A DCT Statistics-Based Blind Image Quality Index

Michele A. Saad, *Student Member, IEEE*, Alan C. Bovik, *Fellow, IEEE*, and Christophe Charrier

***Abstract*—The development of general-purpose no-reference approaches to image quality assessment still lags recent advances in full-reference methods. Additionally, most no-reference or blind approaches are distortion-specific, meaning they assess only a specific type of distortion assumed present in the test image (such as blockiness, blur, or ringing). This limits their application domain. Other approaches rely on training a machine learning algorithm. These methods however, are only as effective as the features used to train their learning machines. Towards ameliorating this we introduce the BLIINDS index (BLind Image Integrity Notator using DCT Statistics) which is a no-reference approach to image quality assessment that does not assume a specific type of distortion of the image. It is based on predicting image quality based on observing the statistics of local discrete cosine transform coefficients, and it requires only minimal training. The method is shown to correlate highly with human perception of quality.**

***Index Terms*—Anisotropy, discrete cosine transform, kurtosis, natural scene statistics, no-reference quality assessment.**

I. Introduction

THE ubiquity of digital visual information (in the form of images and video) in almost every economic sector necessitates reliable and efficient methods for assessing the quality of this visual information. While a number of full-reference image quality assessment (FR-IQA) methods have been established and have shown to perform well (correlating highly with subjective evaluation of image quality), no current no-reference image quality assessment (NR-IQA) algorithm exists that provides consistently reliable, generic performance. Despite the acceptable performance of current FR-IQA algorithms, the need for a reference signal limits their application, and calls for reliable no-reference algorithms. Existing approaches to NR-IQA research are varied and commonly follow one of three trends: 1) Distortion-specific approach: In this approach, the IQA algorithm quantifies a specific distortion in isolation of other factors, and scores an image accordingly. Examples of such NR-IQA algorithms are [1], which computes a *blockiness measure*, [2] and [3], which estimate *blur*, and [4] and [5] which measure *ringing* effects. 2) Feature extraction and learning approach: This approach extracts features from images and trains a learning algorithm to distinguish distorted from undistorted images based on the features extracted. Examples include a support vector machine (SVM) based approach in [6], and a neural networks based approach in [7]. 3) Natural scene statistics (NSS) approach: This approach assumes that natural or undistorted images occupy a subspace of the entire space of possible images, and then seeks to find a distance from the distorted image (which supposedly lies outside of that subspace) to the subspace of natural images. This approach relies on how the statistics of images change as distortions are introduced to them. An example of such an approach is described in [8].

The obvious disadvantage of the first approach is that it is distortion specific and hence also application specific. The number of distortions introduced to images in a wide range of applications is large, which makes it difficult for an algorithm to comprehensively quantify every type of distortion possible. The second approach is only as effective as the features extracted. The more representative the features are of the quality of the image, the more reliable the approach is. Finally, the NSS approach is a very promising one, but relies on extensive statistical modeling and reliable generalization of the models.

In this paper we present the BLIINDS index which is based on a combination of approaches 2) and 3). We seek to observe how certain perceptually relevant statistical features of images change as an image becomes distorted, and then use these features to train a statistical model that we develop to make blind (or no-reference) predictions about the quality of the images. The proposed NR-IQA method is based on a DCT framework entirely. This makes it computationally convenient, uses a commonly used transform, and allows a coherent framework. Our algorithm is tested on the LIVE database of images which contains JPEG2000, JPEG, white noise, Gaussian blur, and fast fading channel distortions. The fast fading channel errors in the LIVE database are simulated by JPEG2000 errors followed by channel errors. The proposed algorithm predicts a *differential mean opinion score* (DMOS) in the range [0,100], as is usually reported in subjective experiments, and as is provided by the LIVE database of images [9] and corresponding reported subjective DMOS scores. Providing a prediction of a quality score in the interval [0,100] makes the method convenient to compare against other algorithms that perform quality prediction and that report a similar type of continuous prediction score. mds

II. Image Representation: Feature Selection

A model is only as effective as the features it relies on to represent the data being modeled. In other words, its performance is a function of the representativeness of the features selected to represent the visual quality of the image being assessed. Consequently, the first issue we need to address is what type of features to extract from the image so as to capture as much of the visual quality of the image as possible.

