No-Reference Quality Assessment Using Natural Scene Statistics: JPEG2000

Hamid Rahim Sheikh, *Member, IEEE*, Alan Conrad Bovik, *Fellow, IEEE*, and Lawrence Cormack

***Abstract*—Measurement of image or video quality is crucial for many image-processing algorithms, such as acquisition, compression, restoration, enhancement, and reproduction. Traditionally, image quality assessment (QA) algorithms interpret image quality as similarity with a "reference" or "perfect" image. The obvious limitation of this approach is that the reference image or video may not be available to the QA algorithm. The field of blind, or no-reference, QA, in which image quality is predicted without the reference image or video, has been largely unexplored, with algorithms focussing mostly on measuring the blocking artifacts. Emerging image and video compression technologies can avoid the dreaded blocking artifact by using various mechanisms, but they introduce other types of distortions, specifically blurring and ringing. In this paper, we propose to use natural scene statistics (NSS) to blindly measure the quality of images compressed by JPEG2000 (or any other wavelet based) image coder. We claim that natural scenes contain nonlinear dependencies that are disturbed by the compression process, and that this disturbance can be quantified and related to human perceptions of quality. We train and test our algorithm with data from human subjects, and show that reasonably comprehensive NSS models can help us in making blind, but accurate, predictions of quality. Our algorithm performs close to the limit imposed on useful prediction by the variability between human subjects.**

***Index Terms*—Blind quality assessment (QA), image QA, JPEG2000, natural scene statistics (NSS), no reference (NR) image QA.**

I. Introduction

MEASUREMENT of visual quality is of fundamental importance to numerous image and video processing applications. Since human beings are the ultimate consumers of almost all of the image and video content in question, the obvious way of measuring quality is to solicit opinion from human observers. This, however, is not a feasible solution because this mechanism cannot be embedded into image and video processing systems whose goal is to maximize visual quality at a given cost. What is needed are automatic quality measurement methods that can assign quality scores to images or videos in meaningful agreement with subjective human assessment of quality.

The standard approach for quality assessment (QA) over the years has been that of image similarity or fidelity measurement. In full-reference (FR) QA, a "reference" image or video of perfect quality (zero loss of fidelity with the original scene) is always assumed to be available, and the loss of quality of a distorted image/video that arises as a result of some processing on the reference, is assumed to be related to its deviation from the perfect reference, but human beings do not need to have access to the reference to make judgements about quality. Given that human beings can do QA so easily without the reference, can we design computer algorithms to do the same? This is the problem of blind or no-reference (NR) QA, and it has recently received a great deal of attention, since the reference signal may not be available for many applications, or may be too expensive to provide. But given the limited success that FR QA has achieved,[1] it should come as little surprise that NR QA is a very hard problem, indeed, and is far from being a mature research area.

The problem of NR QA may seem hopelessly difficult at first. How can a computer be expected to assess the quality of an image or video and to ascertain the degradation in its information content without understanding it first? There is solace, however, in the fact that natural scenes belong to a small set in the space of all possible signals, and many researchers have developed statistical models to describe natural scenes, and that most distortions that are prevalent in image/video processing systems are not natural in terms of such statistics. Thus, one could consider the following alternative philosophy for NR QA: *all images are perfect, regardless of content, unless distorted during acquisition, processing, or reproduction*. This philosophy assigns equal quality to all natural visual stimuli that human beings could possibly come across, and the task of NR QA is reduced to blindly measuring the distortion (using signal or distortion models) that has possibly been introduced during the stages of acquisition, processing or reproduction, and then calibrating this measurement against human judgements of quality.

In this paper, we propose to use NSS models for assessing the quality of images and videos blindly. Using the above philosophy, we will demonstrate the use of an NSS model for blindly measuring the quality of images compressed by JPEG2000. We validate the performance of our algorithm using an extensive

Manuscript received April 3, 2003; revised September 2, 2004. This work was supported by a grant from the National Science Foundation. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Reiner Eschbach.

H. R. Sheikh was with the Laboratory for Image and Video Engineering, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712-1084 USA. He is now with Texas Instruments, Inc., Dallas, TX 75243 USA (e-mail: hamid.sheikh@ieee.org).

A. C. Bovik is with the Laboratory for Image and Video Engineering, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712-1084 USA (e-mail: bovik@ece.utexas.edu).

L. Cormack is with the Department of Psychology, The University of Texas at Austin, Austin, TX 78712-1084 USA (e-mail: cormack@psy.utexas.edu).

Digital Object Identifier 10.1109/TIP.2005.854492

[1]In a recent study published by the Video Quality Expert's Group (VQEG), the performance of the state-of-the-art FR QA algorithms for video was found to be statistically indistinguishable from the peak signal-to-noise ratio (PSNR) when tested across a wide variety of content and distortion types! The second phase of testing was conducted recently, and the draft report does announce proponents that perform better than the PSNR [1].

subjective QA study involving 198 images ranked by about two dozen subjects. This code and the study can be downloaded from [2].

II. Background

A. No-Reference Quality Assessment

Before the widespread use of compression systems, most NR distortion measurements were aimed at measuring the distortions inherent in acquisition or display systems, such as the blur introduced by optics of the capture or display devices, sensor noise, etc. But since the evolution of high-quality image and video acquisition and display devices, and the widespread use of digital images and videos, the emphasis has shifted toward distortions that are introduced during the stages of compression or transmission. Indeed, the experiments conducted by VQEG in Phase-I and Phase-II of the testing consisted of videos distorted mainly with compression artifacts and transmission errors [1].

Since the predominant mode for image and video coding and transmission is using block-based video compression algorithms, blind measurement of the blocking artifact has been the main emphasis of NR QA research [3]–[8]. Blocking, which typically arises from block-based compression algorithms running at low bit rates, is a very regular artifact consisting of periodic horizontal and vertical edges and is, therefore, relatively easy to detect and quantify [9]. Most blocking measurement methods quantify blocking either in the spatial domain [3]–[5] or in the frequency domain [6]–[8]. The main theme of these methods is that images and videos are mostly smooth and not blocky in nature. Hence, the presence of a periodic 8×8 edge structure is solely due to the blocking artifact. A quantification of this artifact is then adjusted for image statistics, such as spatial or temporal activity, to yield a final quality score for the image or video. Thus, only the distortion is explicitly modeled, while the signals are just assumed to be nonblocky and smooth.

