

# Tuning JPEG2000 Image Compression for Graphics Regions

Serene Banerjee and Brian L. Evans

Dept. of ECE, The University of Texas, Austin, TX 78712-1084 USA

{serene,bevans}@ece.utexas.edu

## Abstract

*The new image compression standard, JPEG2000, provides higher compression rates for the same visual quality for grayscale and color images than JPEG. JPEG2000 is being adopted for image compression and transmission, e.g. in mobile phones, PDAs, and wearable computers. Depending on the application, images may contain formatted text and graphics data. For graphics images at the same low bit rates, graphics compression methods, such as the graphics interchange format (GIF) and portable network graphics (PNG), outperform JPEG2000 in visual quality. In this paper, we describe problems associated with compressing graphics data with JPEG2000, and propose modifications to a JPEG2000 encoder to minimize distortion for color graphics data by using a model of the human visual system. The visual improvements are quantified by a new measure of distortion.*

## 1 Introduction

JPEG2000 is an ISO/ITU-T still image compression standard that supports lossy and lossless compression of single component (e.g. grayscale) and multiple component (e.g. color) images. When compared to JPEG, JPEG2000 has higher compression rates for the same visual quality over a wide range of images. For wireless imaging applications, JPEG2000 offers progressive transmission and rendering with region of interest coding. JPEG2000 has a flexible file structure, which is useful for Web browsing, display on PDAs, and high resolution printing. Images may contain formatted text, such as subtitles and computer menus, and graphics data, such as cartoons. For graphics data, however, JPEG2000 performs worse at low bit rates than prevalent graphics compression codecs.

In this paper, Section 2 summarizes the JPEG 2000 standard. Section 3 analyzes the reasons for poor performance of JPEG2000 on graphics images. Section

4 develops standard-compliant modifications in computation of the quantization steps in order to reduce perceived visual distortions for graphics data at low bit rates. Section 5 proposes a low-complexity post-processing filter to reduce ringing. Section 6 proposes a new distortion measure to quantify ringing distortions. Section 7 concludes the paper. The source code and full-sized images are available at

[signal.ece.utexas.edu/~serene/software/ssiai02/](http://signal.ece.utexas.edu/~serene/software/ssiai02/)

In generating JPEG2000 compressed images, we use version 3.08 of the Kakadu JPEG2000 codec [1], which is a C++ developers toolkit.

## 2 JPEG2000

The JPEG2000 [2, 3] standard supports lossless and lossy compression. When processing a color image, the red, green and blue (RGB) components are transformed to luminance and chrominance (YCbCr) components, as the human visual system has a lower spatial bandwidth for chrominance components. However, instead of downsampling the chrominance components by two, as in JPEG or MPEG coding standards, JPEG2000 uses weighting tables to give less importance to these components. The color component transformation can be either irreversible or reversible for lossy compression, but only reversible for lossless compression. The Y, Cb, and Cr components are processed separately.

In JPEG2000, an image can be broken up into non-overlapping rectangular tiles, and each tile is coded independently. A wavelet transform is performed on each tile, which creates decomposition levels. These decomposition levels are subbands of coefficients that characterize the local frequency of the tiles and that are quantized and organized in rectangular arrays as coded blocks. A group of three spatially consistent rectangles, i.e. one from each subband at each resolution level, is called a packet partition location or precinct.

To achieve compression by exploiting spatial redundancies of an image, the bit planes of these coded blocks are entropy coded in three coding passes—significance propagation, magnitude refinement, and cleanup. In the significance propagation pass, a bit is coded if its location is not significant, but at least one of its eight connected neighbors is significant. In the magnitude refinement pass, the bits that became significant in the previous bit plane are encoded. In the cleanup pass, all the bits that were not encoded in the previous passes are encoded. The cleanup pass uses both neighborhood context and runlength context.

To incorporate region of interest (ROI) coding, preference is given to the order in which these subbands of coefficients are coded. Also, certain ROIs can be coded with higher quality than the background. Markers can be added to the bitstream for error resilience.

The tile components are preceded by a tile header. And, the entire codestream is preceded by the main header, which describes the attributes of the original image; the decompositions and coding styles that would be used to locate, extract, decode, and reconstruct the image with the desired fidelity, resolution, region of interest; and other characteristics. A description of the meaning of the image and its components in the context of the application can be described in an optional file format.

In JPEG2000 decoding, the main and tile headers are read to obtain the characteristics of the image and its tile components. The bitstream is entropy decoded, in the significance, refinement, and cleanup decoding passes, to obtain code blocks of quantized coefficients. These code blocks are then inverse quantized and inverse wavelet transformed to reconstruct the image.