Proceeding towards this goal, we first note that humans perform the task of blind IQA quite well, without the need for a reference image to do so. This leads us to believe that the human visual system (HVS) is sensitive to a number of features that

Manuscript received December 02, 2009; revised February 26, 2010. Date of publication March 15, 2010; date of current version April 30, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Dimitrios Androutsos.

M. A. Saad and A. C. Bovik are with the Laboratory of Image and Video Engineering, Department of Electrical and Computer Engineering, The University of Texas, Austin, TX 78712-0240 USA (e-mail: michele.saad@mail.utexas.edu; bovik@ece.utexas.edu).

C. Charrier is with The University of Caen Basse-Normandie, 14000 Caen, France (e-mail: Christophe.Charrier@unicaen.fr).

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.2010.2045550

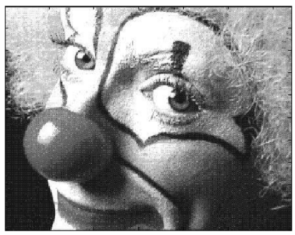

Fig. 1. Low contrast image.

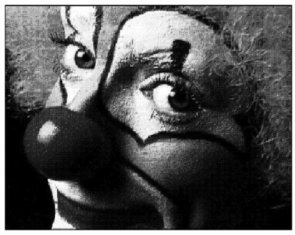

Fig. 2. High contrast image.

Fig. 3. Sharp object on blurred background.

distinguish high visual quality images from distorted ones. Further it is believed that the HVS has evolved to adapt to the statistics of the projected natural world, and therefore embodies mechanisms that process images in accordance with these statistics. Our feature selection process relies on the basic fundamental fact that natural images are highly structured (as was hypothesized in [10]), in the sense that their pixels exhibit strong dependencies, and these dependencies carry important information about the visual scene. Since it has been noted that the visual system is highly adapted to its natural environment, and since natural images are highly structured, the HVS is thus assumed to be adapted to extracting structural information from the visual field. Towards this end, we choose to extract image features that represent image *structure*. In addition to structure, the HVS is highly sensitive to contrast. In general, high contrast in an image is a desirable property, accentuating image structure and making the image more visually appealing. For instance, a majority of viewers might prefer Fig. 1 over Fig. 2 due to the higher contrast in Fig. 1 relative to Fig. 2. Consequently, we choose to extract a feature representative of the contrast of the images, the quality of which is to be assessed. Additionally, the features we extract are also expected to account for image *sharpness* (without explicitly measuring blur distortion or any other specific distortion) and *orientation anisotropies* (properties of images that the HVS is also highly sensitive to [11]). We note however, that image sharpness for instance, is highly content dependent. For example, the background is blurred in Fig. 3, yet it is a desirable property of this specific image. This is why we do not seek to quantify sharpness or blur as is, but rather explore how the statistics of spatial frequency domain characteristics vary in natural and in distorted images. To do so, we employ the discrete cosine transform (DCT) to extract a number of features and model their statistics.

A. DCT-Based Contrast

Contrast is a basic perceptual attribute of an image. One may distinguish between global contrast measures and ones that are computed locally (and possibly pooled into one measure post local extraction). Several measures of contrast exist in the literature such as the simple global Michelson contrast [12] and the more elaborate Peli's contrast [13]. The Michelson contrast statistics did not correlate with human visual perception of quality in our experiments, while Peli's contrast is computationally too intensive. Instead we resort to computing contrast based on local DCT patches, and show that the statistics of these correlate with human visual perception of distortion. This also conforms to our DCT-only framework.[1]

In the simplest, single-scale implementation, the 2-D DCT is applied to 17×17 image patches centered at every pixel in the image.[2] The local DCT contrast is then defined as the average of the ratio of the non-DC DCT coefficient magnitudes in the local patch normalized by the DC coefficient of that patch. The local contrast scores from all patches of the image are then pooled together by averaging the computed values to obtain a global image contrast value.

B. DCT-Based Structure Features

Structure features are derived locally from the local DCT frequency coefficients. We ignore the DC coefficient whose magnitude is usually much larger than the higher frequency DCT coefficients in the image patch. Ignoring the DC coefficient does not alter the local structural content of the image. We illustrate how the statistics of the higher frequency DCT coefficients change as an image becomes distorted in Fig. 4 and Fig. 5, which show the DCT coefficient histograms of a distortion free and a Gaussian blur distorted image, respectively. Similar trends in the histogram statistics are observed throughout the LIVE database of images, on which we perform our study. Among the observed differences in the histograms is the peakedness at zero, (the distorted images are observed to have a higher histogram peak at zero), and the variance along the support of the histogram. We seek to make use of statistical differences, such as the ones demonstrated above, to develop a NR-IQA index. To capture the statistical traits of the DCT histograms we compute its *kurtosis*, which quantifies the degree of its *peakedness* and tail weight, and is given by:

$$\kappa(x) = \frac{E(x-\mu)^4}{\sigma^4} \tag{1}$$

where μ is the mean of x, and σ is its standard deviation.