The methods described above would obviously fail for any other distortion type, such as ringing or blurring introduced by, say the JPEG2000 image compression algorithm, or the H.264 video compression algorithm. In wavelet based compression algorithms [10]–[16], or the H.264 video compression standard [17], which includes a powerful deblocking filter, the blurring and ringing distortions are much harder to quantify without the reference. This is because blurring and ringing artifacts are strongly dependent on the image content as well as the strength of the distortion. Only a handful of researchers have attempted to quantify without reference the ringing and blurring artifacts that result from compression. Oguz *et al.* propose a visible ringing measure (VRM) that captures the ringing artifact around strong edges [18]. The algorithm is based on constructing an image mask that exposes only those parts of the image that are in the vicinity of strong edges, and the ringing measure is considered to be the pixel intensity variance around the edges in the masked image. However, the VRM has not been compared to (or calibrated against) human judgements of quality. Marziliano *et al.* present a NR blur metric that is based on measuring average edge transition widths, and this blur measure was used to predict the quality of JPEG2000 compressed images [19]. Li presents an algorithm that aims to measure several distortions present in an image blindly: global blur (based on assumed Gaussian blurs of step edges), additive white and impulse noise (based on local smoothness violation), blocking artifact (based on simple block boundary detection), and ringing artifact (based on anisotropic diffusion) [20]. The paper reports results of individual distortion measurements on one image only, with no subjective validation of the metrics reported.

In Section III-A, we present a novel way of assessing the quality of images afflicted with ringing and blurring distortion resulting from JPEG2000 compression. Our method is unique in that it uses NSS models to provide a "reference," against which the distorted images can be assessed. We believe that this is the first such attempt at doing NR QA for images compressed by JPEG2000 (or any other wavelet based) compression algorithm.

B. Natural Scene Statistics

Images and videos of the visual environment captured using high-quality capture devices operating in the visual spectrum are broadly classified as natural scenes. This differentiates them from text, computer generated graphics scenes, cartoons, and animations, paintings and drawings, random noise, or images and videos captured from nonvisual stimuli such as radar and sonar, X-rays, ultrasounds, etc. Natural scenes form an extremely tiny subset of the set of all possible scenes [21], [22]. Many researchers have attempted to understand the structure of this subspace of natural images by studying their statistics and developing statistical models for natural images [21]–[25]. Localized models using principal components analysis (PCA) and independent components analysis (ICA), including local statistics that attract human gaze, have revealed insightful similarities between the statistics of natural scenes and the physiology of the HVS [26]–[29]. Pertinent to our work are nonlinear multiresolution models of natural scenes, such as those in the wavelet domain, which capture the strong correlations that are present in images [23]–[25]. These models capture those nonlinear statistical dependencies that cannot be analyzed using linear methods. It has been claimed that these nonlinear dependencies could be explained solely by the presence of edges in natural images [30]. Furthermore, researchers have claimed that many statistical properties of natural images could be explained by the so-called dead leaves model, which is an image formation model where objects occlude one another [31], [32]. A review of recent NSS models is presented in [33].

NSS have been explicitly incorporated into a number of image processing algorithms: in compression algorithms [10]–[12], [25], denoising algorithms [23], [34], [35], image modeling [36], image segmentation [37], and texture analysis and synthesis [38].

While the characteristics of the distortion processes have been incorporated into NR QA algorithms, the assumptions about the statistics of the images that they afflict are usually quite simplistic. Specifically, most algorithms suffice to assume that the input images are smooth and low-pass in nature. In this paper, we adapt an NSS model using prior information about the distortion process (and, hence, incorporating distortion modeling as well) to construct a simplified model that can characterize

images compressed by JPEG2000 as well as uncompressed natural images. Using this model, we can quantify the departure of an image from natural behavior, and make predictions about its quality. The success of our method shows that human perceptions of image quality and the perceptibility of the distortion are indeed related to the naturalness of the image. This observation has similarly been reported by [39] for images distorted by chroma scaling and color temperatures of the display devices, where the researchers blindly measured the departure from naturalness by using chroma statistics of natural images.

III. Natural Scene Statistics Models for NR Quality Assessment: Application to Images Compressed by JPEG2000

As is apparent from Section II-A, there is a need to expand the scope of current NR QA algorithms. We will adhere to the philosophy of doing NR QA by doing blind distortion measurement with respect to a statistical model for natural images. We now present an algorithm that we have developed for NR QA for JPEG2000 compressed images.

A. JPEG2000 Image Compression Standard

Among the important compression algorithms that one comes across daily is the (lossy) JPEG image compression standard, which is based on the block-based discrete cosine transform (DCT). Over the years, researchers have found that greater compression could be achieved at the same visual quality if the DCT is replaced by the discrete wavelet transform (DWT), and research in wavelet-based image compression has resulted in the JPEG2000 compression standard [12].

While there are many interesting features of JPEG2000, we are only interested in the distortion process that occurs during compression, which is the chief source of visible artifacts in most cases. Although other sources may also be present, such as distortions resulting from bit errors during transmission, we ignore them in this work. In a nutshell, JPEG2000, operating in the baseline lossy mode, computes the DWT using the biorthogonal 9/7 wavelet [12]. The DWT coefficients are quantized using a scalar quantizer, with possibly different step sizes for each subband. This quantization causes many small DWT coefficients to become zero. The result is that the reconstruction from the quantized DWT coefficients contains blurring and ringing artifacts.

The key difference between the distortion introduced in JPEG2000 against that by JPEG is that in the case of JPEG2000, the distortion does not show regularity, such as periodic occurrence or directional preference (blocking occurs as horizontal/vertical edges every eight pixels). Rather, the distortion is image dependent, occurring mostly in highly textured areas or around strong edges, with the structure of the distortion being locally dependent on the image structure and the compression ratio. This lack of regularity in the distortion complicates the task of quantifying it without reference. However, if one were to use NSS models, especially those in the wavelet domain, one could identify and quantify the distortion introduced by the quantization of the wavelet coefficients.

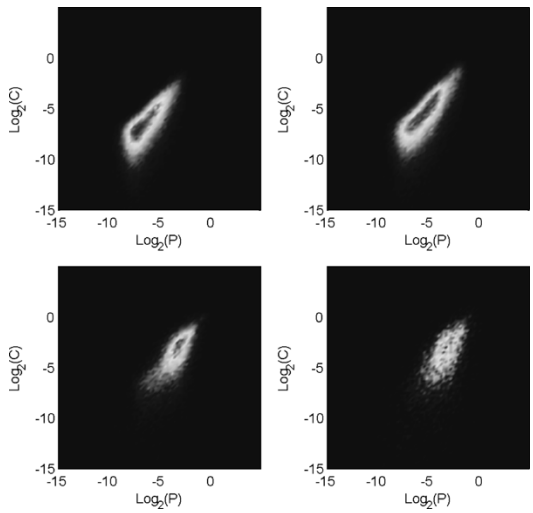


Fig. 1. Joint histograms of $(\log_2 P, \log_2 C)$ for an uncompressed natural image at different scales and orientations of its wavelet decomposition. Top left: Diagonal subband at the finest scale. Top right: Horizontal subband at the finest scale. Bottom left: Vertical subband at the second-finest scale. Bottom right: Diagonal subband at the third-finest scale.

