DCT statistics model-based blind image quality assessment

Michele Saad, Alan C. Bovik, Christophe Charrier

DCT STATISTICS MODEL-BASED BLIND IMAGE QUALITY ASSESSMENT

Michele A. Saad, Alan C. Bovik

The University of Texas at Austin
Department of Electrical and Computer Engineering
1 University Station C0803, Austin, TX 78712-0240, USA

Christophe Charrier

The University of Caen, Basse Normandie
Department of Electrical and Computer Engineering
Esplanade de la Paix, 14000 Caen, France

ABSTRACT

We propose an efficient, general-purpose, distortion-agnostic, blind/no-reference image quality assessment (NR-IQA) algorithm based on a natural scene statistics model of discrete cosine transform (DCT) coefficients. The algorithm is computationally appealing, given the availability of platforms optimized for DCT computation. We propose a generalized parametric model of the extracted DCT coefficients. The parameters of the model are utilized to predict image quality scores. The resulting algorithm, which we name BLIINDS-II, requires minimal training and adopts a simple probabilistic model for score prediction. When tested on the LIVE IQA database, BLIINDS-II is shown to correlate highly with human visual perception of quality, at a level that is even competitive with the powerful full-reference SSIM index.

Index Terms— No-reference image quality assessment, discrete cosine transform, natural scene statistics, generalized Gaussian density.

1. INTRODUCTION

The ubiquity of transmitted digital visual information (in the form of images and video) in every economic sector, and the broad range of applications that rely on it, such as PDAs, high definition televisions, internet video streaming, and video on demand, to name a few, necessitates means to evaluate the visual quality of this information. The various stages of the pipeline through which an image passes, introduce distortions to the image or modify it in one way or another, starting from its capture until its consumption by a viewer. The capture, digitization, compression, storage, transmission, and display processes all introduce modifications to the original image. These modifications, also termed distortions or impairments, may or may not be perceptually visible to the viewer. If they are visible, quantifying how perceptually annoying they are is an important process for improving Quality of Service (QoS) in the applications listed above. Since human raters are generally unavailable or too expensive for these applications, there is a significant need for objective IQA algorithms.

Only recently did full-reference image quality assessment (FR-IQA) methods reach a satisfactory level of performance, as demonstrated by high correlations with human subjective judgements of visual quality. SSIM [1], MS-SSIM [2], VSNR [3], and the VIF index [4] are examples of FR-IQA algorithms, to name a few. These methods require the availability of a reference signal against which to compare the test signal. In many applications, however, the reference signal is not available to perform a comparison against. This strictly limits the application domain of FR-IQA algorithms and points up the need for reliable blind/NR-IQA algorithms. However, no current NR-IQA algorithm exists that has been proven consistently reliable in performance.

Presently, NR-IQA algorithms generally follow one of three trends: 1) Distortion-specific approaches: These algorithms quantify one or more distortions such as blockiness [5], blur [6], or ringing [7] and score the image accordingly. 2) Training-based approaches: these train a model to predict the image quality score based on a number of features extracted from the image [8], [9]. 3) Natural scene statistics (NSS) approaches: these rely on the hypothesis that images of the natural world (i.e. distortion free images) occupy a small subspace in the space of all possible images and seek to find a *distance* between the test image and the subspace of natural images [10].

In this paper, we propose a framework that derives entirely from a simple statistical model of local DCT coefficients. We name our algorithm BLIINDS-II (BLind Image Integrity Notator using DCT Statistics). The new BLIINDS-II index greatly improves upon a preliminary algorithm (BLIINDS-I) [11], 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. We derive a generalized NSS-based model of local DCT coefficients, and transform the model parameters into features used for perceptual image quality score prediction. It is observed that the statistics of the DCT features change as the image quality changes. A generalized probabilistic model is obtained for these features, and is used to make probabilistic predictions of visual quality. We show that the method correlates highly

with human subjective judgements of quality.

The rest of the paper is organized as follows. In Section 2, we provide an overview of the method. In Section 3, we describe the model-based features. In Section 4, we describe the generalized probabilistic prediction model. We present the results in Section 5, and we conclude in Section 6.