The kurtosis of each DCT image patch, (the same 17×17 image patches from which we computed local DCT contrast) is computed, and the resulting values pooled together by averaging the lowest tenth percentile of the obtained values to obtain a global image kurtosis value.

It has been hypothesized that degradation processes damage a scene's directional information. Consequently, anisotropy, which is a directionally dependent quality of images, was shown by Gabarda *et al.* in [11] to decrease as more degradation is added to the image. In [11] anisotropy is computed via the

[1]Transform-based contrast approaches have been studied in the literature and shown to correlate with human visual perception of contrast based on the fact that human contrast sensitivity is a function of spatial frequency [13].

[2]17 was chosen since it is a non-multiple of 4, 8, or 16 to avoid falling on block boundaries that may sometimes appear in JPEG, MPEG-2 or H.264 encoded images.

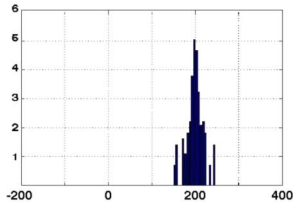

Fig. 4. Original image DCT log-histogram. Horizontal axis is non-DC DCT coefficient magnitudes.

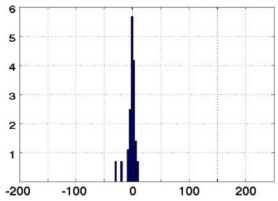

Fig. 5. JPEG2000 distorted image DCT log-histogram. Horizontal axis is non-DC DCT coefficient magnitudes.

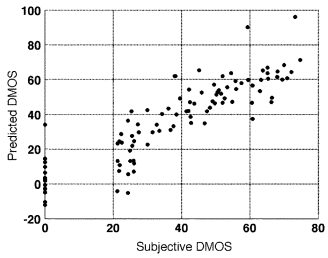

Fig. 6. Predicted DMOS versus subjective DMOS (JPEG2000 LIVE subset).

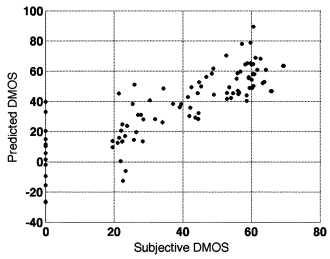

Fig. 7. Predicted DMOS versus subjective DMOS (JPEG LIVE subset).

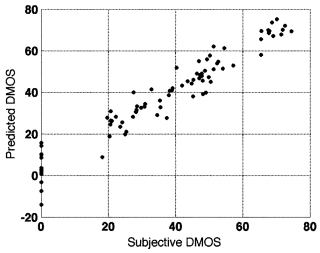

Fig. 8. Predicted DMOS versus subjective DMOS (White noise LIVE subset).

generalized Renyi entropy and a windowed pseudo-Wigner distribution (PWD). In this work, we compute a modified version of the anisotropy measure described in [11]. The anisotropy measure we compute is derived from one dimensional, 17×1 oriented DCT patches. Our anisotropy computation proceeds as follows: DCT image patches are computed along four different orientations (0°, 45°, 90° and 135°). Each patch consists of the DCT coefficients of 17 oriented pixel intensities. (We discard the DC coefficient, since the focus is on directional information). Let the DCT coefficients of a certain patch be denoted by $P[n,k]$, where k is the frequency index of the DCT coefficient $(1 < k \leq 17)$, and n is the spatial index where the

TABLE I
SPEARMAN CORRELATIONS (SUBJECTIVE DMOS VERSUS PREDICTED DMOS) FOR OUR DCT-BASED NR-IQA METHOD AND FOR FR PSNR

LIVE Subset	DCT-Based (NR-IQA)	PSNR (FR-IQA)
JPEG2000	0.9219	0.8765
JPEG	0.8391	0.8937
White Noise	0.9735	0.9560
Gaussian Blur	0.9569	0.8445
Fast Fading	0.7503	0.8617
All Data	0.7996	0.7810

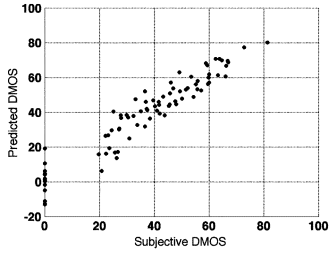

Fig. 9. Predicted DMOS versus subjective DMOS (Gaussian blur LIVE subset).