B. Statistical Model for Natural Images in the Wavelet Domain

One particularly useful model for NSS has been presented in [24] and [25]. It captures the statistics of wavelet coefficients of natural images in a given subband and their correlations with other wavelet coefficients across scales and orientations. We noted that this model is suitable for measuring the effect of quantization of wavelet coefficients of natural images, since quantization pushes wavelet coefficients at finer scales toward zero. This results in a greater probability of zero coefficients in any subband than expected for natural images.

The statistical model proposed in [24] and [25] models the wavelet coefficient's magnitude, C, conditioned on the magnitude of the linear prediction of the coefficient, P, and is given in (1) where M and N are assumed to be independent zero-mean random variables

$$C = MP + N \tag{1}$$

$$P = \sum_{i=1}^{n} l_i C_i \tag{2}$$

where the coefficients C_i come from an n coefficient neighborhood of C in space, scale, and orientation, and l_i are linear prediction coefficients [25].

In [24] and [25], the authors use an empirical distribution for M and assume N to be Gaussian of unknown variance. The linear prediction P comes from a set of neighboring coefficients of C at the same scale and orientation, different orientations at the same scale, and coefficients at the parent scales. Zero-tree based wavelet image coding algorithms [10]–[12] also try to capture the nonlinear relationship of a coefficient with its parent scales and, hence, are conceptually similar to (1).

Fig. 1 shows the joint histograms of $(\log_2(P), \log_2(C))$ of an image at different scales and orientations. The strong nonlinear dependence between C and P is clearly visible on the logarithmic axes. As can be seen from the figure, the model can

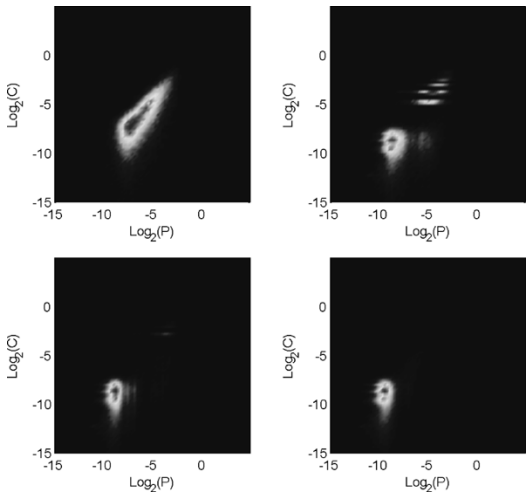


Fig. 2. Joint histograms of $(\log_2 P, \log_2 C)$ for one subband of an image when it is compressed at different bit rates using the JPEG2000 compression algorithm. Top left: No compression. Top right: 2.44 bits/pixel. Bottom left: 0.78 bits/pixel. Bottom right: 0.19 bits/pixel.

describe the statistics across subbands and orientations. The authors of [24] and [25] report (and we observed) that the model is stable across different images as well.

C. Compressed Natural Images

The model in (1) (Fig. 1) is not very useful for modeling compressed images, however, since quantization of the coefficients significantly affects the distribution. Fig. 2 shows the joint histograms of a subband from the uncompressed and compressed versions of an image at different bit rates. The effects of the quantization process are obvious, in that quantization pushes the coefficients toward zero, and disturbs the dependencies between C and P.[2]

We propose to use a simplified two-state model of natural scenes in the wavelet domain. These two states correspond to a coefficient or its predictor being significant or insignificant. The joint two-state model is motivated by the fact that the quantization process in JPEG2000, which occurs in all subbands, results in more of P and C values being insignificant than expected for natural images. Hence, a good indicator for unnaturalness and the perceptual effects of quantization is the proportion of significant P and C. Thus, if the proportion of significant P and C is low in an image, it could be a result of quantization in the wavelet domain.

The details of the model are as follows: Two image-dependent thresholds, one for P and the other for C, are selected for each subband for binarization. The details of threshold computation will be given in Section III-E. A coefficient (or its predictor) is

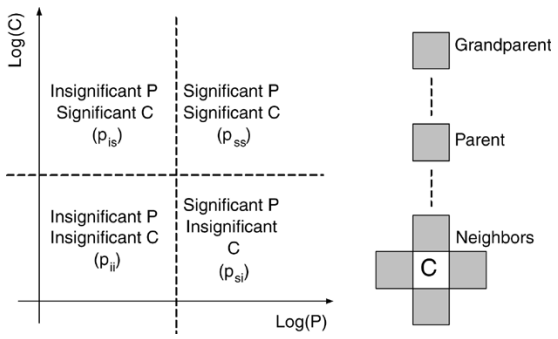


Fig. 3. Partition of the (P, C) space into quadrants. Also, the set of coefficients from which P is calculated in our simulations.

considered to be significant if it is above the threshold.[3] Consequently, we obtain a set of four empirical probabilities, p_{ii}, p_{is}, p_{si}, and p_{ss}, corresponding to the probabilities that the predictor/coefficient pair lies in one of the four quadrants, as depicted in Fig. 3. Obviously, the sum of all these probabilities for a subband is unity. The set of neighbors from which P is calculated in our simulations is also shown in Fig. 3.

D. Features for Blind Quality Assessment

We observed from experiments that the subband probabilities p_{ss} give the best indication of the loss of visual quality, in terms of minimizing the quality prediction error. In our simulations, we compute the p_{ss} feature from six subbands: horizontal, vertical, and diagonal orientations at the second-finest resolution; horizontal, vertical, and diagonal orientations at the finest resolution. Since the subband statistics are different for different scales and orientations, a nonlinear combination of these features is required. This is because the p_{ss} feature from a finer subband decreases faster with increasing compression ratio than the p_{ss} feature from a coarser subband. In order to compensate for these differences, we propose to nonlinearly transform each subband feature independently, before combining them linearly. The nonlinear transformations is designed to improve correspondence among the features coming from different subbands before a weighted average rule is applied.

