NO-REFERENCE PERCEPTUAL QUALITY ASSESSMENT OF JPEG COMPRESSED IMAGES

Zhou Wang, Hamid R. Sheikh and Alan C. Bovik

Laboratory for Image and Video Engineering (LIVE), Dept. of Electrical and Computer Engineering,
The University of Texas at Austin, Austin, TX 78712-1084, USA
Email: zhouwang@ieee.org, {sheikh, bovik}@ece.utexas.edu

ABSTRACT

Human observers can easily assess the quality of a distorted image without examining the original image as a reference. By contrast, designing objective No-Reference (NR) quality measurement algorithms is a very difficult task. Currently, NR quality assessment is feasible only when prior knowledge about the types of image distortion is available.

This research aims to develop NR quality measurement algorithms for JPEG compressed images. First, we established a JPEG image database and subjective experiments were conducted on the database. We show that Peak Signal-to-Noise Ratio (PSNR), which requires the reference images, is a poor indicator of subjective quality. Therefore, tuning an NR measurement model towards PSNR is not an appropriate approach in designing NR quality metrics. Furthermore, we propose a computational and memory efficient NR quality assessment model for JPEG images. Subjective test results are used to train the model, which achieves good quality prediction performance. A Matlab implementation of the proposed method is available at *http://anchovy.ece.utexas.edu/~zwang/research/nr_jpeg_quality/index.html* .

1. INTRODUCTION

In recent years, there has been an increasing need to develop objective measurement techniques that can predict image /video quality automatically. Such methods can have various applications. First, they can be used to *monitor* image/video quality for quality control systems. Second, they can be employed to *benchmark* image/video processing systems and algorithms. Third, they can also be embedded into image/video processing systems to *optimize* algorithms and parameter settings. The most widely used objective image quality/distortion metrics are Peak Signal-to-Nose Ratio (PSNR) and Mean Squared Error (MSE), but they are widely criticized as well for not correlating well with perceived quality measurement. In the past three to four decades, a great deal of effort has been made to develop new objective image/video quality metrics that incorporate perceptual quality measures by considering Human Visual System (HVS) characteristics [1]–[4].

Most of the proposed image quality assessment approaches require the original image as a reference. Interestingly, human observers can easily assess the quality of distorted images without using any reference image. By contrast, designing objective No-Reference (NR) quality measurement algorithms is a very difficult task. This is mainly due to the limited understanding of the HVS, and it is believed that effective NR quality assessment is feasible only when the prior knowledge about the image distortion types is available. Although only a limited number of methods have been proposed in the literature [5]–[9] for objective NR quality assessment, this topic has attracted a great deal of attention recently. For example, the Video Quality Experts Group (VQEG, *http://www.vqeg.org*) considers the standardization of NR and Reduced-Reference (RR) video quality assessment methods as one of its future working directions, where the major source of distortion under consideration is block DCT-based video compression.

The purpose of this research is to develop objective NR quality assessment algorithms for JPEG compressed images. Such algorithms must have the capability to effectively predict perceived JPEG image quality. We consider blurring and blocking as the most significant artifacts generated during the JPEG compression process. An efficient way is proposed to extract features that can be used to reflect the relative magnitudes of these artifacts. The extracted features are combined to constitute a quality prediction model. Subjective experimental results on JPEG compressed images are used to train the model, which achieves very good quality prediction performance.

2. SUBJECTIVE EXPERIMENTS

The subjective test was conducted on 8 bits/pixel gray level images. There are 120 test images in the database. Thirty of them are original images, which are randomly divided into

This research was supported in part by Texas Instruments, Inc., and by the State of Texas Advanced Technology Program (ATP).

Fig. 1. Group I images.

Fig. 2. Group II images.

two groups with 15 images in each group. The two groups of images are shown in Figs. 1 and 2, respectively. The rest of the test images are JPEG-compressed using the “imwrite” routine in the Matlab image processing toolbox. The quality factors are selected randomly between 5 and 100, and the resulting bit rates range from 0.2 to 1.7 bits/pixel. Fifty-three subjects were shown the database; most of them were college students. The subjects were asked to assign each image a quality score between 1 and 10 (10 represents the best quality and 1 the worst). The 53 scores of each image were averaged to a final Mean Opinion Score (MOS) of the image.