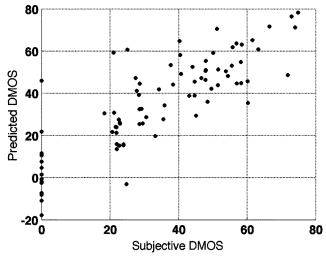

Fig. 10. Predicted DMOS versus subjective DMOS (Fast fading LIVE subset).

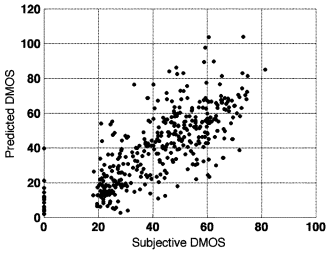

Fig. 11. Predicted DMOS versus subjective DMOS (Entire LIVE database).

DCT patch was computed. Each DCT patch is then subjected to a normalization of the form:

$$\tilde{P}_\theta[n,k] =] \frac{P_\theta[n,k]^2}{\sum_k P_\theta[n,k]^2} \tag{2}$$

where θ is one of the four orientations. The Renyi entropy for that particular image patch is then computed as

$$R_\theta[n] = -\frac{1}{2} log \left(\sum_k \tilde{P}_\theta[n,k]^3 \right). \tag{3}$$

Let M_θ be the number of image patches for orientation θ, then the average per orientation for all patches of orientation θ is obtained. This is denoted as $E[R_\theta]$. The variance across all four orientations (denoted as $var(E[R_\theta])$) along with the maximum $E[R_\theta]$ across the four orientations (denoted as $max(E[R_\theta])$) are chosen as measures of anisotropy.

III. MULTISCALE FEATURE EXTRACTION

Previous work in IQA has shown that extracting features and performing analysis at multiple scales can improve the quality assessment method. This is due to the fact that the perception

of image details depends on the image resolution, the distance from the image plane to the observer, and the acuity of the observer's visual system. A multiscale evaluation accounts for these variable factors. One example is the multiscale structural similarity index (MS-SSIM) [14] which outperforms the single scale SSIM index. We thus extract the same features described above at two scales. The features at the second scale are extracted, in the same manner as explained in the previous sections, after performing a down-sampling operation (by a factor of two in each spatial dimension) on the image in the spatial domain.

IV. Probabilistic Prediction Model

Let $X_i = [x_1, x_2, \ldots x_n]$ be the vector of features extracted from the image, where i is the index of the image being assessed, and n is the number of features extracted (in our case $n = 8$: 4 features at each scale for 2 scales). Additionally, let $DMOS_i$ be the subjective $DMOS$ associated with the image i. We model the distribution of the pair $(X_i, DMOS_i)$.

The probabilistic model is trained on a subset of the LIVE image database to determine the parameters of the probabilistic model by distribution fitting. Two probabilistic models are chosen and have been found to perform almost identically. These are the multivariate Gaussian distribution and the multivariate Laplacian distribution. These two models were chosen because of the simplicity with which they can be parameterized. Parameter estimation of these two models only requires the mean and covariance of the empirical data from the test set. The probabilistic model $P(X, DMOS)$ is designed by distribution fitting to the empirical data of the training set. The training and test sets are completely content independent, in the sense that no two images of the same scene are present in both sets. The LIVE database is derived from 29 reference images, the training set contains images derived from 15 reference images, and the test set contains the images derived from the other 14. The probabilistic model is then used to perform prediction by maximizing the quantity $P(DMOS_i/X_i)$. This is equivalent to maximizing the joint distribution of X and $DMOS$, $P(X, DMOS)$ since $P(X, DMOS) = P(DMOS/X)p(X)$.

V. Results

Our method was tested on Release 2 of the LIVE database of images [9]. The LIVE database consists of five subsets of 5 types of distortions: 1) JPEG2000 distortions (227 images), 2) JPEG distortions (233 images), 3) White noise distortions (174 images), 4) Gaussian blur distortions (174 images), and 5) Fast-fading Rayleigh channel distortions (which are simulated with JPEG2000 compression followed by channel bit-errors) (174 images).