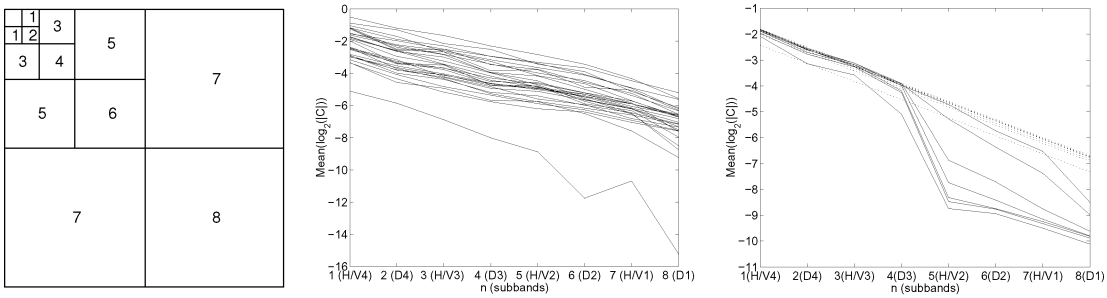
We train the transformations using training data by fitting image quality to the p_{ss} features from individual subbands as follows:

$$q_i = K_i \left(1 - \exp\left(-\frac{(p_{ss,i} - u_i)}{T_i}\right)\right) \tag{3}$$

where q_i is the transformed feature (predicted image quality) for the ith subband, $p_{ss,i}$ is the p_{ss} probability for the ith subband, and K_i, T_i, and u_i are curve-fitting parameters for the ith subband that are learned from the training data. Since each subband feature is mapped to subjective quality by a nonlinearity for that subband, the transformed features from all subbands are

[2]The histograms for compressed images shown in Fig. 2 have open bins to take care of $\log_2(0)$. Since we computed the DWT on reconstructed JPEG2000 images, the bulk of the insignificant coefficients are not exactly zero due to computational residues in the calculations of DWT/inverse-DWT and color transformations. However, our algorithm will not be affected by these small residues, as will become apparent shortly.

[3]The binarized model may remind some readers of two-state hidden Markov tree models for the wavelet coefficients of images, in which a coefficient is associated with the state of a two-state Markov process describing whether the coefficient is insignificant or significant [35]. However, our model is conceptually and computationally simpler, since it does not require complicated parameter estimation associated with hidden Markov tree models, but still provides good performance.



(a) (b) (c)

Fig. 4. $\mathrm{Mean}(\log_2(C))$ versus subband enumeration index. The means of horizontal and vertical subbands at a given scale are averaged. (a) The subband enumeration n used in (b) and (c). (b) Uncompressed natural images. $\mathrm{Mean}(\log_2(C))$ falls off approximately linearly with n. (c) $\mathrm{Mean}(\log_2(C))$ for an image at different bit rates (solid) and the corresponding linear fits (dotted). The fits are computed from $n = 1 \ldots 4$ only. Note that the estimated fits are quite close to the uncompressed image represented by the top most solid line. These linear fits are used for computing the image-dependent thresholds for the corresponding images.

approximately aligned with subjective quality and, hence, with each other, as well.

A weighted average of the transformed feature is used for quality prediction. Due to the similarity in the statistics of horizontal and vertical subbands at a particular scale, we constrain the weights to be the same for these orientations at a given scale. Thus, we modify the six-dimensional subband quality vector, $q = \{q_i | i \in 1 \ldots 6\}$, into a four-dimensional vector $q\prime$ by averaging the quality predictions from horizontal and vertical subbands at a given scale, and the final quality prediction is taken to be a weighted average of $q\prime$

$$\begin{bmatrix} q\prime_1 \\ q\prime_2 \\ q\prime_3 \\ q\prime_4 \end{bmatrix} = \begin{bmatrix} \frac{(q_1+q_2)}{2} \\ q_3 \\ \frac{(q_4+q_5)}{2} \\ q_6 \end{bmatrix}$$

$$Q = q\prime^T w \tag{4}$$

where the weights w are learned by minimizing quality prediction error over the training set.

E. Image-Dependent Threshold Calculations

One important prerequisite for a feature for QA is that it should adequately reflect changes in visual quality due to compression while being relatively robust to changes in the image content. However, the feature proposed in Section III-C varies not only with the amount of quantization, but also with the variation in image content. For example, if the thresholds were held constant, then a low p_{ss} feature could either signify a heavily quantized image, or a relatively smooth image (say the sky and the clouds) that was lightly quantized. In order to make our feature more robust to changes in the image content, we propose to adapt the thresholds, so that they are lower for smooth images, and higher for highly textured images. Once again, we will use NSS models in conjunction with distortion modeling to achieve this goal.

It is an interesting observation that when the means of ($\log_2$ of) subband coefficient amplitudes are plotted against an *enumeration* of the subbands, the plot is approximately linear. This is shown in Fig. 4(b) for a number of uncompressed natural images. The graphs have approximately the same slope, while the intercept on the y-axis varies from image to image. This approximately linear fall-off is expected for natural images since it is well known that natural image amplitude spectra fall off approximately as $1/f$, which is a straight line on log–log axes. Another aspect of NSS is that horizontal and vertical subbands have approximately the same energy, whereas the diagonal subbands have lower energy at the same scale. The interesting observation is that a diagonal subband sits between the horizontal/vertical subband at its scale and the one finer to it.

Quantized images are not natural, however, and, hence, the corresponding plots for them will not have approximately linear fall off [Fig. 4(c), solid lines]. However, we know that the quantization process in wavelet based compression systems (such as the JPEG2000) is designed in such a way that the subband means for coarser subbands are less affected by it, whereas the means for finer subbands are affected more. Hence, from the coarser subbands, one could predict the line that describes the energy fall off for the image by estimating its intercept (assuming that all natural images have the same slope). This line yields the estimated means for the finer subbands in the unquantized image from which the compressed image whose quality is being evaluated was derived. This is shown in Fig. 4(c) as well, where the means of ($\log_2$ of) subband coefficients are plotted for an image compressed at different bit rates (as well as the uncompressed image). Notice that predicted subband means (shown by dotted lines) are quite close to the actual means of the uncompressed image (top solid line).

We use the above observation to calculate the image dependent thresholds as follows:

$$\text{Threshold} = \text{Estimated Subband Mean} + \text{Offset}. \tag{5}$$

The slope of the line can be learned from uncompressed natural images in the training set, while the offsets (one for P and one for C for each subband) can be learned by numerical minimization that attempts to minimize the error in the quality predictions over the training set. In this way, our algorithm utilizes NSS models in concert with modeling the salient features of the distortion process to make the QA feature more robust to changes in the image content.

