

BLIND QUALITY ASSESSMENT OF JPEG2000 COMPRESSED IMAGES USING NATURAL SCENE STATISTICS

Hamid R. Sheikh, Alan C. Bovik and Lawrence Cormack

Laboratory for Image and Video Engineering, Department of Electrical and Computer Engineering,
The University of Texas at Austin, Austin, TX 78712-1084, USA.

Email: hamid.sheikh@ieee.org, bovik@ece.utexas.edu, cormack@psy.utexas.edu

ABSTRACT

Measurement of image quality is crucial for many image-processing algorithms, such as acquisition, compression, restoration, enhancement and reproduction. Traditionally, researchers in image quality assessment have focused on equating image quality with similarity to a ‘reference’ or ‘perfect’ image. The field of blind, or no-reference, quality assessment, in which image quality is predicted without the reference image, has been largely unexplored. In this paper, we present a blind quality assessment algorithm for images compressed by JPEG2000 using natural scene statistics (NSS) modelling. We show how 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.

1. INTRODUCTION

Digital images are now a part of our everyday lives, and the problem of automatically quantifying their quality has received tremendous attention in the research community. Most image quality assessment algorithms in the literature are *image fidelity metrics*, and assume that a ‘reference’ image is available against which a distorted or processed image can be compared. However, human observers can readily judge the quality of images without explicit knowledge of the reference images. We were thus motivated to consider *blind quality assessment*, also known as no-reference quality assessment, in which an algorithm assigns quality scores that are consistent with human perceptions of quality, but without any explicit comparisons with the reference image.

Blind Quality Assessment is a very hard problem since many unquantifiable factors play a role in human perceptions of quality, such as aesthetics, cognitive relevance, learning, context etc. In our previous work, we presented the following philosophy for circumventing these challenges of blind image quality assessment: *all images are perfect, regardless of content, until distorted by acquisition, processing or reproduction* [1]. Hence, the task of blind quality measurement simplifies into blindly measuring the distortion that has possibly been introduced into the image during the stages of acquisition, processing or reproduction. The reference for measuring this distortion would be the statistics of ‘perfect’ natural images.

Lossy image compression algorithms running at low bit rates are common sources of image distortions. JPEG2000 is a recent

image compression standard that is based on the Discrete Wavelet Transform. Like all lossy compression algorithms, it too suffers from artifacts at higher compression ratios, specifically blurring and ringing distortions [2]. In [1], we presented an algorithm for blindly measuring the quality of images compressed by JPEG2000 using NSS modelling. In this paper, we present enhancements to our algorithm to improve its prediction accuracy.

In our research, we make use of NSS modelling in the wavelet domain to quantify the loss in image quality due to the quantization of the wavelet coefficients. Specifically, we claim that the quantization process disturbs the non-linear dependencies that are present in natural images thereby making them ‘unnatural’, and that this perturbation can be quantified using statistical models for natural scenes, as well as by modelling the salient features of the quantization process in JPEG2000 (or any other Wavelet based) image codec. Using human subject data, this quantification can then be related to human perceptions of quality.

2. SUBJECTIVE EXPERIMENTS

A comprehensive psychometric study was conducted in order to train and test our algorithms. The study was conducted in collaboration with the Psychology Department at the University of Texas at Austin. A total of 198 images compressed by JPEG2000 to different bit rates were evaluated by about two dozen human subjects, who were asked to provide their perceptions of quality on a continuous linear scale. The average standard deviation of the processed subject scores over all images was found to be 7.04 (on a scale of 1-100). This quantifies the variability between human subjects, and can be used to gauge the performance of our blind metric. The database is available free of charge to the research community, along with details of the experiment [3].

3. STATISTICAL MODEL FOR IMAGES IN THE WAVELET DOMAIN

In [1], we asserted that the NSS model presented in [4, 5] is useful for blind quality assessment of JPEG2000 (or any other Wavelet based) image codec. The model captures the statistics of wavelet coefficients of natural images in a given subband and their non-linear correlations with other wavelet coefficients across scales and orientations. We observed from experiments that this model is suitable for blindly quantifying the perceptual effects of quantization of the wavelet coefficients of natural images, since quantization disturbs the non-linear dependencies between them.

