Perceptual Criteria on JPEG2000 Quantization

Jaime Moreno, Xavier Otazu and Maria Vanrell

Computer Science Department, Computer Vision Center, Autonomous University of Barcelona
08193, Bellaterra, Cerdanyola del Vallès, Barcelona, Spain
E-mail: jmoreno@cvc.uab.es

Abstract

The aim of this work is to explain how the Brightness Induction Wavelet Model used as a perceptual quantizer can be useful for improving a lossy compression and to introduce preliminary results. In fact, this approach consists in quantizing wavelet transform coefficients using the human visual system behavior properties. When compressing images, noise is fatal to compression performance, it can be both annoying for the observer and consuming excessive amounts of bandwidth when the imagery is transmitted. The perceptual quantization based on a chromatic induction reduces unperceivable details and thus improve both visual impression and transmission properties. The comparison between JPEG2000 without and with perceptual quantization shows that the latter is not favorable in PSNR, but the recovered image is more compressed at the same or even better visual quality measured with a weighted PSNR.

Keywords: Human Visual System, Contrast Sensitivity Function, Perceived Images, Wavelet Transform, Compression Algorithms, Bandwidth Reduction, Peak Signal-to-Noise Ratio.

1 Introduction

Digital image compression has been a topic of research for many years and a number of image compression standards has been created for different applications. The JPEG2000 [2] is intended to provide rate-distortion and subjective image quality performance superior to existing standards, as well as to supply functionality. However JPEG2000 do not provide the most relevant characteristics of the human visual system, since for removing information in order to compress the image only the information theory criteria are applied. This information removal introduces artifacts to the image that are visible at high compression rates, since the compression is based in a data loss from a numerical threshold, because of the discard of many pixels with high perceptual significance.

Hence it is necessary an advanced model that removes information from its perceptual content, which preserves the pixels with high perceptual relevance regardless of the numerical information. The Brightness Induction Wavelet Model (BIWaM) is suitable for it. Both BIWaM and JPEG2000 use wavelet transform, but BIWaM uses it in order to generate an approximation to how every pixel is perceived from a certain distance taking into account the value of its neighboring pixels. By contrast, JPEG2000 applies a perceptual criteria for all coefficients in a certain spatial frequency independently of the values of its surrounding ones.

BIWaM attenuates the details that the human visual system is not able to perceive, enhances those that are perceptually relevant and produces an approximation to the image that the brain detects. At long distances, as Figure 3(d) depicts, the lack of information does not produce the well-known compression artifacts, rather it is presented as a softened version, where the details with high perceptual value remain.

The paper is organized as follows: Section 2 specifies quantization and dequantization model used by JPEG2000 for encoding and reconstruction of wavelet coefficients, thereby is described the Deadzone Uniform Scalar Quantizer and the Visual Frequency Weighting. Section 3 describes the Brightness Induction Wavelet Model. In Section 4 the proposed method of quantization will be discussed. Experimental results applied for some test images are given in section 5. Ultimately, section 6 is where the conclu-

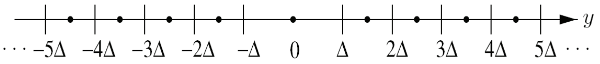


Figure 1: Dead-zone uniform scalar quantizer with step size Δ.

sions and future work will be exposed.

2 JPEG2000 Quantization Review

2.1 Dead-zone Uniform Scalar Quantizer

A uniform scalar quantizer is a function that maps each element in a subset of the real line to a particular value, which ensures that more zeros result [3]. In this way all thresholds are uniformly spaced by step size Δ, except for the interval containing zero, which is called the dead-zone and extends from $-\Delta$ to $+\Delta$, thus a dead-zone means that the quantization range about 0 is 2Δ.