F. Simplified Marginal Model

Since the computation of coefficient predictions P is expensive, we also consider the marginal distribution of the binarized wavelet coefficients C, as opposed to the joint distribution of

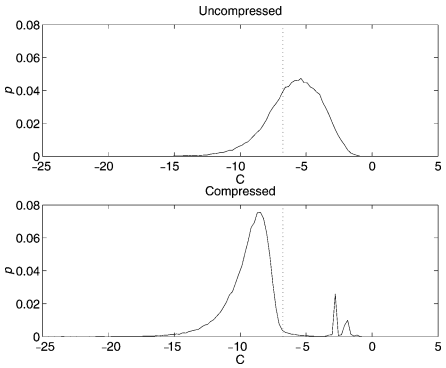


Fig. 5. Histograms of the logarithm (to the base 2) of the horizontal subband the finest scale for one image, before and after compression (at 0.75 bits per pixel). The dotted line denotes the threshold at -6.76.

binarized P and C. Fig. 5 shows the histogram of the logarithm (to the base 2) of the horizontal wavelet coefficient magnitude for one image, and the histogram for the same compressed image at the same scale and orientation. Quantization shifts the histogram of $\log_2(C)$ toward lower values. We divide the histogram into two regions: insignificant coefficients and significant coefficients. Again, the probability of a coefficient being in one of the regions at a certain scale and orientation is a good feature to represent the effect of quantization. In [40], the probabilities for different subbands constituted the feature vector whose dimensionality was reduced by PCA to one dimension, which was then fitted to perceptual quality. In this paper, we map the probabilities by (3) and then take a weighted average as in (4).

G. Subjective Experiments for Training and Testing

In order to calibrate, train, and test the NR QA algorithm, an extensive psychometric study was conducted [2]. In these experiments, a number of human subjects were asked to assign each image with a score indicating their assessment of the quality of that image, defined as the extent to which the artifacts were visible. Twenty nine high-resolution 24-bits/pixel RGB color images (typically 768×512) were compressed using JPEG2000 with different compression ratios to yield a database of 198 images, 29 of which were the original (uncompressed) images. Observers were asked to provide their perception of quality on a continuous linear scale that was divided into five equal regions marked with adjectives "Bad," "Poor," "Fair," "Good," and "Excellent." About 25 human observers rated each image. The raw scores for each subject were converted to Z scores [41] and then scaled and shifted to the full range (1–100). Mean opinion scores (MOS) were then computed for each image. The average root mean-squared error (RMSE) between the processed subject scores and MOS (averaged over all images) was found to be 7.04 (on a scale of 1–100), and the average linear correlation coefficient over all subjects was 0.92. Further details about the experiment and the processing of raw scores is available at [2].

IV. Results

A. Simulation Details

For training and testing, the database is divided into two parts. The training database consists of fifteen randomly selected images (from the total 29) and all of their distorted versions. The testing database consists of the other fourteen images and their distorted versions. This way, there is no overlap between the training and the testing databases. The algorithm is run several times, each time with a different (and random) subset of the original 29 images for training and testing, with 15 and 14 images (and their distorted versions) in the training and testing sets, respectively.

The algorithm is run on the luminance component of the images only, which is normalized to have a root-mean-squared (RMS) value of 1.0 per pixel. The biorthogonal 9/7 wavelet with four levels of decomposition is used for the transform. The slope of the line for estimating the subband coefficient means in (5) is learned from the uncompressed images in the training set. The weights w in (4) are learned using nonnegatively constrained least-squares fit over the training data (MATLAB command *lsqnonneg*). The minimization over the threshold offsets in (5), as well as for the fitting parameters in (3), is done by unconstrained nonlinear minimization (MATLAB command *fminsearch*).

B. Quality Calibration

It is generally acceptable for a QA method to stably predict subjective quality within a nonlinear mapping, since the mapping can be compensated for easily. Moreover, since the mapping is likely to depend upon the subjective validation/application scope and methodology, it is best to leave it to the final application, and not to make it part of the QA algorithm. Thus, in both the VQEG Phase-I and Phase-II testing and validation, a nonlinear mapping between the objective and the subjective scores was allowed, and all the performance validation metrics were computed *after* compensating for it [1]. This is true for the results presented in this section, where a five-parameter nonlinearity (a logistic function with additive linear term) is used. The mapping function used is given in (6), while the fitting was done using MATLAB's *fminsearch*

$$\text{Quality}(x) = \beta_1 \text{logistic}\left(\beta_2, (x - \beta_3)\right) + \beta_4 x + \beta_5 \quad (6)$$

$$\text{logistic}(\tau, x) = \frac{1}{2} - \frac{1}{1 + \exp(\tau x)}. \quad (7)$$

C. Results

Fig. 6(a) shows the predictions of the algorithm on the testing data for one of the runs against the MOS for the joint binary simplification of the model. Fig. 6(b) shows the normalized histogram of the RMSE (which we use as a measure of the performance of the NR metric) between the quality prediction and the MOS, for a number of runs of the algorithm. The mean RMSE is 8.05, with a standard deviation of 0.66. The average linear correlation coefficient between the prediction and the MOS for all the runs is 0.92 with a standard deviation of 0.013. Fig. 6(c) shows the RMSE histogram for the marginal binary simplification of the model. The mean RMSE is 8.54 and the standard deviation is 0.83. The average linear correlation coefficient is 0.91 with a standard deviation of 0.018.

The normalized histograms of the correlation coefficient for the joint and marginal models are shown in Fig. 7(a) and (b).

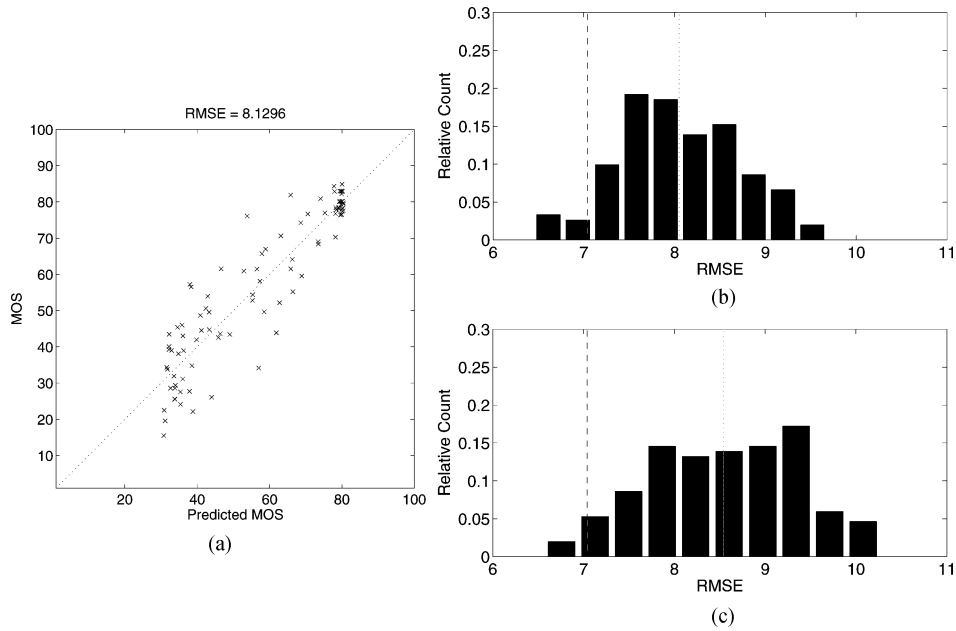


Fig. 6. Results. (a) Quality predictions versus MOS for one run of the algorithm. Normalized histograms of the RMSE for several runs of the algorithm using the joint statistics of P and C (b) and using the marginal statistics of C only (c). The dotted line in the histograms show the mean value. The dashed line shows the average standard deviation of MOS scores.

