Local Perceptual Weighting in JPEG2000 for Color Images

Jaime Moreno, Xavier Otazu and Maria Vanrell; Centre de Visió per Computador, Computer Science Department, Universitat Autònoma de Barcelona, Edifici O, Campus UAB (Bellaterra), C.P.08193, Barcelona, Spain.

Abstract

The aim of this work is to explain how to apply perceptual concepts to define a perceptual pre-quantizer and to improve JPEG2000 compressor. The approach consists in quantizing wavelet transform coefficients using some of the human visual system behavior properties. Noise is fatal to image compression performance, because it can be both annoying for the observer and consumes excessive bandwidth when the imagery is transmitted. Perceptual pre-quantization reduces unperceivable details and thus improve both visual impression and transmission properties. The comparison between JPEG2000 without and with perceptual pre-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. Perceptual criteria were taken from the CIWaM (Chromatic Induction Wavelet Model).

Introduction

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

Hence it is necessary an advanced model that removes information according to perceptual criteria, preserving the pixels with high perceptual relevance regardless of the numerical information. The Chromatic Induction Wavelet Model presents some perceptual concepts that can be suitable for it. Both CIWaM and JPEG2000 use wavelet transform. CIWaM 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. In other words, JPEG2000 performs a global transformation of wavelet coefficients, while CIWaM performs a local one.

CIWaM attenuates the details that the human visual system is not able to perceive, enhances those that are perceptually relevant and produces an approximation of the image that the brain visual cortex perceives. At long distances, as Figure 3d 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 (for example, some edges).

This paper is organized as follows: Section *JPEG2000 Quantization Overview* specifies quantization and dequantization model used by JPEG2000 for encoding and reconstruction of wavelet coefficients, thereby is described the Dead-zone Uniform Scalar Quantizer and the Global Visual Frequency Weighting for JPEG2000. Section *Chromatic Induction Wavelet Model* describes the Chromatic assimilation/contrast phenomena. In Section *Local Perceptual Weighting*, the proposed method of pre-quantization will be discussed. Experimental results applied for some test images are given in the next section. The last section is where the conclusions and future work will be exposed.

JPEG2000 Quantization Overview

Dead-zone Uniform Scalar Quantizer

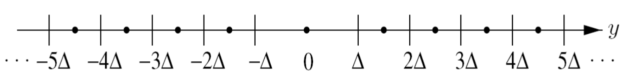
In 2002, Marcellin et. al. in [5] summarize, among others, the uniform scalar quantizer. This quantizer is described as a function that maps each element in a subset of the real line to a particular value, which ensures that more zeros result. In this way all thresholds are uniformly spaced by step size 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.

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

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 quantized index. Figure 1 illustrates such a quantizer with step size Δ.

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

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

The International Organization for Standardization recommends in [2] to adopt the mid-point reconstruction value, setting

$\delta = 0.5$. Experience indicates that some small improvements can be obtained by selecting a slightly smaller value, as Pearlman and Said suggest [9] $\delta = 0.375$, especially for higher frequency subbands. 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.

JPEG2000 Global Visual Frequency Weighting

In JPEG2000, only one set of weights is chosen and applied to wavelet coefficients 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 low frequencies are perceptually more important.

The table 1 specifies a set of weights which was designed for the luminance component based on the Contrast Sensitivity Function (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 weight for *LL* is not included in the table, because it is always 1. Levels 1, 2, … , 5 denote the spatial frequency levels in low to high frequency order with three spatial orientations, *horizontal, vertical* and *diagonal*.

Recommended JPEG2000 frequency weighting for 400 dpi's

s	*horizontal*	*vertical*	*diagonal*
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

Chromatic Induction Wavelet Model

In order to explain the Chromatic assimilation/contrast phenomena as a unique perceptual process, Otazu et al. propose in [7] a low-level Chromatic induction model , which combines three important stimulus properties: spatial frequency, spatial orientation and surround contrast.

