JPEG vs. JPEG2000:
An Objective Comparison of Image Encoding Quality

Farzad Ebrahimi, Matthieu Chamik, Stefan Winkler

Genista Corporation
Rue du Théâtre 5, 1820 Montreux, Switzerland

ABSTRACT

This paper describes an objective comparison of the image quality of different encoders. Our approach is based on estimating the visual impact of compression artifacts on perceived quality. We present a tool that measures these artifacts in an image and uses them to compute a prediction of the Mean Opinion Score (MOS) obtained in subjective experiments. We show that the MOS predictions by our proposed tool are a better indicator of perceived image quality than PSNR, especially for highly compressed images.

For the encoder comparison, we compress a set of 29 test images with two JPEG encoders (Adobe Photoshop and IrfanView) and three JPEG2000 encoders (JasPer, Kakadu, and IrfanView) at various compression ratios. We compute blockiness, blur, and MOS predictions as well as PSNR of the compressed images. Our results show that the IrfanView JPEG encoder produces consistently better images than the Adobe Photoshop JPEG encoder at the same data rate. The differences between the JPEG2000 encoders in our test are less pronounced; JasPer comes out as the best codec, closely followed by IrfanView and Kakadu. Comparing the JPEG- and JPEG2000-encoding quality of IrfanView, we find that JPEG has a slight edge at low compression ratios, while JPEG2000 is the clear winner at medium and high compression ratios.

Keywords: Image compression, codec comparison, perceptual quality metrics, blockiness, blur.

1. INTRODUCTION

Advances in computer and communication technologies have led to a proliferation of digital media content. However, digital images and videos are still demanding in terms of storage space and transmission bandwidth. Lossy compression is necessary to bring these demands down to a manageable level, but it introduces various types of artifacts, such as blockiness, blur, ringing, noise etc. In order to optimize imaging systems and to improve the perceptual quality of the delivered content, metrics are needed that identify these artifacts and measure their perceptual impact.

Well-known metrics like the Peak Signal-to-Noise Ratio (PSNR) or bit error rate tell us how well a signal is reproduced with respect to a reference signal. They are pure signal fidelity metrics and do not measure how impairments are perceived by viewers. In contrast to these, perceptual metrics take the human visual system (HVS) into account and can provide accurate estimates of the attributes and the extent of visible distortions.

In this paper, we present a practical approach for measuring image quality in the framework of encoder comparison, specifically with JPEG and JPEG2000 codecs. We use two publicly available JPEG2000 encoders, JasPer and Kakadu, as well as the encoders of two well-known image manipulation programs, Adobe Photoshop and IrfanView, to compress a set of test images. We then measure the perceptual impairments in these images and use the results for a comparative analysis of encoding quality. A comparison of features, complexity and robustness of JPEG and JPEG2000 can be found elsewhere.[1]

The paper is organized as follows. Section 2 introduces the perceptual metrics that we use to characterize the quality of images. Section 3 describes our experimental setup. In Section 4, we present and discuss our results, comparing the JPEG encoders, the JPEG2000 encoders, and JPEG vs. JPEG2000. Finally, Section 5 concludes the paper.

E-mail of corresponding author: stefan.winkler@genista.com

2. ARTIFACTS AND METRICS

It is useful to define perceptual metrics for the numerical measurement of the visible damage caused by image compression. We describe two artifact metrics for blockiness and blur as well as an objective prediction of overall image quality that is based on these artifact metrics.

2.1. Blockiness

Blockiness is a perceptual measure of the block structure that is common to all block-DCT based image and video compression techniques, as for example JPEG.[2] The DCT is typically performed on 8×8 pixel blocks in the frame; the coefficients in each block are quantized separately. This leads to artificial horizontal and vertical borders between these blocks, which can be detected.[3] Blockiness can also be caused by transmission errors, which often affect entire slices of blocks in an image.