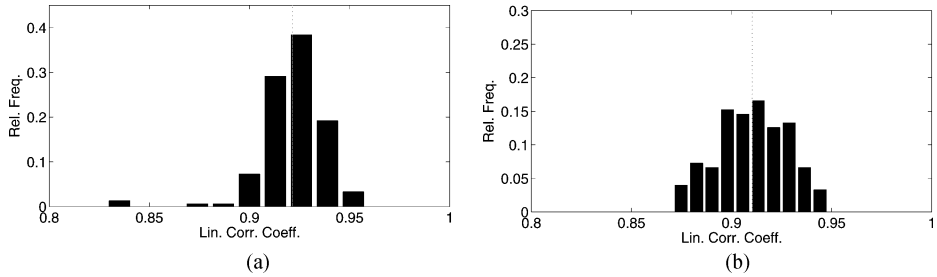


Fig. 7. Results. Normalized histograms of the linear correlation coefficient for several runs of the algorithm using the joint statistics of P and C (a) and using the marginal statistics of C only (b). The dotted lines in the histograms show the mean value.

D. Discussion

It is apparent from the above figures that our NR algorithm is able to make predictions of the quality of images compressed with JPEG2000 that are consistent with human evaluations. The average error in quality assignment for a human subject is 7.04, while, for our algorithm, it is 8.05. *We are, therefore, performing close to the limit imposed on useful prediction by the variability between human subjects*, that is the variability between our algorithm and MOS is comparable to the variability between a human and the MOS on average. The average gap between an average human and the algorithm is only about 1.0 on a scale of 1–100. Another interesting figure is the standard deviation of the RMSE of 0.66 on a scale of 1–100, which indicates that our algorithm's performance is stable to changes in the training database. As a comparison, we can compare the performance of our algorithm against a NR perceptual blur metric presented in [19]. This blur metric was not specifically designed for JPEG2000, but nevertheless reports a correlation coefficient of 0.85 on the same database. It is also insightful to compare the performance of our method against PSNR. This comparison is unfair since PSNR is a full-reference QA method, while the proposed method is NR. Nevertheless, researchers are often interested in comparing QA methods against it. On the same database, we obtain an RMSE[4] of 7.63 for the entire database (including the reference images), and a linear correlation coefficient of 0.93.

It is also interesting to analyze the performance of the algorithm for different images in order to expose any content-dependence. Fig. 8(a) shows the prediction-MOS graph for one image for which the algorithm consistently performed well over the entire range of quality. Fig. 8(b) shows the case where the algorithm consistently over-predicts the quality and Fig. 8(c) shows the case where the algorithm consistently under-predicts the quality. We believe that the spatial *spread* of image details and texture over the image affects the performance of our algorithm and makes it content dependent. We are working on the development of a spatially adaptive version of our algorithm.

[4]We used a four-parameter logistic curve for the nonlinear regression on PSNR, since the data set on which the PSNR was calculated *included* the reference images as well. The saturating logistic nonlinearity could be used for distorted as well as reference images. However, inclusion of the reference images unfairly biases the results in favor of PSNR, but we report this number since our NR algorithm included the reference images as well.

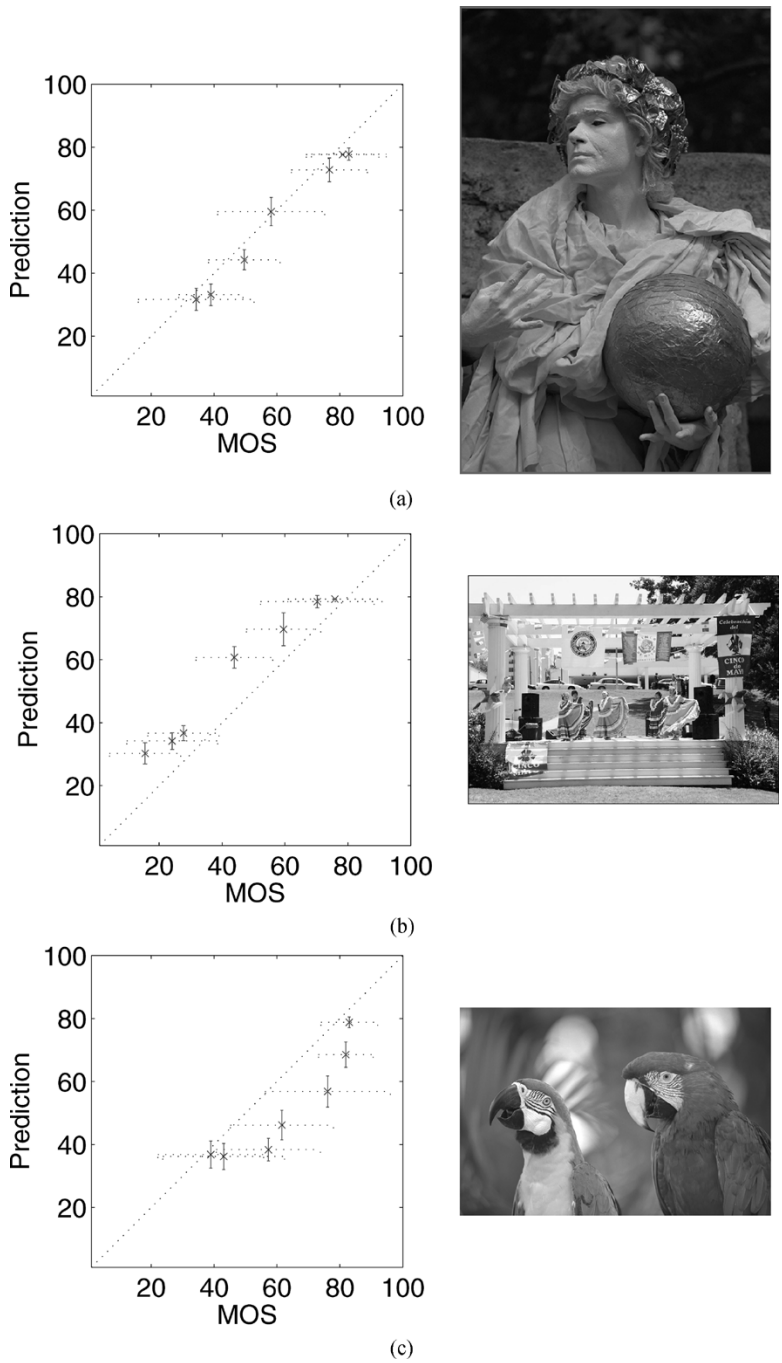


Fig. 8. Content-dependence of the performance of the proposed NR QA algorithm over the quality range. (a) Algorithm performs consistently well over the range of quality. (b) The algorithm consistently over-predicts the quality and (c) the algorithm consistently under-predicts the quality. The MOS-Prediction point "$\times$" and the ± 2 standard deviation intervals are shown for each figure. The variance in the MOS is due to inter-human variability and the variation in the prediction is due to changes in the training set.