H. R. Sheikh and A. C. Bovik are affiliated with the Laboratory for Image and Video Engineering, and L. Cormack is affiliated with the Center for Perceptual Systems, The University of Texas at Austin.

3.1. Natural Images

The statistical model proposed in [4, 5] 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:

$$C = M \times P + N \quad (1)$$

In [4, 5], an empirical distribution for M is used and N is assumed 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. Figure 1 shows the joint histograms of $(\log_2(P), \log_2(C))$ of an image at different scales and orientations, the strong non-linear dependence between P and C is clearly visible on the logarithmic axes.

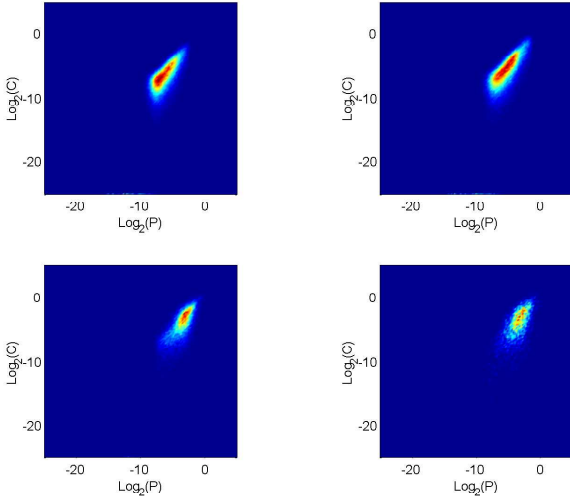


Fig. 1. Joint histograms of $(\log_2 P, \log_2 C)$ for an uncompressed natural image at different scales and orientations of its Wavelet decomposition. Top left: Diagonal subband at the finest scale. Top right: Horizontal subband at the finest scale. Bottom left: Vertical subband at the second-finest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

3.2. Compressed Natural Images

The model in (1) (Figure 1) is not very useful for compressed images, since quantization significantly affect the distributions. Figure 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 clearly visible: quantization pushes the coefficients towards zero, and disturbs the dependencies between P and C .

The blind quality assessment algorithm presented in [1] was based on a two-state statistical model for C only. In this paper, we will extend the method by using the joint distribution of P and C . We propose to use a discrete two-state model for P and C that can characterize uncompressed as well as compressed natural images. These two states correspond to the coefficient and its predictor being significant or insignificant. The two-state model is motivated by the fact that the quantization process in JPEG2000,

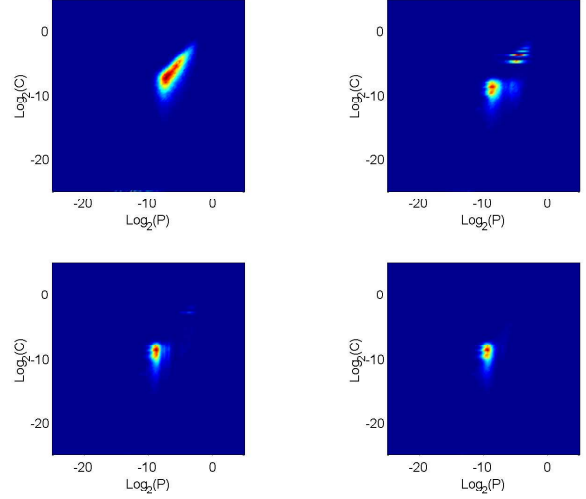


Fig. 2. Joint histograms of $(\log_2 P, \log_2 C)$ for one subband of an image when it is compressed at different bit rates using the JPEG2000 compression algorithm. Top left: No compression. Top right: 2.44 bits/pixel. Bottom left: 0.78 bits/pixel. Bottom right: 0.19 bits/pixel.

which occurs in all subbands, results in more of P and C values being insignificant than expected for uncompressed natural images. Hence, a good indicator of the unnaturalness and the perceptual effect of quantization is the percentage of significant P and C .

The details of the two-state 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 3.3.1. A coefficient (or its predictor) is considered to be significant if it is above threshold. Consequently, we obtain a set of four probabilities, $p_{ii}, p_{is}, p_{si}, p_{ss}$, corresponding to the probabilities that the predictor/coefficient pair lies in one of the four quadrants, as depicted in Figure 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 Figure 3.