2.2. Blur

Blur is a perceptual measure of the loss of fine detail and the smearing of edges.[4] It is due to the attenuation of high frequencies at some stage of the recording or encoding process. It is one of the main artifacts of wavelet-based compression techniques such as JPEG2000,[5] for which transmission errors or packet loss can also induce blur. DCT-based compression schemes exhibit blur too, even if it is not the primary distortion. Other important sources of blur are low-pass filtering (e.g. analog VHS tape recording) or out-of-focus shots.

Figure 1 shows a portion of one of the images used in our tests encoded with JPEG and JPEG2000 at a compression ratio of 100:1. Blockiness and blur are visible in the JPEG- and JPEG2000-encoded images, respectively.

(a) Original image   (b) JPEG-encoded image   (c) JPEG2000-encoded image

Figure 1. The original image and the JPEG- and JPEG2000-encoded images at a compression ratio of 100:1. The respective artifacts of blockiness and blur are visible in the compressed images.

2.3. Overall Quality

The overall quality of an image is characterized by the Mean Opinion Score (MOS) obtained from experiments with human subjects. MOS is the average quality rating over a number of human observers that have been asked to score an image, often on the scale from 1 (worst) to 5 (best). An objective MOS prediction is a metric that correlates with human perception of image quality and thus with the output of subjective test results. Our MOS prediction uses perceptual metrics for blockiness and blur in combination with a few others to estimate the perceived quality of an image.

2.4. Image Quality Measurement

We use Genista's Image Optimacy software[6] to compute perceptual metrics for each encoded image. Image Optimacy is a software tool to objectively measure image quality, incorporating ANSI metrics,[7] fidelity metrics and artifact metrics based on human visual perception (including the ones discussed above). It thus enables users to objectively evaluate the perceptual impact of image artifacts introduced by compression and image processing. It has two modes, one for full-reference (FR) analysis and one for no-reference (NR) analysis. In full-reference mode, the relative degradation of a processed image with respect to a reference (typically the original image) is measured. In no-reference mode, the reference image is not needed; instead, an absolute quality rating is computed for a given image. This mode is of interest mainly when the reference is not available. Since the focus of this paper is encoder comparison, the full-reference metrics are used here.

To verify the MOS predictions made by Image Optimacy, we used the subjective MOS data from the LIVE Image Quality database.[8] This database comprises 175 JPEG-encoded images and 169 JPEG2000-encoded images, with compression ratios ranging from 7.2:1 to 160:1 for JPEG, and from 7.6:1 to 860:1 for JPEG2000. We computed a mapping of the artifact metrics of Image Optimacy to the subjective MOS data for these images. The mapping is based on a multivariate nonlinear regression model. Tuning was performed on a randomly selected half of the data, and the other half was used for evaluation. Figure 2 shows the MOS prediction performance of the full-reference metrics for all 344 compressed images in the LIVE database.
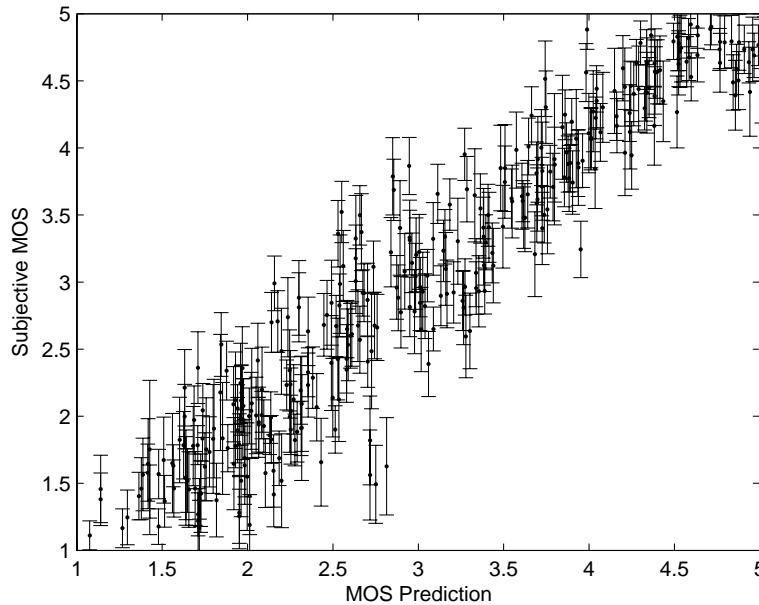
Figure 2. Subjective MOS versus full-reference MOS predictions by Image Optimacy. The error bars indicate the 95% confidence intervals of the subjective ratings.