2. OVERVIEW OF THE METHOD

The framework of the proposed approach is summarized in Fig. 1. An image entering the IQA "pipeline" is first subjected to local 2-dimensional DCT-transform coefficient computation. This stage of the pipeline consists of partitioning the image into equally sized nxn blocks, henceforth referred to as local image patches, and computing a local 2-dimensional DCT transform on each of the blocks. The coefficient extraction is performed locally in the spatial domain in accordance with the HVS's property of local spatial visual processing (i.e. in accordance with the fact that the HVS processes the visual space locally) [12]. As will be seen, this nxn DCT decomposition may be accomplished across scales. The second stage of the pipeline applies a generalized Gaussian density model to each block of DCT coefficients, as well as for specific partitions within each DCT block.

We next briefly describe the nxn DCT block partitions that are used. In order to capture *directional* information from the local image patches, the DCT block is partitioned directionally as shown in Fig. 2(a) into 3 *oriented* subregions. A generalized Gaussian fit is obtained for each of the *oriented* DCT-coefficient subregions. Another configuration for the DCT block partition is shown in Fig. 2(b). The partition reflects 3 radial frequency subbands in the DCT block. The upper, middle, and lower partitions correspond to the low frequency, mid-frequency, and high frequency DCT subbands respectively. A generalized Gaussian fit is obtained for each of the subregions as well.

The third step of the pipeline computes functions of the derived model parameters. These are the features used to predict image quality scores and are derived from the model parameters.

The fourth and final stage of the pipeline uses a simple Bayesian model to predict a quality score for the image. The Bayesian approach maximizes the probability that the image has a certain quality score given the model-based features extracted from the image. The posterior probability that the image has a certain quality score given the extracted features is modeled as a multidimensional generalized Gaussian distribution.

2.1. The Generalized Probabilistic Model

We model image features using a generalized Gaussian family of distributions which encompasses a wide range of observed behavior of distorted DCT coefficients.

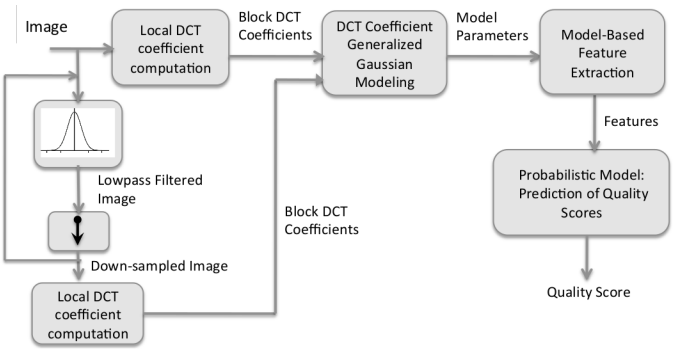

Fig. 1. High level overview of the BLIINDS-II framework

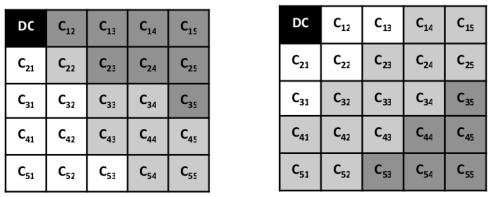

Fig. 2. (a) DCT coefficients, 3 bands, (b) DCT coefficients along 3 orientations

The univariate generalized Gaussian density is given by

$$\alpha e^{-(\beta|x-\mu|)^{\gamma}}, \tag{1}$$

where μ is the mean, β is the scale parameter, γ is the shape parameter, and Γ denotes the gamma function given by

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt, \tag{2}$$

and α and β are normalizing and scaling constants given by

$$\alpha = \frac{\beta\gamma}{2\Gamma(1/\gamma)}, \tag{3}$$

$$\beta = \frac{1}{\sigma}\sqrt{\frac{\Gamma(3/\gamma)}{\Gamma(1/\gamma)}}. \tag{4}$$

This family of distributions includes the Gaussian distribution ($\beta = 2$) and the Laplacian distribution ($\beta = 1$). As $\beta \to \infty$ the distribution converges to a uniform distribution.