3.3. Features for Blind Quality Assessment

We observed from experiments that the feature vector consisting of probabilities p_{ss} from a number of subbands (specifically, the horizontal, vertical and diagonal orientations at the two finest resolutions) gives the best indication of the loss of quality, in terms of minimizing the quality prediction error over the database. We also observed that the feature p_{ss} from each subband is related to image quality by a saturating exponential fit as follows:

$$q_i = K_i \left(1 - \exp \left(- \frac{(p_{ss,i} - u_i)}{T_i} \right) \right) \quad (2)$$

where q_i is the quality prediction for the image using the i -th subband, $p_{ss,i}$ is the p_{ss} probability for the i -th subband, and K_i, T_i and u_i are curve fitting parameters for the i -th subband. In our simulations, $i = 1 \dots 6$ denote the following six subbands: horizontal, vertical and diagonal orientations at the second-finest resolution; horizontal, vertical and diagonal orientations at

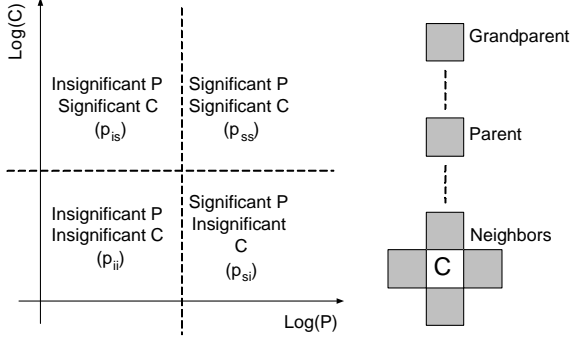


Fig. 3. Partition of the (P, C) space into quadrants. Also, the set of coefficients from which we calculated P .

the finest resolution. The six-dimensional subband quality vector, $q = \{q_i | i \in 1 \dots 6\}$, is modified into a four-dimensional vector q' 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' :

$$\begin{bmatrix} q'_1 \\ q'_2 \\ q'_3 \\ q'_4 \end{bmatrix} = \begin{bmatrix} (q_1 + q_2)/2 \\ q_3 \\ (q_4 + q_5)/2 \\ q_6 \end{bmatrix}$$

$$Q = q'^T w \quad (3)$$

Since the computation of coefficient predictions P is expensive, we also tried to improve the performance of our blind quality assessment algorithm presented in [1] by incorporating image dependent thresholds in it. The parameters K_i , T_i , u_i and w are learned from the training data.

3.3.1. Image-dependent threshold calculations

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 Figure 4 for a number of uncompressed natural images. The graphs have approximately the same slope, while the intercept on the y-axis vary 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 $\frac{1}{f}$, 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 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.

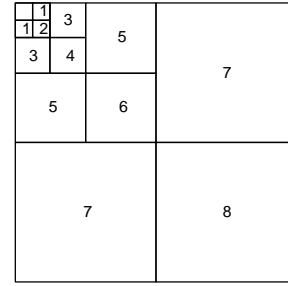
Quantized images are not natural however, and hence the corresponding plots for them will not have approximately linear fall-off (Figure 5, solid lines). However, we know that the quantization process in JPEG2000 is designed in such a way that the subband means for coarser subbands are less affected 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 is derived. This is shown in Figure 5 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).

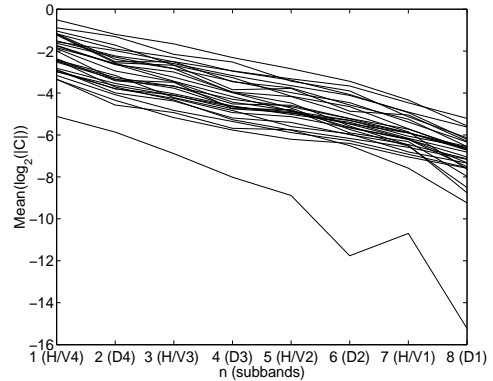
We use the above observation to calculate the image dependent thresholds as follows:

$$\text{Threshold} = \text{Estimated Mean} + \text{Offset} \quad (4)$$

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 a minimization process that attempts to minimize the error in the quality predictions over the training set (using MATLAB's command *fminsearch*). In this way, our algorithm utilizes NSS models in concert with modelling the salient features of the distortion process to make accurate predictions about quality.