The accuracy of the MOS prediction is characterized by a correlation of 95%, which is higher than what can be achieved with PSNR. There are very few outliers, and the RMS prediction error is 0.26 MOS units on the 1-5 scale, which is practically the same as the average confidence interval size of the subjective ratings (0.25 MOS units). Image Optimacy achieves almost the same MOS prediction accuracy with no-reference metrics.[9]

While PSNR also does quite well on this particular database, its problems are amplified when we look at the lower-quality range of the test images (see Figure 3). There PSNR performance decreases dramatically, whereas MOS prediction performance with our metrics remains high. This confirms that PSNR is not a good image quality indicator for high compression ratios.
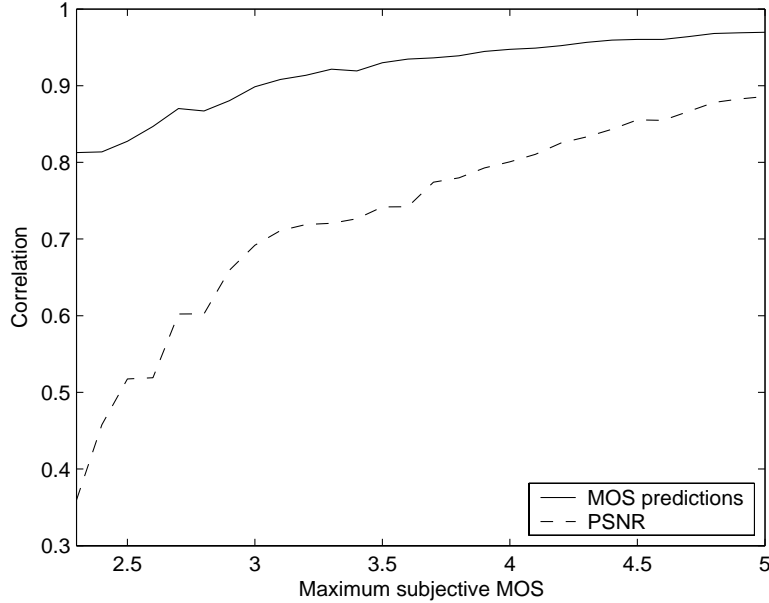
Figure 3. Correlation of PSNR and Image Optimacy MOS predictions as a function of maximum subjective image quality. For the images with a MOS below 3, for example, Image Optimacy achieves a correlation of 0.9 with MOS, compared to only 0.69 for PSNR.

3. EXPERIMENTAL SETUP

29 full color images from the LIVE Image Quality database[8] were used as originals. The resolution of most images is 768×512 pixels (some images are slightly smaller).

For JPEG compression, we used the encoders that come with Adobe Photoshop[10] version 7.0 and IrfanView[11] version 3.91. For our comparisons, we looked at compression ratios between 2:1 and 100:1.

For JPEG2000 compression, we used JasPer[12] version 1.700.0 and Kakadu[13] version 4.2 as well as the JPEG2000 encoder of IrfanView 3.91. Due to the better performance of JPEG2000 at low bitrates, we extended the range of compression ratios up to 300:1.

Genista's Image Optimacy software (see Section 2.4) was then used to evaluate the quality of the test images.