For each spatial frequency s, a basic quantizer step size Δ_s is used to quantize all the coefficients in that spatial frequency according to Equation 1.

$$q = sign(y) \left\lfloor \frac{|y|}{\Delta_s} \right\rfloor \qquad (1)$$

where y is the input to the quantizer or original wavelet coefficient value, $sign(y)$ denotes the sign of y and q is the resulting quantizer index. Figure 1 illustrates such a quantizer with step size Δ, where vertical lines indicate the endpoints of the quantization intervals and heavy dots represent reconstruction values.

The inverse quantizer or the reconstructed $\hat{y}$ is given by the Equation 2, wherein δ is a parameter often set to place the reconstruction value at the centroid of the quantization interval and varies form 0 to 1.

$$\hat{y} = \begin{cases} (q+\delta)\Delta_s, & q > 0 \\ (q-\delta)\Delta_s, & q < 0 \\ 0, & q = 0 \end{cases} \qquad (2)$$

The International Organization for Standardization recommends [2], the δ values are both 0.5 and 0.375, whereas Pearlman and Said suggest [6] $\delta = 0.38$, which places the reconstruction at the intervals midpoint. It is important to realize that when $-\Delta < y < \Delta$, the quantizer level and reconstruction value are both 0. For a spatial frequency, there may be many coefficients usually those of higher frequencies, that are set to 0. The array of quantizer levels q is further encoded losslessly.

s	HL	LH	HH
1	1	1	1
2	1	1	0.731 668
3	0.564 344	0.564 344	0.285 968
4	0.179 609	0.179 609	0.043 903
5	0.014 774	0.014 774	0.000 573

Table 1: Recommended frequency weighting for 400 dpi's

2.2 Visual Frequency Weighting

In JPEG2000, only one set of CSF weights is chosen and applied according to a particular viewing condition (100, 200 or 400 dpi's) with fixed visual weighting. This viewing condition may be truncated depending on the stages of embedding, in other words at low bit rates, the quality of the compressed image is poor and the detailed features of the image are not available since at a relatively large distance the global features are more important. As more bits are received, the image quality improves, which is equivalent to decreasing the viewing distance.

The table 1 specifies a set of CSF weights which was designed for the luminance component based on the CSF value at the mid-frequency of each spatial frequency. The viewing distance is supposed to be 4000 pixels, corresponding to 10 inches for 400 dpi print or display. The table does not include the weight for *LL*, because it is always 1. Levels 1, 2, . . . , 5 denote the spatial frequency levels in low to high frequency order with three spatial orientations (*HL, LH, HH*).

3 Brightness Induction Wavelet Model

In order to explain the brightness assimilation/contrast phenomena as a unique perceptual process, Otazu et al. in [5] proposed a low-level brightness induction model, which combines three important stimulus properties: Spatial frequency, Spatial orientation and Surround contrast.

Thereby the achromatic input image $\mathcal{I}$ is separated into different spatial frequency and orientation components from a multiresolution wavelet decomposi-

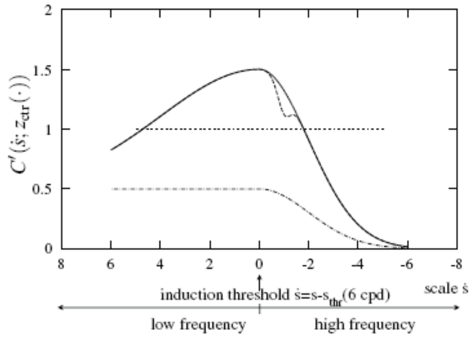


Figure 2: Contrast Sensitivity Function.

tion. Thus every single transformed coefficient is weighted using the response of the contrast sensitivity function (CSF, Figure 2) hence a perceptual brightness image $\mathcal{I}_\rho$ is recovered. The CSF is modified considering both spatial surround information and observation distance, in this way the CSF value decreases when the surround contrast increases and vice versa.