Four features were extracted at two scales. These are 1) the average of the lowest 1% of the DCT coefficients kurtosis, 2) the average of the local DCT contrast values, 3) the DCT coefficient entropy variance across four orientations, and 4) the maximum DCT coefficient entropy across four orientations. Each of the features is raised to the power α_i to assign *importance* to the extracted features, where $1 \leq i \leq 8$ is the index of the feature, and $\sum_{i=1}^{8} \alpha_i = 1$. These exponents are determined from the Spearman correlations of each of the 8 features with subjective data on the training set. The features are then modeled by a multivariate Gaussian distribution.

To evaluate the method, the Spearman correlation was computed between the reported subjective DMOS and the DMOS predicted by our method. The results are displayed in Table I. We also provide the PSNR correlations as well to compare against, even though PSNR is a full-reference approach.[3] A plot of the predicted DMOS versus the subjective one for each of the data set subsets is shown in Figs. 6Figs. 9–11.

VI. Conclusion

In this paper we have developed BLIINDS, a new approach to NR-IQA and an exemplar algorithm that models the evolution of four features extracted from the DCT domain applied to local image patched at two spatial scales. The BLIINDS index correlates well with human visual perception, is computationally convenient as it is based on a DCT-framework entirely, and beats the performance of PSNR (which is a full reference approach). The probabilistic prediction model was trained on a small sample of the data (the training set), and only required the computation of the mean and the covariance of the training data. The MATLAB code for BLIINDS can be found on the LIVE webpage http://live.ece.utexas.edu/.

References

[1] Z. Wang, A. C. Bovik, and B. L. Evans, "Blind measurement of blocking artifacts in images," in *IEEE Int. Conf. Image Processing*, September 2000, vol. 3, p. 981984, IEEE.

[2] Z. M. Parvez Sazzad, Y. Kawayoke, and Y. Horita, "No-reference image quality assessment for jpeg2000 based on spatial features," *Signal Process.: Image Commun.*, vol. 23, no. 4, pp. 257–268, April 2008.

[3] X. Zhu and P. Milanfar, "A no-reference sharpness metric sensitive to blur and noise," in *QoMEX*, 2009.

[4] R. Barland and A. Saadane, "A new reference free approach for the quality assessment of mpeg coded videos," in *7th Int. Conf. Advanced Concepts for Intelligent Vision Systems*, Sep. 2005, vol. 3708, pp. 364–371.

[5] X. Feng and J. P. Allebach, "Measurement of ringing artifacts in jpeg images," in *Proc. SPIE*, Jan. 2006, vol. 6076, pp. 74–83.

[6] C. Charrier, G. Lebrun, and O. Lezoray, "A machine learning-based color image quality metric," in *Third Eur. Conf. Color Graphics, Imaging, and Vision*, June 2006, pp. 251–256.

[7] M. Jung, D. Léger, and M. Gazalet, "Univariant assessment of the quality of images," *J. Electron. Imag.*, vol. 11, no. 3, Jul. 2002.

[8] T. Brandao and M. P. Queluz, "No-reference image quality assessment based on DCT-domain statistics," *Signal Process.*, vol. 88, no. 4, pp. 822–833, April 2008.

[9] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, LIVE Image Quality Assessment Database Release 2 [Online]. Available: http://live.ece.utexas.edu/research/quality 2006

[10] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, April 2004.

[11] S. Gabarda and G. Cristóbal, "Blind image quality assessment through anisotroy," *J. Opt. Soc. Amer.*, vol. 24, no. 12, pp. B42–B51, Dec. 2007.

[12] H. Kukkonen, J. Rovamo, K. Tiipppana, and R. Nasanen, "Michelson contrast, RMS contrast, and energy of various spatial stimuli at threshold," *Vis. Res.*, vol. 33, no. 10, pp. 1431–1436, July 1993.

[13] E. Peli, "Contrast in complex images," *J. Opt. Soc. Amer.*, vol. 7, pp. 2032–2040, 1990.

[14] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity image quality assessment," in *37th Asilomar Conf. Signals, Systems, and Computers*, Nov. 2003, vol. 2, pp. 1398–1402.

[3]While PSNR has higher correlation on the JPEG and Fast Fading subsets, it is important to note that PSNR is a *full-reference* IQA approach. BLIINDS on the other hand is a *no-reference* approach and outperforms PSNR on the entire LIVE image database.