metric for a video quality control loop," *5th World Multiconference on Systemics, Cybernetics and Informatics*, 2001.

[8] J. Caviedes and F. Oberti, "A new sharpness metric based on local kurtosis, edge and energy information," *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 147–161, 2004.

[9] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: Application to JPEG2000," *Sig. Proc.: Image Comm.*, vol. 19, no. 2, pp. 163–172, 2004.

[10] A. Srivastava, AB Lee, EP Simoncelli, and S.C. Zhu, "On advances in statistical modeling of natural images," *J. Math. Imaging Vis.*, vol. 18, no. 1, pp. 17–33, 2003.

[11] A. K. Moorthy and A. C. Bovik, "Statistics of natural image distortions," *IEEE Int'l Conf. Acoustics, Speech and Signal Proc.*, 2010 (to appear).

[12] I. Daubechies, *Ten Lectures on Wavelets*, Society for Industrial Mathematics, 1992.

[13] B.A. Wandell, *Foundations of vision*, Sinauer Associates, 1995.

[14] K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized Gaussiandistributions in subband decompositions of video," *IEEE Tran. Circ. Syst. for Video Tech.*, vol. 5, no. 1, pp. 52–56, 1995.

[15] V.N. Vapnik, *The Nature of Statistical Learning Theory*, Springer Verlag, 2000.

[16] B. Schölkopf, A.J. Smola, R.C. Williamson, and P.L. Bartlett, "New support vector algorithms," *Neural Computation*, vol. 12, no. 5, pp. 1207–1245, 2000.

[17] A. K. Moorthy and A. C. Bovik, "Visual importance pooling for image quality assessment," *IEEE Jnl of Sel. Top. in Sig. Proc., Issue on Visual Media Qual. Ass.*, vol. 3, no. 2, pp. 193–201, April 2009.

[18] H. R. Sheikh, A. C. Bovik, and G. De Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Tran. Image Proc.*, vol. 14, no. 12, pp. 2117–2128, 2005.