Thereby the input image $\mathscr{I}$ is separated into different spatial frequency and orientation components using a multiresolution wavelet decomposition. Thus every single transformed coefficient is weighted using the response of the *extended contrast sensitivity function* (e-CSF, Figure 2), hence a perceptual Chromatic image $\mathscr{I}_\rho$ is recovered. The e-CSF is an extension of the perceptual CSF considering both spatial surround information and observation distance. Particulary the e-CSF value decreases when the surround contrast increases and vice versa.

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

$$\mathscr{I} = \sum_{s=1}^{n} \sum_{o=v,h,d} \omega_s^o + c_n \quad (3)$$

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

The perceptual image $\mathscr{I}_\rho$ recovered from the wavelet planes can be written as:

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

The term $C'(\dot{s}, z_{ctr}(s,o))$ is the e-CSF weighting function, that tries to emulate some perceptual properties of human visual system, as described in [8], has a shape similar to the CSF and can be written as:

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

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:

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

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.

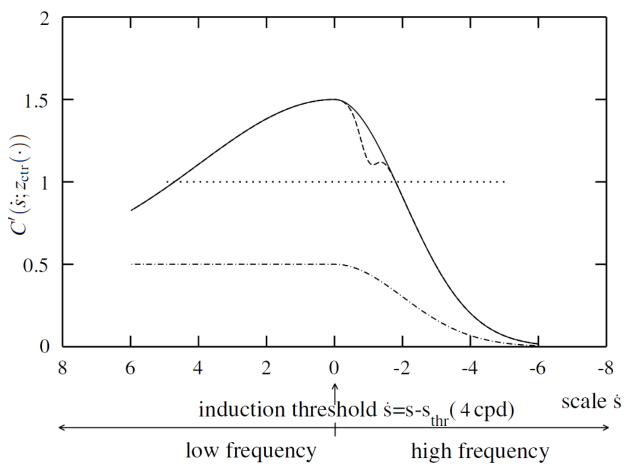
The weighting function $C_d(\dot{s})$ is an approximation to the perceptual CSF, emulates some perceptual properties and is defined as a piecewise Gaussian function [6], such as:

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

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

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

Figure 2. *Continuous function: extended contrast sensitivity function. Dashed function: profile of e-CSF $C'(\dot{s}, z_{ctr}(s,o))$ with $z_{ctr}(x,y;s,o) = 0.75$. Dashed-dotted function: profile of $C_{min}(\dot{s})$. Dotted line: values above this value implies brightness contrast, and values below it implies brightness assimilation.*



taking $\sigma_1 = 2$ and $\sigma_2 = 2\sigma_1$. Both $C_{min}(\dot{s})$ and $C_d(\dot{s})$ depend on the factor s_{thr}, which is the scale associated to 4cpd when an image is observed from a distance d with a pixel size l_p and one visual degree, whose expression is defined by Equation 9.

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

This s_{thr} value is associated to the CSF maximum value. Figure 3 shows three CIWaM images of *Lena*, which are calculated by Equation 4 for a 19 inch monitor with 1280 pixels of horizontal resolution, at 30, 100 and 200 centimeters of distance.

Figure 3. *Perceptual color images of* Lena *performed by CIWaM. The lack of information does not produce compression artifacts, rather it is presented as a softened version.*

(a) Original (b) 30 cm. (c) 100 cm. (d) 200 cm.

Perceptual Local Weighting

In order to compare the JPEG2000 effectiveness and to get each bit-plane, some transformed coefficients of the Original Image or $\mathscr{I}_{org}$ are selected such that $\mathscr{I}_{org} \geq 2^{thr-bpl+1}$, where bpl is the desired bit-plane and thr is the maximum threshold of $\mathscr{I}_{org}$, expressed as follows:

$$thr = \left\lfloor \log_2\left(\max_{(i,j)}\left\{\left|\mathscr{I}_{org\,(i,j)}\right|\right\}\right)\right\rfloor \qquad (10)$$

Figure 4 depicts this process, which is applied for the three components of a opponent color space, i.e. Intensity, Red-Green and Blue-Yellow, thus this selected coefficients are inverse transformed in order to create a new Source of Image Data and to separate the original one in bit-planes.