Although the purpose of this research is to develop NR objective image quality assessment methods and the calculation of PSNR requires the reference images, it is interesting to see how the PSNR measurements correlate with the MOS values because PSNR is widely used in various image processing applications, and has been employed as a reference model to evaluate the effectiveness of other objective image/video quality assessment approaches [1], [9]. The PSNR results versus MOSs of the JPEG-compressed test images are shown in Fig. 3, where each sample point represents one test image. It can be observed that PSNR performs poorly in predicting subjective image quality. This is reflected by the correlation coefficient between PSNR and MOS, which is only 0.3267. Although it is often believed that PSNR is an acceptable quality measure for high-quality (high bit rate) compressed images, its visual quality prediction ability degrades significantly when applied to images with a wide range of compression ratios as in our current test. Therefore, tuning an NR measurement model towards PSNR as in [9] is not an appropriate approach in designing NR quality metrics.

3. OBJECTIVE NR QUALITY ASSESSMENT

JPEG is a block DCT-based lossy image coding technique. It is lossy because of the quantization operation applied to the DCT coefficients in each 8×8 coding block. Both blurring and blocking artifacts may be created during quantization. The blurring effect is mainly due to the loss of high frequency DCT coefficients, which smoothes the image signal within each block. Blocking effect occurs due to the discontinuity at block boundaries, which is generated because the quantization in JPEG is block-based and the blocks are quantized independently.

One effective way to examine both the blurring and blocking effects is to transform the signal into the frequency domain [6]. We denote the test image signal as $x(m, n)$ for $m \in [1, M]$ and $n \in [1, N]$, and calculate a differencing signal along each horizontal line:

$$d_h(m, n) = x(m, n+1) - x(m, n)\,, \ n \in [1, N-1]. \quad (1)$$

Let $f_m(n) = |d_h(m, n)|$ be a 1-D horizontal signal for a fixed value of m. If we compute the power spectrum of $f_m(n)$ for $m = 1, \cdots, M$, and average them together, then we obtain a power spectrum estimation $P_h(l)$ exemplified in Fig. 4, where the blocking effect can be easily identified by the peaks at the feature frequencies (1/8, 2/8, 3/8, and 4/8) and the blurring effect is also characterized by the energy shifting from high frequency to low frequency bands. A disadvantage of the frequency domain method is the involvement of the Fast Fourier Transform (FFT) [6], which has to be calculated many times for each image, and is therefore expensive. FFT also requires more storage space because it cannot be computed locally.

In this paper, we attempt to design a computationally inexpensive and memory efficient feature extraction method.

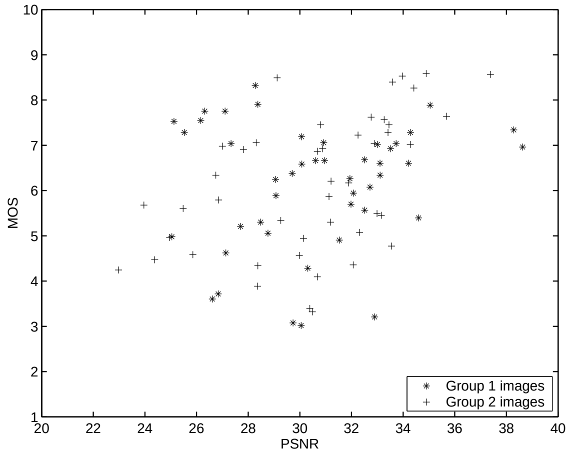

Fig. 3. PSNR results compared with MOS.

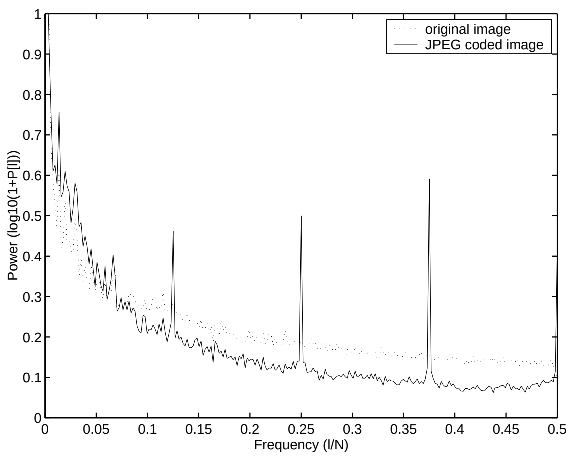

Fig. 4. Power spectrum comparison of the original and JPEG compressed "Lena" images.