$\mathcal{I}$ can be decomposed a set of wavelet planes ω of different spatial frequencies, where each wavelet plane contains details at different resolutions of $\mathcal{I}$ and it is described by:

$$\mathcal{I} = \sum_{s=1}^{n} \sum_{o=v,h,d} \omega_s^o + c_n \tag{3}$$

where n is the number of wavelet planes computed. The term c_n is the residual plane and the index o represents the spatial orientation either vertical, horizontal or diagonal at a certain spatial frequency.

The perceptual image $\mathcal{I}_\rho$ recovered from the wavelet planes is defined by:

$$\mathcal{I}_\rho = \sum_{s=1}^{n} \sum_{o=v,h,d} C'\left(\dot{s}, z_{ctr}\left(s,o\right)\right) \cdot \omega_s^o + c_n \tag{4}$$

The term $C'\left(\dot{s}, z_{ctr}\left(s,o\right)\right)$ is a weighting function, that tries to emulate some perceptual properties of human visual system, has a shape similar to the CSF and can be written as:

$$C'\left(\dot{s}, z_{ctr}\left(s,o\right)\right) = z_{ctr} \cdot C_d(\dot{s}) + C_{min}(\dot{s}) \tag{5}$$

where z_{ctr} is a non-linear function and an estimation of the central feature contrast relative to its surround contrast. Its range oscillates from zero to one and is defined by:

$$z_{ctr} = \frac{\left[\frac{\sigma_{cen}}{\sigma_{sur}}\right]^2}{1 + \left[\frac{\sigma_{cen}}{\sigma_{sur}}\right]^2} \tag{6}$$

being σ_{cen} and σ_{sur} the standard deviation of the wavelet coefficients in two concentric rings, which represent a center−surround interaction around each coefficient.

The weighting function $C_d(\dot{s})$ is an approximation to the perceptual CSF [4] and to emulate some perceptual properties and is defined as a piecewise Gaussian function, such as:

$$C_d(\dot{s}) = \begin{cases} e^{-\frac{\dot{s}^2}{2\sigma_1^2}}, & \dot{s} = s - s_{thr} \le 0, \\ e^{-\frac{\dot{s}^2}{2\sigma_2^2}}, & \dot{s} = s - s_{thr} > 0 \end{cases} \tag{7}$$

The term $C_{min}(\dot{s})$ avoids the $C'\left(\dot{s}, z_{ctr}\left(s,o\right)\right)$ function to be zero and is defined by:

$$C_{min}(\dot{s}) = \begin{cases} \frac{1}{2}\, e^{-\frac{\dot{s}^2}{2\sigma_1^2}}, & \dot{s} = s - s_{thr} \le 0, \\ \frac{1}{2}, & \dot{s} = s - s_{thr} > 0 \end{cases} \tag{8}$$

taking $\sigma_1 = 2$ and $\sigma_2 = 2\sigma_1$ so as to reproduce the approximate profile of the psychophysical functions. Both $C_{min}(\dot{s})$ and $C_d(\dot{s})$ depend on the factor s_{thr}, which is the scale associated to an induction threshold value equal to 4cpd when an image is observed from a distance d with a pixel size l_p and 1 visual degree, whose expression is defined by Equation 9.

$$s_{thr} = \log_2\left(\frac{d \tan(1^\circ)}{4\, l_p}\right) \tag{9}$$

Figure 3 shows three BIWaM images of *Lena*, which were calculated for a 19 inch monitor with 1280 pixels of horizontal resolution, at 30, 100 and 200 centimeters of distance.

4 Perceptual Method of Quantization

The block diagram of the JPEG2000 modification is illustrated in figure 4. To obtain transformed coefficients or $\mathcal{I}$ a Forward Transformation with the 9/7 filter fast wavelet transform is first applied on the source

(a) Original (b) 30 cm.

(c) 100 cm. (d) 200 cm.