4. RESULTS AND DISCUSSION

We divide our analysis into three parts. First, we compare the quality of the JPEG encoders of IrfanView and Adobe Photoshop. Then we compare the quality of the JPEG2000 encoders of JasPer, Kakadu and IrfanView. Finally, we perform a quality comparison between JPEG and JPEG2000 using the respective encoders provided with IrfanView. For the analysis, the MOS predictions and PSNR values are averaged over all 29 source images used in the test.
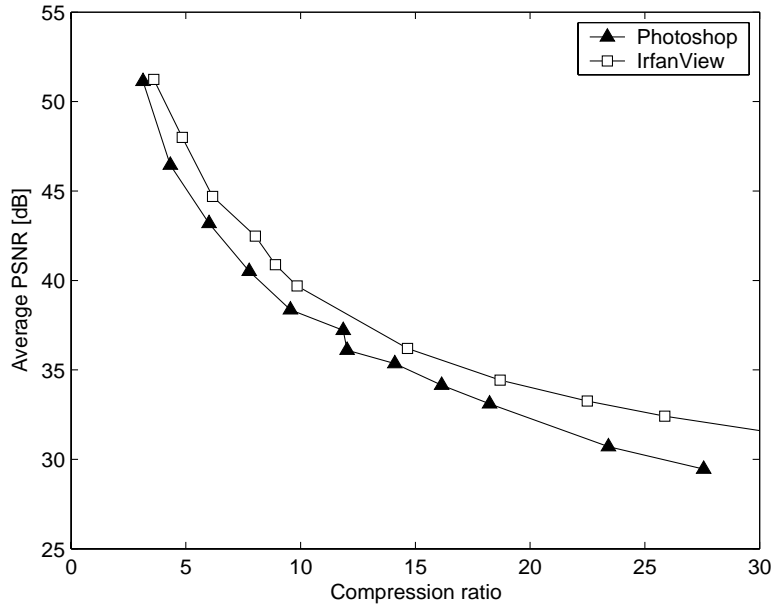
4.1. JPEG Encoder Comparison

Figures 4 and 5 show the average MOS predictions and the average PSNR as a function of compression ratio for encoding with Adobe Photoshop and IrfanView. Using either metric, IrfanView comes out as the superior JPEG encoder compared to Photoshop.

When using Adobe Photoshop for JPEG compression we observed a strange behavior. The compression is chosen by means of a quality factor that ranges from 0 (lowest quality) to 12 (highest quality). Surprisingly, a higher quality factor does not produce better images in all cases. Figure 6 shows the MOS predictions and PSNR values as a function of Photoshop's JPEG quality factor. Indeed a quality factor of 7 produces worse images (both in terms of MOS and PSNR) than a quality factor of 6, while the compression ratio remains almost the
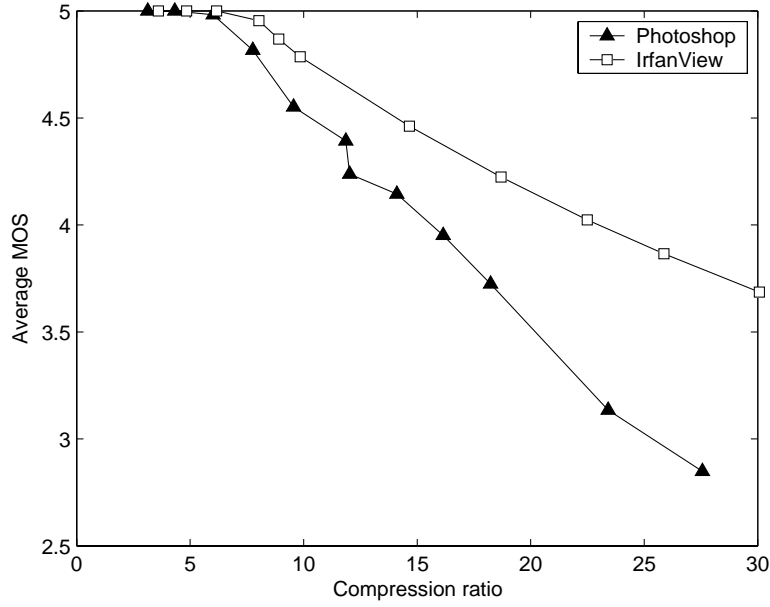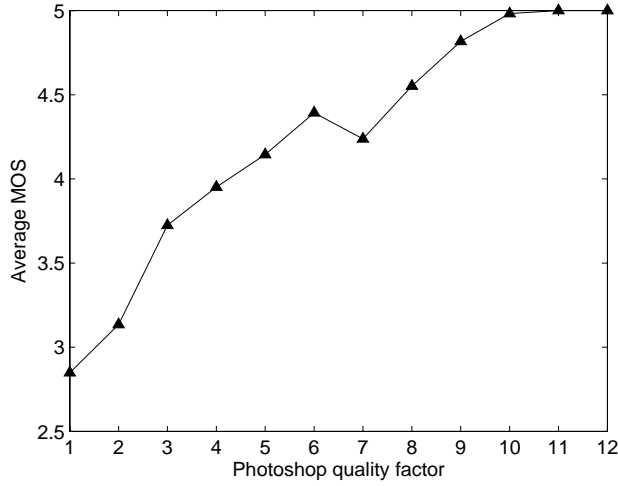
Figure 4. Average MOS predictions as a function of compression ratio for the JPEG-encoders of Adobe Photoshop and IrfanView. The JPEG codec of IrfanView clearly produces better images.