(a) Subband enumeration.



(b) $\text{Mean}(\log_2(C))$ versus subband enumeration index for uncompressed images.

Fig. 4. $\text{Mean}(\log_2(C))$ versus subband enumeration index for uncompressed natural images. The means of horizontal and vertical subbands at a given scale are averaged.

4. RESULTS

For training and testing, the database is divided into two parts. The training database consists of fifteen randomly selected images

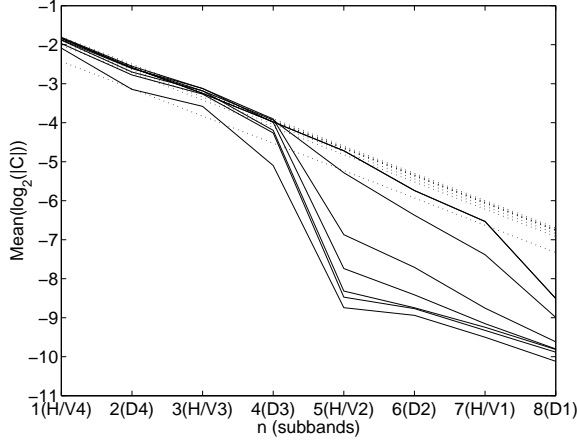
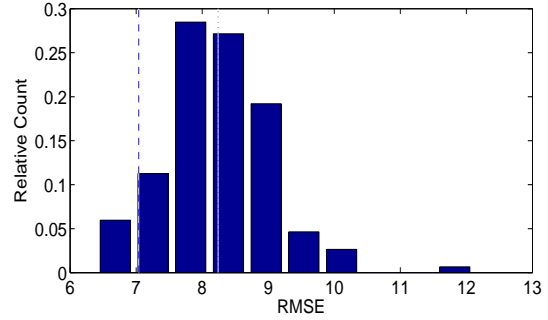


Fig. 5. $\text{Mean}(\log_2(C))$ versus subband enumeration index for an image at different bit rates. Prediction for the means of subbands $n = 5 \dots 8$ are made by estimating the line that describes the fall-off (dotted). This line is estimated by assuming a constant slope for all images, and using subbands $n = 1 \dots 4$ for estimating the intercept. Note that the predicted lines for the image at different bit rates (dotted) is quite close to the actual graph for the uncompressed image (top solid line).

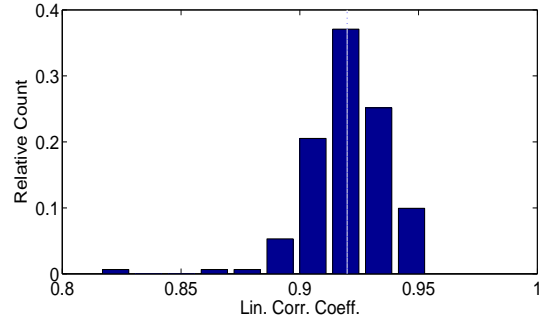
(from the total 29) and all of their distorted versions. The testing database consists of the other fourteen images and their distorted versions. This way there is no overlap between the training and the testing databases. The algorithm is run several times, each time with a different (and random) subset of the original 29 images for training and testing, with 15 and 14 images (and their distorted versions) in the training and the 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 (4) is learned from the uncompressed images in the training set. The weights w in (3) are learned using non-negatively constrained least-squares fit over the training data (MATLAB command *lsqnonneg*). The minimization over the threshold offsets in (4), as well as for the fitting parameters is done by unconstrained non-linear minimization (MATLAB command *fminsearch*).

Figure 6(a) shows the normalized histogram of the Root Mean Squared Error (RMSE) (which we use as a measure of the performance of our algorithm) between the quality prediction and the subjective Mean Opinion Score (MOS), for a number of runs of the algorithm. The mean RMSE is 8.24, with a standard deviation of 0.83. Figure 6(b) shows the normalized histogram of the linear correlation coefficient, with a mean linear correlation coefficient of 0.92. The RMSE should be compared against a mean standard deviation of 7.04 for human assignments, and against an RMSE of 26.9 for uniform random predictions over the range of MOS scores. Figure 7(a) shows the RMSE histogram obtained using the binarized marginal model for C . The mean RMSE is 9.00 and the standard deviation is 0.84, with a mean linear correlation coefficient of 0.91.