Figure 3: BIWaM images of *Lena*.

image data. Then the perceptual quantized coefficients or $\mathcal{Q}$ with a known viewing distance are calculated by:

$$\mathcal{Q} = \sum_{s=1}^{n} \sum_{o=v,h,d} sign(\omega_s^o) \left\lfloor \frac{\left|C'\left(\dot{s}, z_{ctr}\left(s,o\right)\right) \cdot \omega_s^o\right|}{\Delta_s} \right\rfloor + \quad sign(c_n) \left\lfloor \frac{|c_n|}{\Delta_n} \right\rfloor \qquad (10)$$

This expression is similar to Equation 1, but introduces a perceptual criteria. A normalized quantization step size Δ equal to $1/128$ is used, namely the range between the minimal and maximal values at $\mathcal{I}_\rho$ is divided into 128 intervals. Finally, the perceptual quantized coefficients are entropy coded, before forming the output code stream or bitstream.

At the decoder, the code stream is first entropy decoded in order to reconstruct the perceptual quantized coefficients $\widehat{\mathcal{Q}}$. Second it is dequantized using 2 with a normalized quantization step size Δ equal to $1/128$ and δ equal to $3/8$. Finally, an inverse discrete transformed to recover $\widehat{\mathcal{I}_\rho}$, thus providing the reconstructed perceived image data.

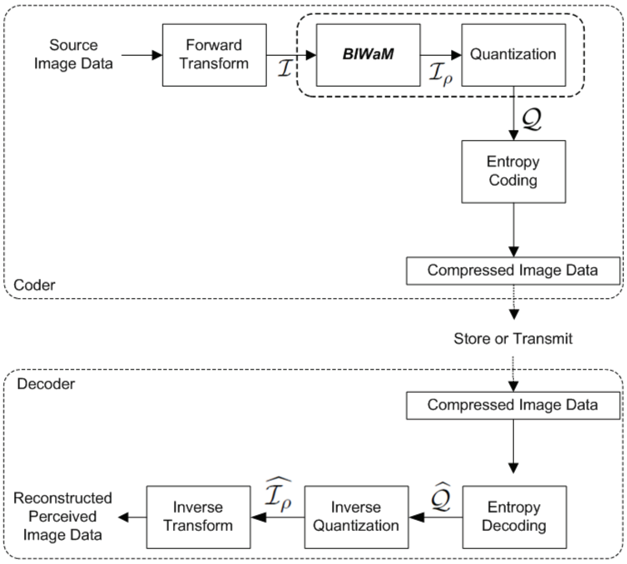


Figure 4: General block diagram of JPEG2000 compression.

5 Experimental Results

The Perceptual Criteria on JPEG2000 Quantization were tested on 44 images, but only results of the images *Peppers* and *Baboon* are reported, which are 256 gray-scale images and 512×512 of resolution (Figure 5). The BIWaM images were calculated for a 19 inch monitor with 1280 pixels of horizontal resolution at 50 centimeters of viewing distance.

(a) Peppers (b) Baboon

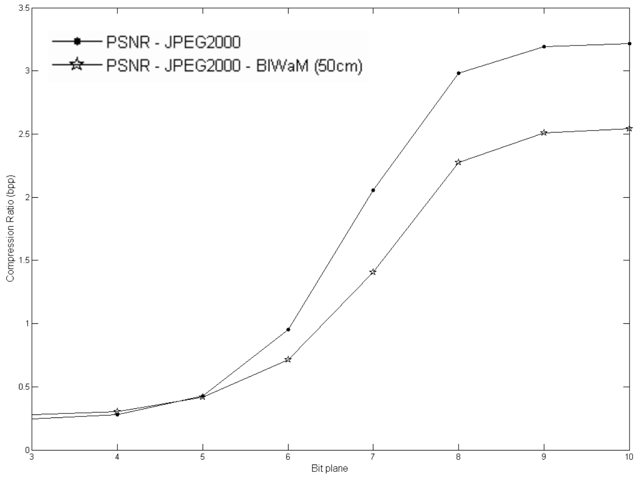
Figure 5: Tested Images.