3. MODEL-BASED DCT DOMAIN NSS FEATURES

3.1. The Generalized Gaussian Model Shape Parameter

We deploy a generalized Gaussian model of the non-DC DCT coefficients. In other words, we model the DCT coefficients in an nxn block, omitting the DC coefficient. The generalized Gaussian density in (1) is parametrized by mean μ, scale parameter β, and shape parameter γ. The shape parameter γ is used as a model-based feature. This feature is computed over all blocks in the image.

The overall shape parameter-based quality feature used is found by computing the lowest 10^{th} percentile average of the local block shape scores (γ) across the image. The reason for this pooling of the local histogram shape features, as opposed to simple averaging, is that percentile pooling has been observed to result in high correlations with subjective perception of quality [13]. Percentile pooling is motivated by the observation that the "worst" distortions in an image dominate subjective impressions. We choose 10% as a round number to avoid the possibility of "training". In addition, we compute the 100^{th} percentile average (regular mean) of the local γ scores across the image.

3.2. The Coefficient of Frequency Variation

The next feature is the *coefficient of frequency variation feature*:

$$\zeta = \frac{\sigma_{|X|}}{\mu_{|X|}}, \tag{5}$$

The feature ζ is computed for all blocks in the image. The highest 10^{th} percentile average and the mean (100^{th} percentile) of the local block scores across the image are then computed. Both pooling results (10% and 100%) are used as pooled features, since the difference between these is a compact but rich form of information.

3.3. Energy Subband Ratio Measure

Image distortions often modify the local spectral signatures of an image in unnatural ways. Towards measuring this, define a local DCT energy-subband ratio measure. Moving along diagonal lines on Fig. 2(b) from the top-left corner of the matrix to the bottom-right corner, the DCT coefficients represent higher radial spatial frequencies in the block. Consequently, we define 3 frequency bands in the block, as depicted in Fig. 2(b). Let Ω_n denote the set of coefficients belonging to band n, where $n = 1, 2, 3$, (lower, middle, higher). Then define the average energy in frequency band n as the model variance σ_n^2 corresponding to band n:

$$E_n = \sigma_n^2. \tag{6}$$

This is found by fitting the DCT data histogram in each band to the generalized Gaussian density model (1), then using the σ_n^2 value from the fit. We then compute the ratio of the difference between the average energy in frequency band n and the average energy up to frequency band n, over the sum of these two quantities:

$$R_n = \frac{|E_n - \frac{1}{n-1}\sum_{j<n} E_j|}{E_n + \frac{1}{n-1}\sum_{j<n} E_j} \tag{7}$$

This feature is computed for all blocks in the image. We compute the highest 10^{th} percentile average and the 100^{th} percentile average (regular mean) of the local block scores across all the image.

3.4. Orientation Model-Based Feature

Image distortions often modify local orientation energy in an unnatural manner. To capture directional information in the image that may correlate with changes in human subjective impressions of quality, we model block DCT coefficients along 3 orientations as shown in Fig. 2(a) below. The 3 shaded areas represent the DCT coefficients along 3 orientations. A generalized Gaussian model is fit to each shaded region in the block, and γ and ζ are obtained from the model histogram fits for each orientation. We then compute the variance of ζ along the 3 orientations. The variance of ζ across the 3 orientations from all the blocks in the image is then averaged (highest 10^{th} percentile and 100^{th} percentile) to obtain two numbers per image.

4. PREDICTION MODEL

We have found that a simple probabilistic predictive model is quite adequate for training the features used in BLIINDS-II. The efficacy of this simple predictor points up the power of the NSS-based features we have defined. Let $X_i = [x_1, x_2, ...x_m]$ be the vector of features extracted from the image, where i is the index of the image being assessed, and m is the number of features extracted (in our case $m = 10$ per scale). 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 IQA Database, which includes DMOS scores, to determine the parameters of the probabilistic model by distribution fitting. A multivariate generalized Gaussian model is used to model the data. Parameter estimation of the model only requires the mean and covariance of the empirical data from the training 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 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 $P(X, DMOS) = P(DMOS|X)p(X)$.