### 3 Visual Artifacts

JPEG2000 supports frequency weighting [4, 5], which is based on a (lowpass) contrast sensitivity function (CSF) of the human visual system, and pixel masking [6, 7] to minimize the perceived distortion in compressed natural images. A CSF, which is based on the human eye’s reduced sensitivity to higher spatial frequencies, weights the frequency coefficients. The CSF has a stronger bias to luminance components than the Cb and Cr components. The quantizer step sizes, or the distortion measures computed during quantization, are calculated as a function of the CSF. Previous work derives optimal CSF tables for natural images [4, 5, 6, 7]. These CSF weights are determined based on just noticeable detection (JND) thresholds [8], which assume visually perceived near-lossless conditions. These tables are not optimized for graphic im-

ages.

Due to truncation of high frequency wavelet coefficients in JPEG2000, at low bit rates, spurious oscillations are observed in the vicinity of edges in the decompressed image [9]. This is known as *ringing artifacts*. However, for graphics blocks, a large number of adjacent pixels often have the same luminance values, or adjacent pixels have a large amplitude difference in luminance values [10]. The second property gives rise to a large number of edges in a graphics image. So, JPEG2000 decompressed graphic images suffer from ringing artifacts.

Fig. 1 shows four, 24 bits per pixel (bpp) natural images compressed with JPEG2000 at 0.3 bpp. These appear lossless to the human eye. Fig. 2 shows graphics images compressed at the same rate. Ringing artifacts are visible in the results. If the graphic images were coded losslessly with JPEG2000, then the number of results bits obtained would be much higher than if they had been coded losslessly with Portable Network Graphics (PNG) or Graphics Interchange Format (GIF) coders.

## 4 Encoder Optimizations

Zeng *et al.* suggests modifications to a JPEG2000 encoder based on a human visual system model in order to reduce perceived artifacts. These modifications are optimized for grayscale natural images. We evaluate how these modifications perform on color graphic images. The modifications can be classified into two categories: (1) visual frequency weighting, and (2) visual frequency masking.

In visual frequency weighting, the wavelet coefficients at different levels are weighted by the contrast sensitivity function (CSF) modeling the human eye. These weights can either be fixed or progressively varying based on the image content. Typical weights assigned to the coefficients are shown in Fig. 4, where the low-frequency coefficients have a weight of one, and the other weights decrease with increasing frequency. However, this does not work well for graphic images. The perceived distortions in these images are visible and are more than the just noticeable difference (JND), as shown in Fig. 3. Three visual masking techniques are applied—self-contrast, neighborhood, and point-wise extension. In self-contrast masking, the coefficients are modulated by a non-linear function, and then quantized uniformly, which gives rise to non-uniform quantization of coefficients. This non-uniform quantization takes into account the masking property of human visual system. The transformation is

$$x_i \rightarrow y_i = x_i^\alpha, \quad 0 < \alpha \leq 1 \quad (1)$$

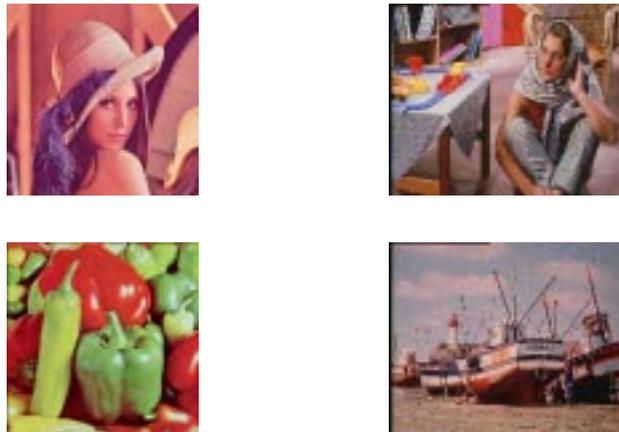


Figure 1. No visible distortion artifacts noticed in JPEG2000 compressed natural images at 0.3 bpp.



Figure 2. Noticeable ringing artifacts in JPEG2000 compressed graphic images at 0.3 bpp.

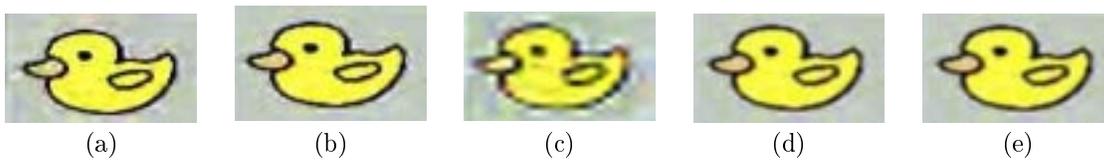


Figure 3. Performance on graphic images compressed at 0.3 bpp with and without visual optimizations: (a) Compressed with minimizing mean-squared error; (b) Contrast sensitivity function (CSF) weighted compression; (c) Visual masking optimization enabled result; (d) CSF weighted image post-processed with lower threshold; and (e) CSF weighted image post-processed with higher threshold.