In order to measure the distortion between the original image $f(i,j)$ and the reconstructed image $\hat{f}(i,j)$ The Peak Signal to Noise Ratio was employed, however PSNR does not calculate perceptual quality measures. Therefore, it is necessary to weight each PSNR term by means of its local activity factor, taking into account the local variance of the neighbors of the studied wavelet coefficients, thus defining a weighted PSNR or wPSNR [1]. The wPSNR increases with increasing variance and vice versa as:

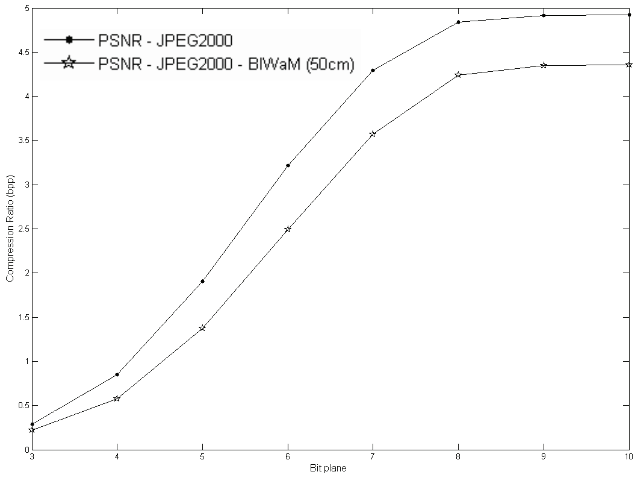
$$wPSNR = 10\log_{10}\left(\frac{\mathcal{G}_{max}{}^{2}}{wMSE}\right) \qquad (11)$$

where $\mathcal{G}_{max}$ is the maximum possible intensity value in $f(i,j)$ ($M \times N$ size) and weighted MSE (wMSE) is defined as:

$$wMSE = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\left[\frac{f(i,j)-\hat{f}(i,j)}{1+Var(i,j)}\right]^2 \qquad (12)$$

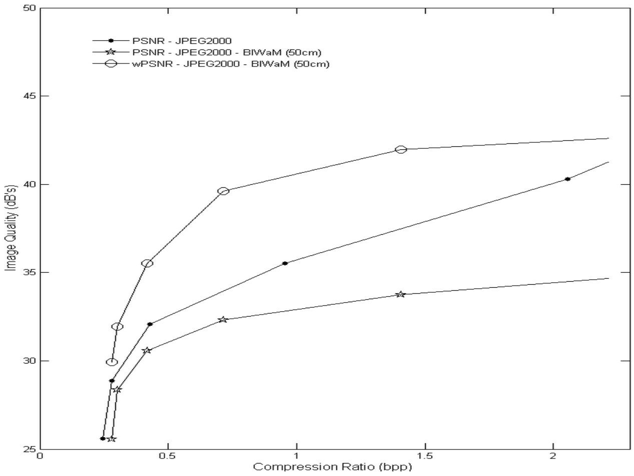


(a) Peppers.

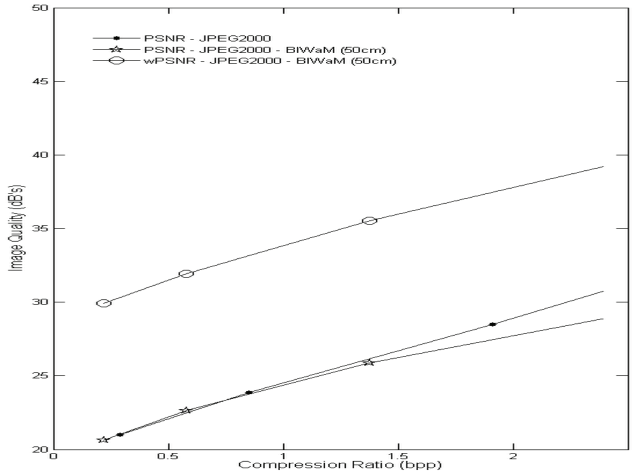


(b) Baboon.