Figure 5. Average PSNR as a function of compression ratio for the JPEG-encoders of Adobe Photoshop and IrfanView.

same (cf. Figures 4 and 5). This behavior can be observed for all images and different versions of Photoshop. These results suggest that there is an inconsistency in the quality factor parameter. It is thus best to avoid a quality factor of 7 in Photoshop, because a quality factor of 6 still achieves almost the same compression at a higher quality.
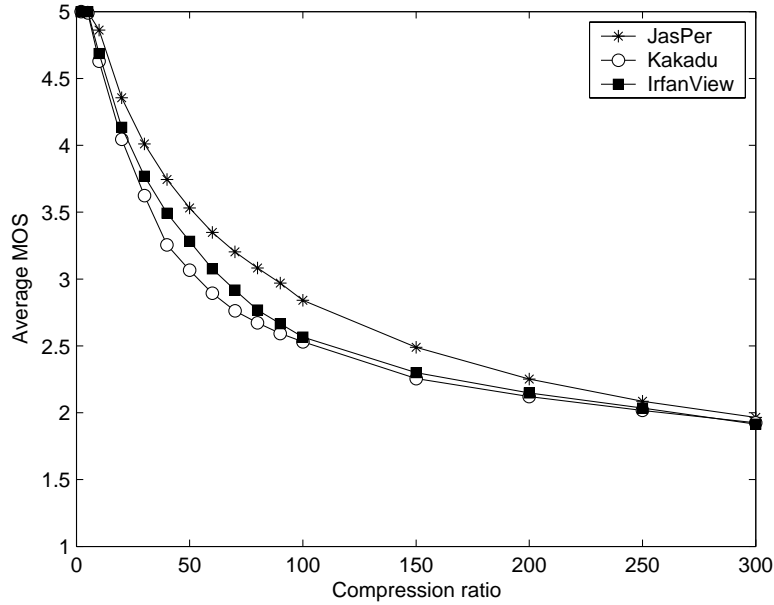
(a) MOS predictions
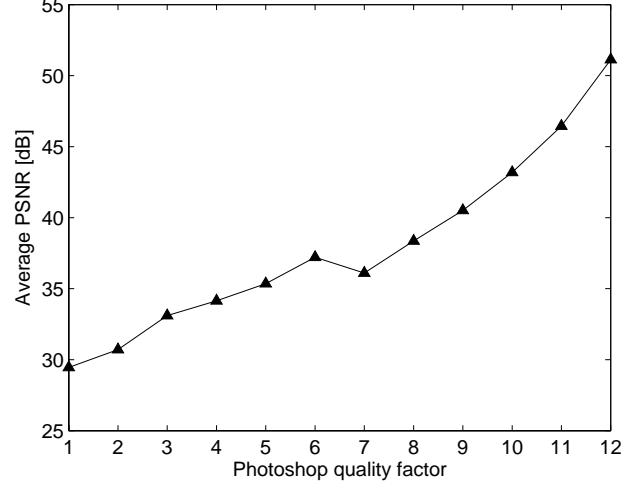
(b) PSNR

Figure 6. Average MOS predictions and PSNR as a function of the quality factor of the Adobe Photoshop JPEG encoder. Note the counterintuitive decrease in quality from factor 6 to 7.