The features are calculated horizontally and then vertically. First, the blockiness is estimated as the average differences across block boundaries:

$$B_h = \frac{1}{M(\lfloor N/8 \rfloor - 1)} \sum_{i=1}^{M} \sum_{j=1}^{\lfloor N/8 \rfloor -1} |d_h(i, 8j)| \qquad (2)$$

Second, we estimate the activity of the image signal. Although blurring is difficult to be evaluated without the reference image, it causes the reduction of signal activity, and combining the blockiness and activity measures gives more insight into the relative blur in the image. The activity is measured using two factors. The first is the average absolute difference between in-block image samples:

$$A_h = \frac{1}{7} \left[\frac{8}{M(N-1)} \sum_{i=1}^{M} \sum_{j=1}^{N-1} |d_h(i,j)| - B_h \right] \qquad (3)$$

Table 1. RMS between MOS and model prediction

Training images	Testing Images		
	Group 1	Group 2	All
Group 1	0.7756	0.7627	0.7692
Group 2	0.8947	0.5894	0.7576
All	0.8113	0.6283	0.7256

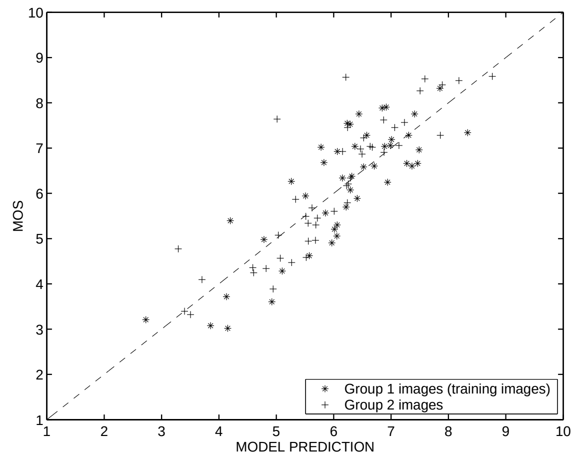

Fig. 5. Model prediction results using Group I images as the training images.

The second activity measure is the zero-crossing (ZC) rate. We define for $n \in [1, N-2]$,

$$z_h(m,n) = \begin{cases} 1 & horizontal\ ZC\ at\ d_h(m,n) \\ 0 & otherwise \end{cases} \qquad (4)$$

The horizontal ZC rate then can be estimated as:

$$Z_h = \frac{1}{M(N-2)} \sum_{i=1}^{M} \sum_{j=1}^{N-2} z_h(m,n) \qquad (5)$$

Using similar methods, we calculate the the vertical features of B_v, A_v, and Z_v. Finally, the overall features are given by:

$$B = \frac{B_h + B_v}{2}, \ A = \frac{A_h + A_v}{2}, \ Z = \frac{Z_h + Z_v}{2}. \qquad (6)$$

There are many different ways to combine the features to constitute a quality assessment model. One method we find that gives good prediction performance is given by

$$S = \alpha + \beta B^{\gamma_1} A^{\gamma_2} Z^{\gamma_3}, \qquad (7)$$

where α, β, γ_1, γ_2, and γ_3 are the model parameters that must be estimated with the subjective test data. The nonlinear regression routine "nlinfit" in the Matlab statistics toolbox is used to find the best parameters for (7). It is important that the model is not overtrained, in which case,

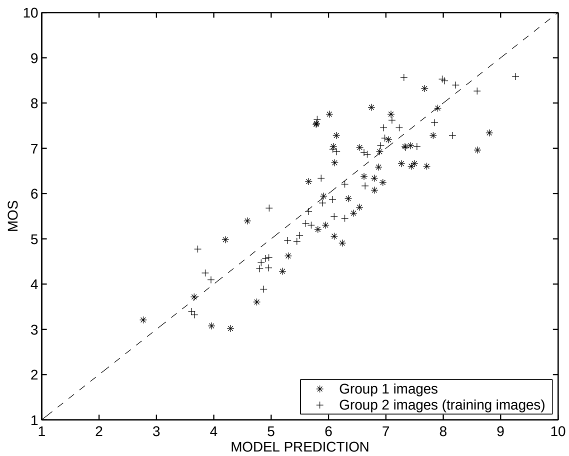

Fig. 6. Model prediction results using Group II images as the training images.