Figure 6: Bit-plane compression ratio.



(a) Peppers.



(b) Baboon.

Figure 7: Comparison between compression ratio and image quality.

Figure 6 shows the assessment results of the compression performance at every bit-plane for a Deadzone Uniform Scalar Quantizer and also for a BIWaM Quantizer. In both Figure 6(a) and Figure 6(b) a BIWaM Quantizer achieves better compression ratios with the same threshold, that is because BIWaM reduces unperceivable coefficients. For example at the tenth bit-plane of *Peppers* a BIWaM Quantizer diminishes 21 percent less bits per pixel than a Scalar Quantizer, namely 22.3KB of information is perceptually irrelevant at 50 centimeters.

The comparison between compression ratio and image quality is depicted by the Figure 7, which shows that the reconstructed images quantized by BIWaM has less PSNR but higher wPSNR than the ones quantized by a scalar way, i.e. even if the reconstructed image has a lower objective quality, this image could have a higher perceptual quality.

(a) JPEG2000. (b) Upper-left Quadrant.

(c) JPEG2000-BIWaM. (d) Upper-left Quadrant.

Figure 8: Reconstructed images compressed at 0.29 bpp.

The Figures 8(a) and 8(c) shows an example of reconstructed images compressed at 0.29 bits per pixel by means of JPEG2000 without and with perceptual Quantization, respectively. PSNR in 8(a) is 21.01 decibels and in 8(c) is 20.32 decibels but wPSNR is equal to 29.08 decibels, namely the reconstructed image quantized by BIWaM is perceptually better than the one quantized by a Scalar Quantizer, since the latter has more compression artifacts, as Figures 8(b) and 8(d) illustrate.

6 Conclusions and Future Work

This paper proposes an alternative of quantization for JPEG2000 using BIWaM. In order to measure the effectiveness of the perceptual quantization a performance analysis is done using the PSNR and wPSNR measured between reconstructed and original images. Unlike PSNR, wPSNR uses not only a single coefficient but also its neighbors as well as its psycho-visual properties. The experimental results show that a BIWaM Quantization can help to improve the compression and image perceptual quality. One of the future tasks is the use of a threshold based on the CSF properties, namely a threshold based on perceptual importance that a coefficient has regardless of its numerical value.

Acknowledgements

This work was partially supported by The Spanish Ministry of Education and Science: Project TIN2007-64577 and The Mexican Science and Technology National Council: Financial Aid number 207950.

References

[1] M. Bertini, R. Cucchiara, A. D. Bimbo, and A. Prati, “An integrated framework for semantic annotation and adaptation,” *Multimedia Tools and Applications*, vol. 26, no. 3, pp. 345–363, August 2005.

[2] M. Boliek, C. Christopoulos, and E. Majani, *Information Technology: JPEG2000 Image Coding System*, JPEG 2000 Part I final committee draft version 1.0 ed., ISO/IEC JTC1/SC29 WG1, JPEG 2000, April 2000.

[3] M. W. Marcellin, M. A. Lepley, A. Bilgin, T. J. Flohr, T. T. Chinen, and J. H. Kasner, “An overview of quantizartion of JPEG2000,” *Signal Processing: Image Communication*, vol. 17, no. 1, pp. 73–84, Jan. 2002.

[4] K. T. Mullen, “The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings.” *The Journal of Physiology*, vol. 359, pp. 381–400, February 1985.

[5] X. Otazu, M. Vanrell, and C. Parraga, “Multiresolution wavelet framework models brightness induction effects,” *Vision Research*, vol. 48, pp. 733–751, 2007.

[6] W. A. Pearlman and A. Said, “Image wavelet coding systems: Part II of set partition coding and image wavelet coding systems,” *Foundations and Trends in Signal Processing*, vol. 2, no. 3, pp. 181–246, 2008.