4.2. JPEG2000 Encoder Comparison

We compare the three JPEG2000 encoders JasPer, Kakadu and IrfanView. Figures 7 and 8 show the average MOS predictions and the average PSNR of these encoders as a function of compression ratio. The metrics indicate that JasPer is the JPEG2000 encoder producing the best quality, whereas Kakadu comes out last. However, the differences between the three codecs are less significant than for JPEG.
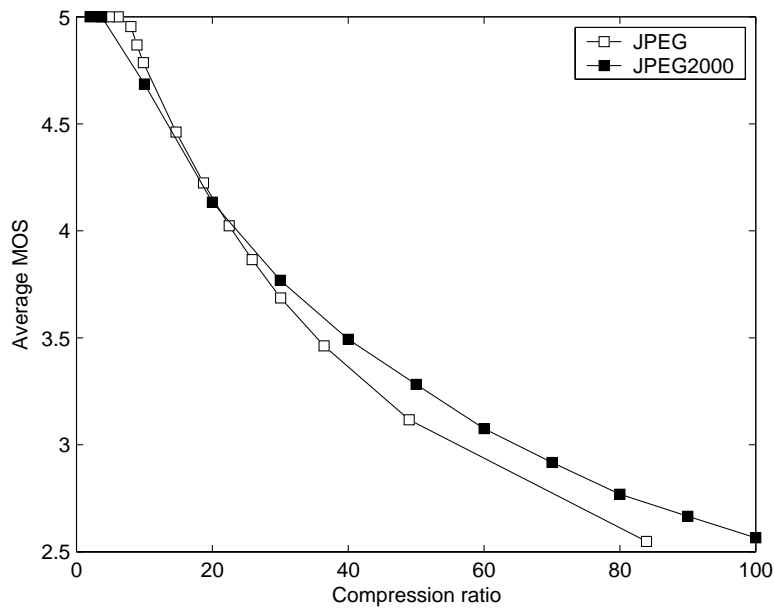

Figure 7. Average MOS predictions as a function of compression ratio for JPEG2000 encoders JasPer, Kakadu, and IrfanView.
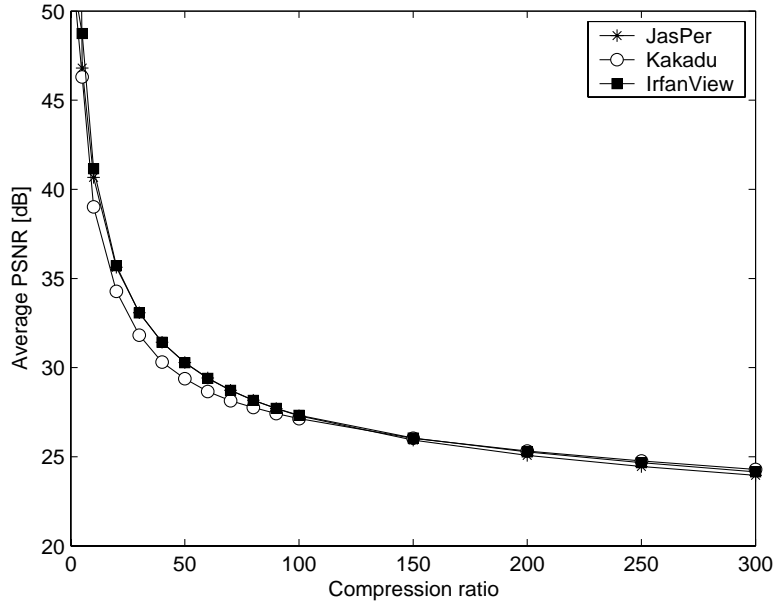
Figure 8. Average PSNR as a function of compression ratio for JPEG2000 encoders JasPer, Kakadu, and IrfanView.

4.3. JPEG vs. JPEG2000

Figure 9 shows the average MOS of IrfanView's JPEG and JPEG2000 encoders as a function of compression ratio. For low compression ratios (below 20:1) JPEG produces slightly better images (mainly due to the greater perceived sharpness, cf. Figure 11b below), whereas for medium to high compression ratios we obtain higher quality with JPEG2000. This confirms the common observation that JPEG2000 works best for high compression ratios. JPEG breaks down at compression ratios above 100:1, i.e. the images generally get distorted beyond recognition, whereas JPEG2000 still produces more or less acceptable images at this rate.