The modification of JPEG2000 core is illustrated in the block diagram of Figure 5. To obtain wavelet coefficients of $\mathscr{I}$ a Forward Transformation with the 9/7 filter fast wavelet transform is first applied on the source image data. Then, the perceptual quantized coefficients $\mathscr{Q}$, calculated from a known viewing distance d as follows:

$$\mathscr{Q} = \sum_{s=1}^{n}\sum_{o=v,h,d} sign(\omega_s^o)\left\lfloor\frac{\left|C'\left(\dot{s}, z_{ctr}\left(s,o\right)\right)\cdot\omega_s^o\right|}{\Delta_s}\right\rfloor + \left\lfloor\frac{c_n}{\Delta_n}\right\rfloor \qquad (11)$$

This expression is similar to Equation 1, but introduces perceptual criteria to each coefficient contrary to the classical Global Visual Frequency Weighting. A normalized quantization step size $\Delta = 1/128$ is used, namely the range between the minimal and maximal values at $\mathscr{I}_\rho$ is divided into 128 intervals. Finally, the perceptual quantized coefficients are entropy coded, before forming the output code stream or bitstream.

Figure 4. *Bit-plane selection. Some coefficients are selected provided that they fulfil the current threshold.*

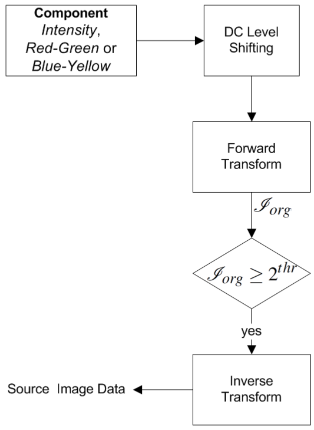
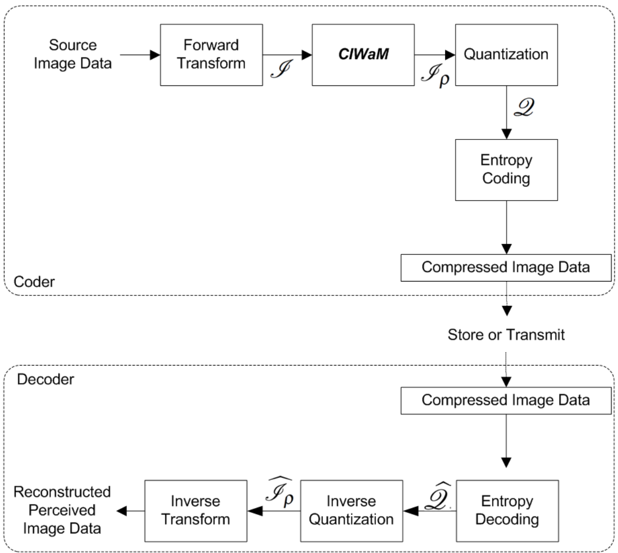


Figure 5. *General block diagram of JPEG2000 compression, applying between Transformation and Quantization steps a perceptual local weighting function into the wavelet coefficients.*



At the decoder, the code stream is, first, entropy decoded in order to reconstruct the perceptual quantized coefficients $\widehat{\mathscr{Q}}$, second, dequantized using Equation 2 with a normalized quantization step size $\Delta = 1/128$ and $\delta = 3/8$. Finally, an inverse discrete wavelet transform is applied to recover $\widehat{\mathscr{I}_\rho}$, thus providing the reconstructed perceived image data.

Experimental Results

The software used to obtain a JPEG2000 compression for the experiments was *JJ2000*, developed by Cannon Research, École Polytechnique Fédérale de Lausanne and Ericsson, available at http://jj2000.epfl.ch.

The Perceptual Local Weighting in JPEG2000 was tested on all the color images of the *Miscellaneous volume* of the University of Southern California Image Data Base, available at http://sipi.usc.edu/database/. The data sets were eight 256×256 pixel images (Figure 6) and eight 512×512 pixel images (Figure 7), but only visual results of the well-known images *Lena*, *F-16* and *Baboon* are depicted, which are 24-bit color images and 512×512 of resolution. The CIWaM images were calculated for a 19 inch monitor with 1280 pixels of horizontal resolution at 50 centimeters of viewing distance.