E. Implementation Complexity

The computational complexity of our NR algorithm is small, especially our simplification based on the wavelet coefficients marginals. Our algorithm takes about 20 s per image for the joint model and about 5 s per image for the marginal model with un-optimized MATLAB implementation running on a Pentium-III, 1-GHz machine. If the DWT coefficients are already available, say directly from the JPEG2000 stream, the marginal model computation reduces essentially to counting the number of significant coefficients, which is computationally very efficient.

F. Future Improvements

We are continuing efforts into improving the performance of our algorithm to further narrow the gap between human and machine assessment. To this end, we are investigating a number of extensions.

1) Although NSS models are global, images seldom have stationary characteristics. Our algorithm is, as yet, blind

to the spread of spatial detail in images, or to the presence of different textures within an image, which may have different statistics from one another. We are currently working on incorporation of spatial adaptation into our algorithm by constructing "quality maps," and investigations into optimal pooling methods could improve algorithm performance.

2) Our simplification of the NSS model reduces continuous variables to binary ones. More sophisticated models for characterizing natural images and their compressed versions could result in more stable features and improved algorithm performance.

V. Conclusion

In this paper, we proposed to use NSS for measuring the quality of images and videos without the reference image. We claimed that since natural images and videos come from a tiny subset of the set of all possible signals, and since the distortion processes disturb the statistics that are expected of them, a deviation of a signal from the expected natural statistics can be quantified and related to its visual quality. We presented an implementation of this philosophy for images distorted by JPEG2000 (or any other wavelet based) compression. JPEG2000 compression disturbs the nonlinear dependencies that are present in natural images. Adapting a nonlinear statistical model for natural images by incorporating quantization distortion modeling, we presented an algorithm for quantifying the departure of compressed images from expected natural behavior, and calibrated this quantification against human judgements of quality. We demonstrated our metric on a data set of 198 images, calibrated and trained our algorithm on data from human subjects, and demonstrated the stability and accuracy of our algorithm to changes in the training/testing content. Our algorithm performs close to the limit imposed by variability between human subjects on the prediction accuracy of an NR algorithm. Specifically, we achieved an average error of 8.05 (on a scale of 1–100) from the MOS, whereas subjective human opinion is expected be deviant by 7.04 from the MOS on average.

We feel that NSS modeling needs to be an important component of image and video processing algorithms that operate on natural signals. We are continuing efforts into designing NR QA algorithms that can predict quality for a broader class of distortion types, such as noise, wireless channel errors, watermarking, etc., by posing the problem in an estimation-theoretic framework based on statistical models for natural scenes.

References

[1] (2000) Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment. VQEG. [Online]. Available: http://www.vqeg.org/

[2] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik. (2003) LIVE Image Quality Assessment Database. [Online]. Available: http://live.ece.utexas.edu/research/quality

[3] H. R. Wu and M. Yuen, "A generalized block-edge impairment metric for video coding," *IEEE Signal Process. Lett.*, vol. 4, no. 11, pp. 317–320, Nov. 1997.

[4] V.-M. Liu, J.-Y. Lin, and K.-G. C.-N. Wang, "Objective image quality measure for block-based DCT coding," *IEEE Trans. Consum. Electron.*, vol. 43, no. 3, pp. 511–516, Jun. 1997.

[5] L. Meesters and J.-B. Martens, "A single-ended blockiness measure for JPEG-coded images," *Signal Process.*, vol. 82, pp. 369–387, 2002.

[6] Z. Wang, A. C. Bovik, and B. L. Evans, "Blind measurement of blocking artifacts in images," in *Proc. IEEE Int. Conf. Image Processing*, vol. 3, Sep. 2000, pp. 981–984.

[7] K. T. Tan and M. Ghanbari, "Frequency domain measurement of blockiness in MPEG-2 coded video," in *Proc. IEEE Int. Conf. Image Processing*, vol. 3, Sep. 2000, pp. 977–980.

[8] A. C. Bovik and S. Liu, "DCT-domain blind measurement of blocking artifacts in DCT-coded images," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, vol. 3, May 2001, pp. 1725–1728.

[9] M. Yuen and H. R. Wu, "A survey of hybrid MC/DPCM/DCT video coding distortions," *Signal Process.*, vol. 70, no. 3, pp. 247–278, Nov. 1998.

[10] J. M. Shapiro, "Embedded image coding using zerotrees of wavelets coefficients," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3445–3462, Dec. 1993.

[11] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 6, pp. 243–250, Dec. 1996.

[12] D. S. Taubman and M. W. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards, and Practice*. Norwell, MA: Kluwer, 2001.

[13] C. Podilchuk, N. Jayant, and N. Farvardin, "Three-dimensional subband coding of video," *IEEE Trans. Image Process.*, vol. 4, no. 2, pp. 125–139, Feb. 1995.

[14] K. S. Shen and E. J. Delp, "Wavelet based rate scalable video compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 1, pp. 109–122, Jan. 1999.

[15] S. Cho and W. A. Pearlman, "A full-featured, error-resilient, scalable wavelet video codec based on the set partitioning in hierarchical trees (SPIHT) algorithm," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 3, pp. 157–171, Mar. 2002.

[16] B.-J. Kim and W. A. Pearlman, "An embedded wavelet video coder using three-dimensional set partitioning in hierarchical trees (SPIHT)," in *Proc. Data Compression Conf.*, 1997, pp. 251–260.

[17] "Special issue on the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 557–725, Jul. 2003.

[18] S. H. Oguz, Y. H. Hu, and T. Q. Nguyen, "Image coding ringing artifact reduction using morphological post-filtering," in *Proc. IEEE 2nd Workshop Multimedia Signal Processing*, 1998, pp. 628–633.

[19] 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, Feb. 2004.

[20] X. Li, "Blind image quality assessment," presented at the IEEE Int. Conf. Image Processing, Rochester, NY, Sep. 2002.