With  $\alpha = 0.5$ , distortion is reduced for diagonal edges. With  $\alpha$  approaching one, distortions in horizontal edges are reduced. Self-contrast masking works well for natural textures, but does not work well for graphics regions.

In neighborhood masking, the distortion measure in the rate-distortion optimization is modified. The distortion is weighted by a visual masking factor that is a function of the neighborhood coefficients. This is better for images with edges. Mathematically, the distortion estimates are given by

$$V_i' = \frac{V_i}{M_i}, \quad \text{where} \quad M_i = A \sum_{k \text{ near } i} |V_k|^\gamma \quad (2)$$

and  $V_i$  is the distortion.

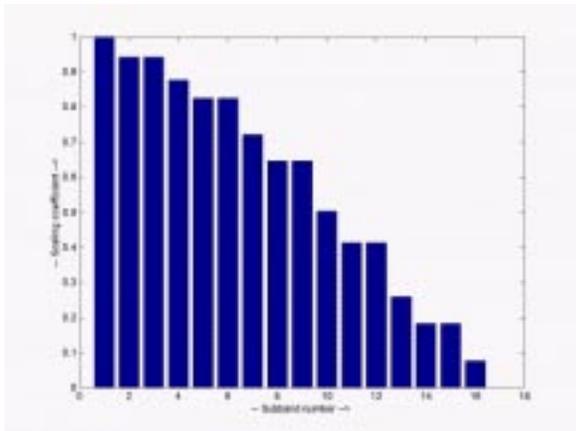
Pointwise extended masking exploits the properties of both self-contrast and neighborhood masking, by first modulating the quantization coefficients and then applying neighborhood masking. Mathematically,

$$x_i \rightarrow y_i = \text{sign}(x_i)|x_i|^\alpha \quad (3)$$

The distortion is then calculated as

$$z_i = \frac{y_i}{(1 + a \sum_{k \text{ near } i} |\hat{x}_k|^\beta / |\Phi_i|)}, \quad (4)$$

where  $|\Phi_i|$  is the size of the neighborhood. The constants  $\alpha$  and  $\beta$  were varied to generate the best visual results, which are shown in Fig. 3.



**Figure 4. An example of predetermined contrast sensitivity function weights. The horizontal axis is in units of subband decomposition number.**

## 5 Post-processing Technique

The designed low-complexity post-processing filter removes the ringing artifacts noticeable in Fig. 3. Previous work in by Yang *et al.* [9] uses parameter estimation and maximum likelihood approach for solving the problem. Graphics images are characterized by slightly varying pixel values on the whole and abrupt changes in sample values at edges. So, a minimum and maximum threshold is defined which can characterize an edge in the graphics image. Initially the thresholds are computed from row-wise, column-wise and diagonal scanning of the original image. Ringing artifacts create false edges. However, in a four-by-four window the pixel-to-pixel difference at these ringing artifacts is either below the minimum threshold or above the maximum threshold. So, if the pixel-to-pixel difference of a pixel and its four connected neighbors lies beyond the threshold values, the value of the pixel is substituted by a value from its neighbor. The output image is thus cleaned up of ringing artifacts. For more visually appealing results the generated image is lowpass filtered with a  $3 \times 3$  filter with coefficients

1	2	1
2	4	2
1	2	1

Fig. 3 shows results before and after post-filtering.

## 6 Quantifying Ringing Distortion

For distortion measurement, we only consider the Y (luminance) component. The peak signal-to-noise ratio (PSNR) improvement does not correlate well with the human eye, as it assumes that the error image is independent, whereas ringing artifacts are correlated with the edge map of the original image. Since PSNR is a common figure of merit, we show the PSNR values before and after masking in Table 1. In this section, we define a new measure to quantify ringing distortion.

Due to the masking effect of the human eye, distortions in the vicinity of an edge will not appear too objectionable [8]. But, if ringing propagates to the smooth regions, then the distortion will be visible. First the residual image is computed as the difference between the original and the compressed image. The residual image is masked with the edge map of the original image. Then, correlation between the new residual image and the edge map is used as a measure of distortion. Since ringing is highly correlated with the edge map, decorrelating the noise from the edge map will appear to be better. The results with this new

measure are tabulated in Table 2. The correlation values of post-filtered image being less than that of CSF weighted image, indicates post-filtered image has less distortion. However, as visual masking and minimizing the mean square error (MSE) approaches produce artifacts other than ringing, such as blocking and abrupt change in pixel values, this measure cannot be used to quantify the distortions in those two cases. Arslan *et al.* [11] also use correlation with edge map as a measure to quantify distortion in synthetic aperture radar images.