Figure 6. *Tested 256 x 256 pixel 24-bit Color Images, obtained from the University of Southern California Image Data Base.*

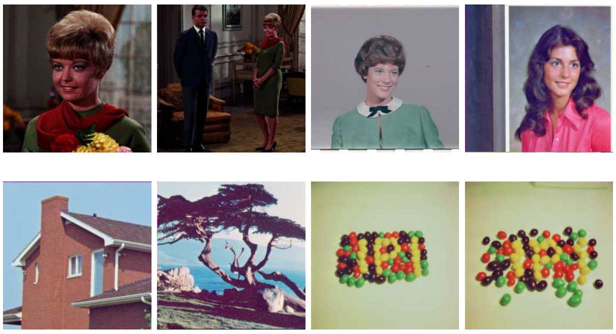
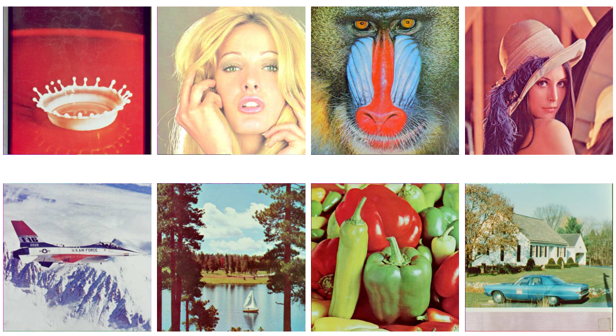


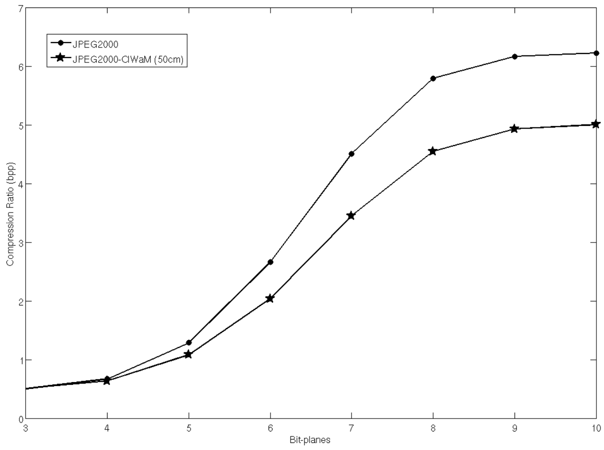
Figure 7. *Tested 512 x 512 pixel 24-bit Color Images, obtained from the University of Southern California Image Data Base.*



The Peak Signal to Noise Ratio or PSNR [4] between the original image $f(i,j)$ and the reconstructed image $\hat{f}(i,j)$ was developed to calculate objectively the degradations introduced into the compression process. PSNR is a function of the Mean Square Error (MSE) and is defined as:

$$PSNR = 10\log_{10}\left(\frac{\mathscr{G}_{max}{}^2}{MSE}\right) \tag{12}$$

Figure 8. *JPEG2000 Compression ratio by Bit-plane. Continuous function with heavy dots: JPEG2000 only quantized by the dead-zone uniform scalar manner. Continuous function with heavy stars: JPEG2000 pre-quantized by the chromatic induction wavelet model, in addition to a dead-zone uniform scalar quantification.*



where $\mathscr{G}_{max}$ is the maximum possible intensity value in $f(i,j)$ ($M \times N$ size) and the MSE has the form:

$$MSE = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\left[f(i,j) - \hat{f}(i,j)\right]^2 \tag{13}$$

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, 3]. The wPSNR increases with increasing variance and vice versa as:

$$wPSNR = 10\log_{10}\left(\frac{\mathscr{G}_{max}{}^2}{wMSE}\right) \tag{14}$$

where the weighted MSE (wMSE) is defined as:

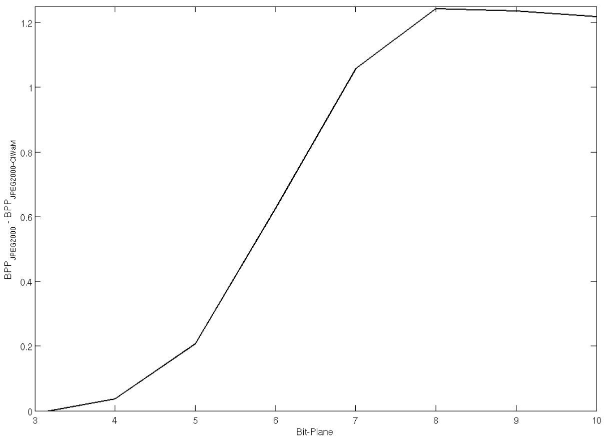
$$wMSE = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\left[\frac{f(i,j) - \hat{f}(i,j)}{1 + Var(i,j)}\right]^2 \tag{15}$$

Figure 8 shows the assessment results of the average performance of color image compression for each bit-plane using a Dead-zone Uniform Scalar Quantizer (continuous function with heavy dots), and adding to it a previous quantization step developed by CIWaM (continuous function with heavy stars).

CIWaM used as a method of pre-quantization, achieves better compression ratios with the same threshold, reaching better results at the highest bit-planes, since CIWaM reduces unperceivable coefficients. Figure 9 shows the contribution of CIWaM in the JPEG2000 compression ratio, for example at the eighth bit-plane, CIWaM diminishes 1.2423 bits per pixel less than without it, namely in a 512×512 pixel color image, CIWaM estimates that 39.75KB of information is perceptually irrelevant at 50 centimeters.

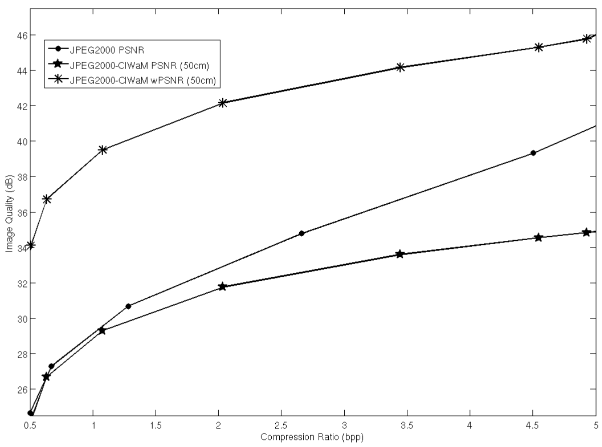
The comparison between compression ratio and image quality is depicted by the Figure 10, which shows that the reconstructed images pre-quantized by CIWaM (continuous function

Figure 9. Contribution of a CIWaM pre-quantification over the JPEG2000 compression ratio by each Bit-plane.



with heavy stars) has less PSNR but higher wPSNR (continuous function with heavy asterisks) than the ones quantized just by a scalar way (continuous function with heavy dots), i.e. even if the reconstructed image has a lower objective quality, this image could have a higher perceptual quality.

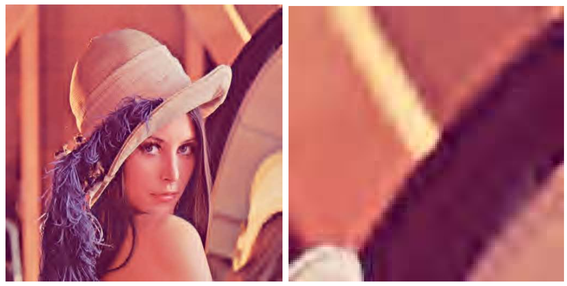
Figure 10. Comparison between compression ratio and image quality. Continuous function with heavy stars: objective quality when a CIWaM pre-quantification and a dead-zone uniform scalar quantification are used. Continuous function with heavy dots: objective quality only quantizing with the dead-zone uniform scalar method. Continuous function with heavy asterisks: subjective quality when jointly a CIWaM pre-quantification and a dead-zone uniform scalar quantification are performed.