(a) RMSE histogram for Joint Model



(b) CC histogram for Joint Model

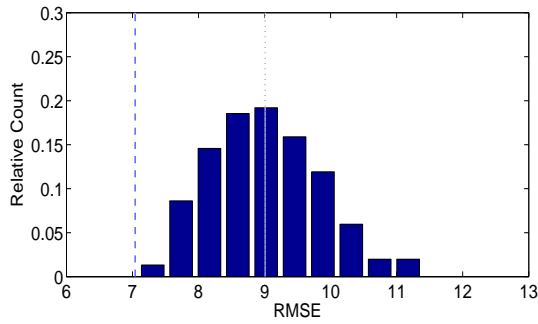
Fig. 6. Results: Normalized histograms of the RMSE and linear correlation coefficient for several runs of the algorithm using the joint statistics of P and C . The mean performance is shown as well (dotted). The standard deviation among human judgements is shown on the RMSE histogram (dashed).

5. DISCUSSION

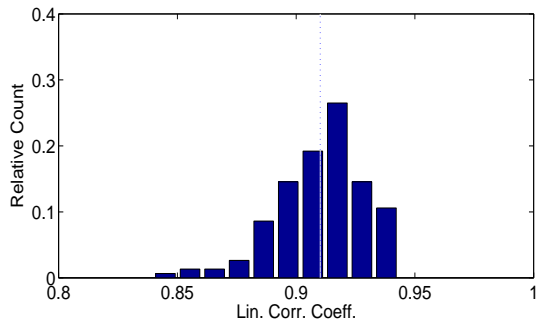
It is apparent from the above figures that our algorithm is able to make predictions of the quality of images compressed with JPEG2000 that are consistent with human evaluations. The average ‘error’ in quality assignment for a human subject is 7.04, while for our algorithm it is 8.24 (in [1], we reported a mean RMSE performance of 9.9). We are therefore performing close to the limit imposed on useful prediction by the variability between human subjects. The average gap between an average human and the algorithm is 1.2 on the a scale of 1 - 100. Another interesting figure is the standard deviation of the RMSE of 0.83 on a scale of 1 - 100 (in [1] we reported a standard deviation of 1.4), which indicates that our algorithm’s performance is stable to changes in the training database.

6. CONCLUSIONS

In this paper we have presented an algorithm for blindly determining the quality of images that have been compressed by JPEG2000. The algorithm utilizes NSS modelling in concert with modelling the salient features of the distortion process to blindly measure the distortion present in an image. The algorithm is trained and tested



(a) RMSE histogram for Marginal Model



(b) CC histogram for Marginal Model

Fig. 7. Results: Normalized histograms of the RMSE and linear correlation coefficient for several runs of the algorithm using the marginal statistics of C only. The mean performance is shown as well (dotted). The standard deviation among human judgements is shown on the RMSE histogram (dashed).

on data obtained from human observers. On a scale of 1-100, an average RMSE of approximately 8.24 between the quality predictions and human evaluations is reported, which is close to the average standard deviation of 7.04 for the quality scores assigned by human observers. We are continuing research into using higher-order models of natural image statistics in the wavelet domain to achieve better performance.

7. REFERENCES

- [1] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "Blind quality assessment for JPEG2000 compressed images," in *Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers*, Nov. 2002.
- [2] D. S. Taubman and M. W. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards, and Practice*. Kluwer Academic Publishers, 2001.
- [3] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "LIVE image quality assessment database." <http://live.ece.utexas.edu/research/quality>.
- [4] E. P. Simoncelli, "Statistical models for images: Compression,

restoration and synthesis," in *Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers*, Nov. 1997.

- [5] R. W. Buccigrossi and E. P. Simoncelli, "Image compression via joint statistical characterization in the wavelet domain," *IEEE Trans. Image Processing*, vol. 8, pp. 1688–1701, Dec. 1999.