Image (80:1 comp.)	Min MSE	CSF weight	Visual mask	Post filter
Graphic_school	52.8	49.4	39.3	41.8
See_saw	40.1	39.1	30.4	34.9
Show_n_tell	47.4	45.1	37.8	41.4
Swing	52.9	53.0	46.8	49.7
Welcome	36.3	35.9	28.7	32.2
Caterpillar	39.9	39.8	33.8	37.1

**Table 1. PSNR measurements for six compressed graphic images at 0.3 bpp, with and without visual optimizations.**

## 7 Conclusion

This paper explores visual improvements for graphics images and presents a low-complexity post-processing filter to minimize ringing distortions, for graphics images compressed with JPEG2000 at low bit rates. It also proposes a new method to quantify ringing in graphics images, based on psychovisual characteristics of the human eye. Although post-processing techniques give better results, it will add to the complexity of the JPEG2000 codec.

The visual frequency weighting and visual masking optimizations work well for natural images, but fail to provide noticeable improvement for graphics. This is due to the abrupt changes in pixel values for graphics, leading to high-frequency content in the image. Ringing distortion arises when these high frequency coefficients are quantized to achieve low bit rates.

## References

[1] D. Taubman, "Kakadu," <http://maestro.ee.unsw.edu.au/~taubman/kakadu/>, Nov. 2001.

Image (80:1 comp.)	Min MSE	CSF weight	Visual mask	Post filter
Graphic_school	0.0104	0.0201	0.0135	0.0138
See_saw	0.0170	0.0215	0.0096	0.0248
Show_n_tell	0.0066	0.0180	0.0004	0.0106
Swing	0.0131	0.0155	0.0019	0.0164
Welcome	0.0109	0.0209	0.0106	0.0140
Caterpillar	0.0092	0.0129	0.0006	0.0110

**Table 2. Edge correlation values for six compressed graphic images at 0.3 bpp, with and without visual optimizations.**

- [2] D. T. Lee and M. Boliek, "Information Technology - JPEG 2000 Image Coding System: Compound Image File Format," *ISO/IEC JTC 1/SC 29/WG 1 N2268*, July 2001.
- [3] C. Christopoulos, A. Skodras, and T. Ebrahimi, "The JPEG2000 Still Image Coding System: An Overview," *IEEE Trans. Consumer Elect.*, vol. 46, pp. 1103–1127, Nov. 2000.
- [4] W. Zeng and S. Lei, "CSF Weighting Strategy for Visual Progressive Coding," *ISO/IEC JTC 1/SC 29/WG 1 N1584*, Mar. 2000.
- [5] W. Zeng and T. Chinen, "Evaluation of the Distortion-Adaptive Progressive CSF Weighting Technique," *ISO/IEC JTC 1/SC 29/WG 1 N1716*, July 2000.
- [6] S. Daly, W. Zeng, J. Li, and S. Lei, "Visual Masking in Wavelet Compression for JPEG2000," in *Proc. SPIE Conf. Image & Video Comm. & Proc.*, vol. 3974, Jan. 2000.
- [7] W. Zeng, S. Daly, and S. Lei, "Point-wise Extended Visual Masking for JPEG2000 Image Compression," in *Proc. IEEE Int. Conf. Image Proc.*, vol. 1, pp. 657–660, Sept. 2000.
- [8] W. Zeng, S. Daly, and S. Lei, "An Overview of the Visual Optimization Tools in JPEG2000," *IEEE Trans. Circuits & Sys. Video Tech.*, vol. 17, Oct. 2001.
- [9] S. Yang, Y.-H. Hu, T. Q. Nguyen, and D. L. Tull, "Maximum-Likelihood Parameter Estimation for Image Ringing-Artifact Removal," *IEEE Trans. Circuits & Sys. Video Tech.*, vol. 11, pp. 963–973, Aug. 2001.
- [10] M. van der Schaar and P. H. N. de With, "Hybrid Compression of Video with Graphics in DTV Communication Systems," *IEEE Trans. Consumer Elect.*, vol. 46, pp. 1007–1017, Nov. 2000.
- [11] G. Arslan, M. Valliappan, and B. L. Evans, "Quality Assessment of Compression Techniques for Synthetic Aperture Radar Images," in *Proc. IEEE Int. Conf. Image Proc.*, vol. 3, pp. 857–861, Oct. 1999.