5. EXPERIMENTS AND RESULTS

It is well understood that images are naturally multiscale [4], [14], and that the early visual system involves decompositions over scales [12]. Towards this end, we implement the BLIINDS-II concept over multiple scales. Specifically, the feature extraction is repeated after lowpass filtering the image and subsampling it by a factor of 2. Prior to downsampling, the image is filtered by the rotationally symmetric discrete $3x3$ Gaussian filter kernel. This defines a multiscale feature

LIVE Subset	BLIINDS-II	BIQI	SSIM	PSNR
JPEG2000	0.9506	0.7995	0.9496	0.8658
JPEG	0.9411	0.8914	0.9664	0.8889
White Noise	0.9783	0.9510	0.9644	0.9791
GBlur	0.9435	0.8463	0.9315	0.7887
Fast Fading	0.9268	0.7067	0.9415	0.8986
ALL	0.9202	0.8190	0.9225	0.8669

Table 1. Median SROCC correlations for 1000 iterations of train and test sets (subjective DMOS vs predicted DMOS)

extraction approach, which enables BLIINDS-II to deal with changes in the image resolution, with distance from the image display to the observer, and with variations in the acuity of the observer's visual system.

BLIINDS-II was rigorously tested on the LIVE IQA Database [15] which contains 29 reference images, each impaired by many levels of 5 distortion types: JPEG2000, JPEG, white noise, Gaussian blur, and fast-fading channel distortions (simulated by JPEG2000 compression followed by channel bit errors.). The total number of distorted images (excluding the 29 reference images) is 779 images. Multiple train-test sequences were run. In each, the image database was subdivided into distinct training and test sets (completely content-separate). In each train-test sequence, 80% of the LIVE IQA Database content was chosen for training, and the remaining 20% for testing. Specifically, each training set contained images derived from 23 reference images, while each test set contained the images derived from the remaining 6 reference images. 1000 randomly chosen training and test sets were obtained and the prediction of the quality scores was run over the 1000 iterations. The code for BLIINDS-II is available at `http://live.ece.utexas.edu/research/quality/BLIINDS_release.zip`.

The Spearman rank-order correlation coefficient (SROCC) between predicted DMOS and subjective DMOS is reported in Table 1 for BLIINDS-II (at 3 scales), BIQI [16] (a recent NR-IQA method), and the *full-reference* SSIM and PSNR.

6. CONCLUSION

We have proposed a model-based, general (non-distortion specific) approach to NR-IQA using a minimal number of features extracted entirely from the DCT-domain which is also computationally convenient. We have shown that the new BLIINDS-II algorithm can be easily trained and it employs a simple probabilistic model for prediction. The method correlates highly with human visual perception of quality, and outperforms the *full-reference* PSNR measure and the recent no-reference BIQI index, and approaches the performance of the *full-reference* SSIM index.

7. REFERENCES

[1] 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.

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

[3] D.M. Chandler and S.S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284–2298, September 2007.

[4] H.R. Sheikh, A.C. Bovik, and G. de Veciana, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, February 2006.

[5] Z. Wang, A.C. Bovik, and B.L. Evans, "Blind measurement of blocking artifacts in images," in *IEEE Int'l Conf. Image Process.* IEEE, September 2000, vol. 3, pp. 981–984.

[6] 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.

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

[8] M. Jung, D. Lger, and M. Gazalet, "Univariant assessment of the quality of images," *J. Elect. Imaging*, vol. 11, no. 3, pp. 354–364, July 2002.

[9] C. Charrier, G. Lebrun, and O. Lezoray, "A machine learning-based color image quality metric," in *3rd Euro. Conf. Color Graphics, Imaging, Vision*, June 2006, pp. 251–256.

[10] 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.

[11] M.A. Saad, A.C. Bovik, and C. Charrier, "A DCT statistics-based blind image quality index," *IEEE Signal Process. Lett.*, vol. 17, no. 6, pp. 583–586, June 2010.

[12] R. Blake and R. Sekuler, *Perception*, McGraw Hill, 5th edition, 2006.

[13] M.H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Trans. Broadcasting*, vol. 10, no. 3, pp. 312–322,, September 2004.

[14] W. S. Geisler, "Visual perception and the statistical properties of natural scenes," *Ann. Revi. Psychology*, vol. 59, pp. 167–192, 2008.

[15] H.R. Sheikh, Z. Wang, L. Cormack, and A.C. Bovik, "LIVE image quality assessment database release 2," *http://live.ece.utexas.edu/research/quality*, 2006.

[16] A.K. Moorthy and A.C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 513–516, May 2010.