[21] D. L. Ruderman, "The statistics of natural images," *Network: Comput. Neur. Syst.*, vol. 5, no. 4, pp. 517–548, Nov. 1994.

[22] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Amer.*, vol. 4, no. 12, pp. 2379–2394, 1987.

[23] M. J. Wainwright, E. P. Simoncelli, and A. S. Wilsky, "Random cascades on wavelet trees and their use in analyzing and modeling natural images," *Appl. Comput. Harmon. Anal.*, vol. 11, pp. 89–123, 2001.

[24] E. P. Simoncelli, "Statistical models for images: Compression, restoration and synthesis," presented at the IEEE Asilomar Conf. Signals, Systems, and Computers, Nov. 1997.

[25] R. W. Buccigrossi and E. P. Simoncelli, "Image compression via joint statistical characterization in the wavelet domain," *IEEE Trans. Image Process.*, vol. 8, no. 12, pp. 1688–1701, Dec. 1999.

[26] P. J. B. Hancock, R. J. Baddeley, and L. S. Smith, "The principal components of natural images," *Network: Comput. Neur. Syst.*, vol. 3, pp. 61–70, 1992.

[27] P. Reinagel and A. M. Zador, "Natural scene statistics at the center of gaze," *Network: Comput. Neur. Syst.*, vol. 10, pp. 1–10, 1999.

[28] J. H. van Hateren and A. van der Schaaf, "Independent component filters of natural images compared with simple cells in primary visual cortex," in *Proc. Roy. Soc. London, Ser. B*, 1998, pp. 265:359–366.

[29] U. Rajashekar, L. K. Cormack, A. C. Bovik, and W. S. Geisler, "Image properties that draw fixation [abstract]," *J. Vis.*, vol. 2, no. 7, p. 730a, 2002.

[30] D. L. Donoho and A. G. Felisia, "Can recent innovations in harmonic analysis 'explain' key findings in natural image statistics," *Vis. Res.*, vol. 12, no. 3, pp. 371–393, 2001.

[31] A. B. Lee, D. Mumford, and J. Huang, "Occlusion models for natural images: A statistical study of a scale-invariant dead leaves model," *Int. J. Comput. Vis.*, vol. 41, no. 1/2, pp. 35–59, 2001.

[32] M. G. A. Thomson, "Beats, kurtosis and visual coding," *Network: Comput. Neur. Syst.*, vol. 12, pp. 271–287, 2001.
[33] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S.-C. Zhu, "On advances in statistical modeling of natural images," *J. Math. Imag. Vis.*, vol. 18, pp. 17–33, 2003.
[34] M. K. Mihçak, I. Kozintsev, K. Ramachandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Process. Lett.*, vol. 6, no. 12, pp. 300–303, Dec. 1999.
[35] J. K. Romberg, H. Choi, and R. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden Markov models," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 1056–1068, Jul. 2001.
[36] E. Y. Lam and J. W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1661–66, Oct. 2000.
[37] H. Choi and R. G. Baraniuk, "Multiscale image segmentation using wavelet-domain hidden Markov models," *IEEE Trans. Image Process.*, vol. 10, no. 9, pp. 1309–1321, Sep. 2001.
[38] J. Portilla and E. P. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *Int. J. Comput. Vis.*, vol. 40, no. 1, pp. 49–71, 2000.
[39] T. J. W. M. Jannsen and F. J. J. Blommaert, "Predicting the usefulness and naturalness of color reproductions," *J. Imag. Sci. Technol.*, vol. 44, no. 2, pp. 93–104, 2000.
[40] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "Blind quality assessment for JPEG2000 compressed images," presented at the IEEE Asilomar Conf. Signals, Systems, and Computers, Nov. 2002.
[41] A. M. van Dijk, J. B. Martens, and A. B. Watson, "Quality assessment of coded images using numerical category scaling," *Proc. SPIE*, vol. 2451, pp. 90–101, Mar. 1995.

Hamid Rahim Sheikh (S'93–M'04) received the B.S. degree in electrical engineering from the University of Engineering and Technology, Lahore, Pakistan, in 1998, and the M.S. and Ph.D. degrees from The University of Texas at Austin in 2001 and 2004, respectively.

His research interests include full-reference and no-reference quality assessment, application of natural scene statistics models and human visual system models for solving image and video processing problems, and image and video codecs and their embedded implementation.

Alan Conrad Bovik (S'80–M'81–SM'89–F'96) received the B.S., M.S., and Ph.D. degrees in electrical and computer engineering from the University of Illinois, Urbana-Champaign, in 1980, 1982, and 1984, respectively.

He is currently the Cullen Trust for Higher Education Endowed Professor in the Department of Electrical and Computer Engineering, The University of Texas at Austin, where he is the Director of the Laboratory for Image and Video Engineering (LIVE) in the Center for Perceptual Systems. During the Spring of 1992, he held a visiting position in the Division of Applied Sciences, Harvard University, Cambridge, MA. He is the editor/author of the *Handbook of Image and Video Processing* (New York: Academic, 2000). His research interests include digital video, image processing, and computational aspects of visual perception, and he has published over 350 technical articles in these areas and holds two U.S. patents.

Dr. Bovik was named Distinguished Lecturer of the IEEE Signal Processing Society in 2000, received the IEEE Signal Processing Society Meritorious Service Award in 1998, the IEEE Third Millennium Medal in 2000, the University of Texas Engineering Foundation Halliburton Award in 1991, and is a two-time Honorable Mention winner of the International Pattern Recognition Society Award for Outstanding Contribution (1988 and 1993). He was named a Dean's Fellow in the College of Engineering in 2001. He is a Fellow of the IEEE and has been involved in numerous professional society activities, including Board of Governors, IEEE Signal Processing Society, 1996–1998; Editor-in-Chief, IEEE TRANSACTIONS ON IMAGE PROCESSING, 1996–2002; Editorial Board, The PROCEEDINGS OF THE IEEE, 1998–present; and Founding General Chairman, First IEEE International Conference on Image Processing, held in Austin in November 1994. He is a Registered Professional Engineer in the State of Texas and is a frequent consultant to legal, industrial, and academic institutions.

Lawrence Cormack received the B.S. degree from the University of Florida, Gainesville, and the Ph.D. degree in physiological optics from the University of California, Berkeley.

He is currently an Associate Professor of psychology and neuroscience at the University of Texas (UT) at Austin, an Adjunct Associate Professor of vision science at the University of Houston, Houston, TX, and an active member of the UT Center for Perceptual Systems. His early research interest was in contrast processing in stereoscopic vision, which is a good model system in which to study how the brain combines signals in general. More recently, his laboratory has been investigating the image properties that attract gaze when viewing natural scenes and when searching for targets imbedded in naturalistic background noise.