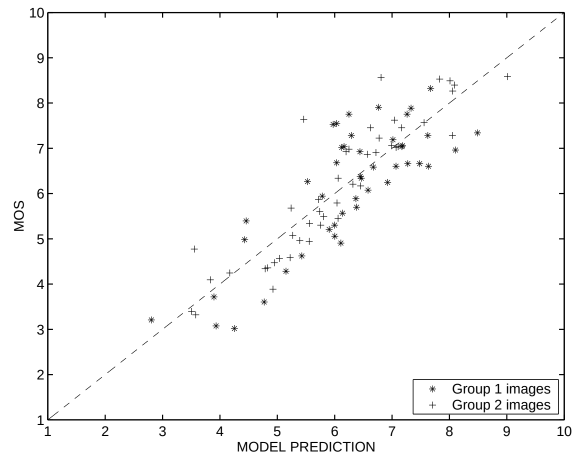

Fig. 7. Model prediction results using both groups of images as the training images.

although very good fitting is obtained for the training data, the model's generalization ability is affected. Therefore, we use different groups of images as the training images. The results shown in Figs. 5, 6 and 7 are obtained using Group I (Fig. 1), Group II (Fig. 2), and both groups of images as the training images, respectively. The model performs well in all three tests, which implies that the model is robust. This is confirmed by Table 1, where the Root Mean Squared error (RMS) between the model prediction score and MOS is given. The parameters obtained with all test images are $\alpha = -245.9$, $\beta = 261.9$, $\gamma_1 = -0.0240$, $\gamma_2 = 0.0160$, and $\gamma_3 = 0.0064$, respectively.

4. CONCLUSIONS

We demonstrate a novel NR perceptual quality assessment scheme for JPEG compressed images. Subjective experiments were conducted to evaluate the quality of JPEG compressed images. The features described in the paper effectively capture the artifacts introduced by JPEG, and the nonlinear fitting gives good agreement with MOS scores.

The method is computationally efficient since no complicated transforms are computed and the algorithm can be implemented without storing the entire image (or even a row of pixels) in memory, which makes embedded implementations easier. The basic methodology of the proposed method can also been used to develop NR quality assessment methods for H.26x/MPEG compressed video.

A Matlab implementation of the proposed method is available at *http://anchovy.ece.utexas.edu/˜zwang/research/nr_jpeg_quality/index.html* . We are also continuing our subjective experiments with more test images, subjects and types of image distortions, and will make the test database available to the public in the near future.

5. REFERENCES

[1] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," *http://www.vqeg.org/*, Mar. 2000.

[2] T. N. Pappas and R. J. Safranek, "Perceptual criteria for image quality evaluation," in *Handbook of Image & Video Proc.* (A. Bovik, ed.), Academic Press, 2000.

[3] C. J. van den Branden Lambrecht, Ed., "Special issue on image and video quality metrics," *Signal Processing*, vol. 70, Nov. 1998.

[4] Z. Wang and A. C. Bovik, "Why is image quality assessment so difficult?," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, May 2002.

[5] H. R. Wu and M. Yuen, "A generalized block-edge impairment metric for video coding," *IEEE Signal Processing Letters*, vol. 4, pp. 317–320, Nov. 1997.

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

[7] 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, pp. 1725 –1728, May 2001.

[8] P. Gastaldo, S. Rovetta, and R. Zunino, "Objective assessment of MPEG-video quality: a neural-network approach," in *Proc. IJCNN*, vol. 2, pp. 1432 –1437, 2001.

[9] M. Knee, "A robust, efficient and accurate signal-ended picture quality measure for MPEG-2." available at *http://www-ext.crc.ca/vqeg/frames.html*, 2001.