Figure 9. Average MOS predictions as a function of compression ratio for JPEG and JPEG2000.

If we just consider PSNR to compare JPEG and JPEG2000, which is shown in Figure 10, we find JPEG2000 to be a better encoder than JPEG for all compression ratios. This confirms again that the MOS prediction by Image Optimacy is a better indicator of image quality than PSNR.
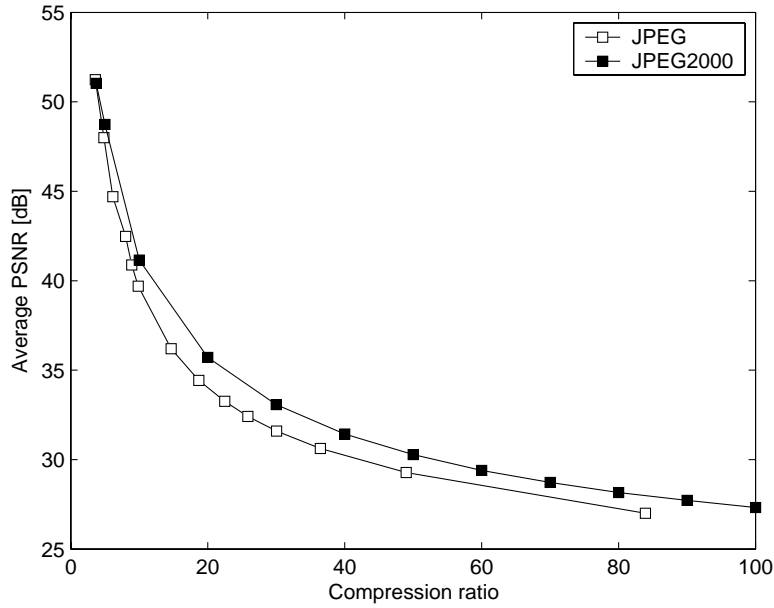
Figure 10. Average PSNR as a function of compression ratio for JPEG and JPEG2000.

The average blockiness and the average blur are shown as a function of the compression ratio in Figure 11. As expected, the JPEG2000 encoder produces a lot of blur but virtually no blockiness, while for JPEG-encoded images blockiness is the main artifact. This is in line with the encoder-specific compression algorithms (wavelets vs. block-based DCT) and demonstrates the correct responses of the two artifact metrics.

(a) Blockiness

(b) Blur
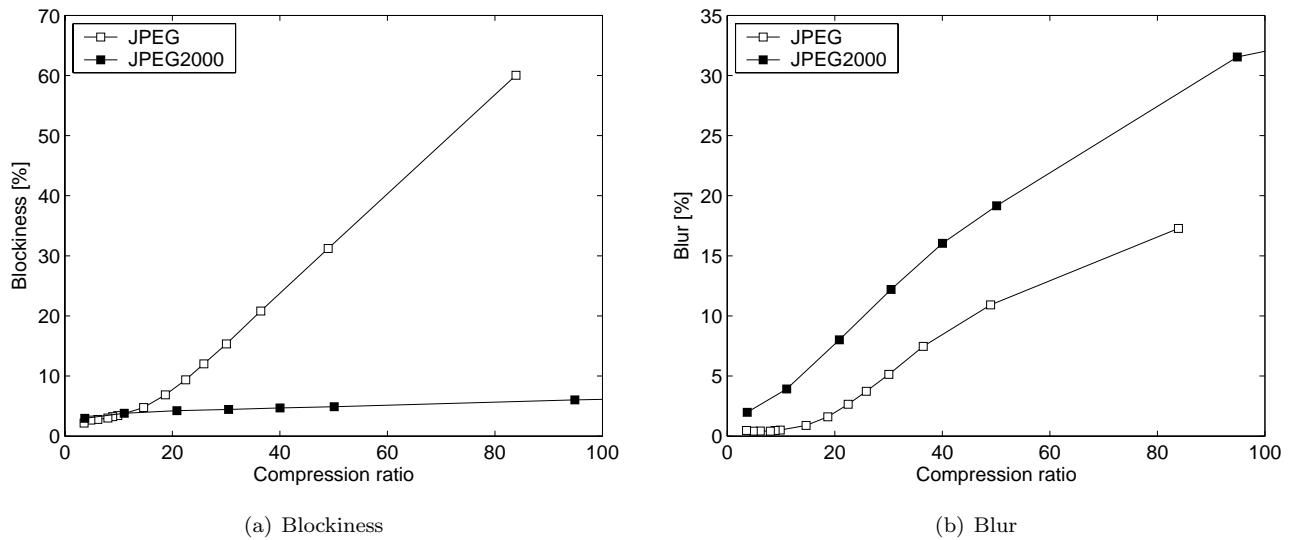
Figure 11. Average blockiness (a) and blur (b) as a function of compression ratio for JPEG and JPEG2000.

5. CONCLUSIONS

In this paper, we presented a tool that measures perceptual artifacts in an image and uses them to estimate the Mean Opinion Score (MOS) obtained in subjective experiments. We demonstrated that the MOS predictions by the proposed tool are a better indicator of perceived image quality than PSNR, especially for highly compressed images.

For the encoder comparison, we compressed a set of 29 test images with two JPEG encoders and three JPEG2000 encoders at various compression ratios. Using our Image Optimacy tool, we computed blockiness, blur, and MOS predictions as well as PSNR of the compressed images.

Our comparison of JPEG encoders showed that the IrfanView JPEG encoder produces consistently better images than the Adobe Photoshop JPEG encoder at the same data rate. We also found an inconsistency with the quality factor settings for the Adobe Photoshop JPEG encoder.

The differences between the three JPEG2000 encoders in our test were less pronounced. According to our measurements, JasPer was the best codec, closely followed by IrfanView and Kakadu.

Comparing JPEG- and JPEG2000-encoding with IrfanView, JPEG had a slight quality edge at low compression ratios (below 20:1), while JPEG2000 was the clear winner at medium and high compression ratios.

REFERENCES

1. D. Santa-Cruz, R. Grosbois, and T. Ebrahimi, "JPEG 2000 performance evaluation and assessment," *Signal Processing: Image Communication* **17**, pp. 113–130, Jan. 2002.
2. G. K. Wallace, "The JPEG still picture compression standard," *Comm. ACM* **34**, pp. 30–44, April 1991.
3. S. Winkler, A. Sharma, and D. McNally, "Perceptual video quality and blockiness metrics for multimedia streaming applications," in *Proc. International Symposium on Wireless Personal Multimedia Communications*, pp. 547–552, (Aalborg, Denmark), Sept. 9–12, 2001.
4. P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: Application to JPEG2000," *Signal Processing: Image Communication* **19**, pp. 163–172, Feb. 2004.
5. M. Rabbani and R. Joshi, "An overview of the JPEG2000 still image compression standard," *Signal Processing: Image Communication* **17**, pp. 3–48, Jan. 2002.
6. Genista Corporation. http://www.genista.com/.
7. ANSI T1.801.03, "Digital transport of one-way video signals – parameters for objective performance assessment." American National Standards Institute, New York, NY, 1996.
8. H. R. Sheikh, A. C. Bovik, L. Cormack, and Z. Wang, "LIVE image quality assessment database." http://live.ece.utexas.edu/research/quality, 2003.
9. S. Süsstrunk and S. Winkler, "Color image quality on the Internet," in *Proc. SPIE*, **5304**, pp. 118–131, (San Jose, CA), Jan. 19–22, 2004.
10. Adobe Photoshop. http://www.adobe.com/products/photoshop/main.html.
11. IrfanView. http://www.irfanview.com/.
12. JasPer JPEG2000 codec. http://www.ece.uvic.ca/~mdadams/jasper/.
13. Kakadu JPEG2000 codec. http://www.kakadusoftware.com/.