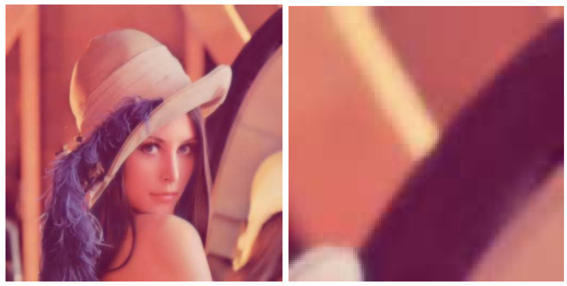
Both Figure 11 and 12 depict examples of reconstructed images compressed at 0.9 and 0.4 bits per pixel, respectively, by means of JPEG2000 without (a) and with perceptual pre-quantization (b). Also this figures demonstrate that the CIWaM subjective quality is higher than the objective one.

The Figure 13 shows examples of reconstructed images of *Baboon* compressed at 0.59, 0.54 and 0.45 bits per pixel by means of JPEG2000 without (a) and with perceptual pre-quantization (b and c). PSNR in 13a is 26.18dB and in 13b is 26.15dB but wPSNR is equal to 34.08 decibels, namely the reconstructed image pre-quantized by CIWaM is perceptually better than the one just quantized by a Scalar Quantizer, since the latter has more compression artifacts, even the result at

Figure 11. Examples of reconstructed images of Lena compressed at 0.9 bpp.



(a) JPEG2000 31.19dB.



(b) JPEG2000-CIWaM 27.57dB.

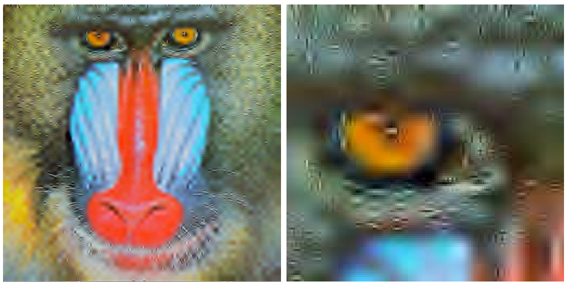
Figure 12. Examples of reconstructed images of F-16 compressed at 0.4 bpp.

(a) JPEG2000 25.12dB.

(b) JPEG2000-CIWaM 24.57dB.

0.45bpp (Figure 13c) has less artifacts, showing for example that the *Baboon*'s eye is softer and better defined and saving additionally 4.48KB of information.

Figure 13. *Examples of reconstructed images of Baboon.*



(a) JPEG2000 compressed at 0.59 bpp.

(b) JPEG2000-CIWaM compressed at 0.54 bpp.

(c) JPEG2000-CIWaM compressed at 0.45 bpp.

Conclusions and Future Work

This work proposes the incorporation of a pre-quantization step to JPEG2000 using CIWaM. 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 CIWaM Quantization improves the compression and image perceptual quality and impacts, on the average,with about 20 per cent . One of the future tasks is the use of a threshold based on the e-CSF properties, namely a threshold based on the perceptual importance of a coefficient, 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] R. Cucchiara, C. Granay, and A. Prati, "Semantic video transcoding using classes of relevance," *International Journal of Image and Graphics*, vol. 3, no. 1, pp. 145–169, January 2003.

[4] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electronics Letters*, vol. 44, no. 13, pp. 800–801, 2008.

[5] 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.

[6] 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.

[7] X. Otazu, C. A. Parraga, and M. Vanrell, "A unified model for chromatic induction," 2009, under review in Journal of Vision.

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

[9] 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.

Author Biographies

Jaime Moreno received his BSc degree in Telecommunications and Electronics Engineering in 2002 from the Instituto Politécnico Nacional, Mexico and his MSc degree in Advanced Informatics in 2008 from the Universitat Autònoma de Barcelona, Spain. Currently, he is a Phd. Student in the Computer Science Department and a researcher in the Computer Vision Center, where he is developing his PhD Thesis in perceptual criteria applied in colour image compression.

Xavier Otazu received the Ph.D. degree in physics from the Universitat de Barcelona, in 2001. He is also currently a Researcher at the Computer Vision Center and a Lecturer in the Computer Science Department, Universitat Autònoma de Barcelona, working in color and texture image analysis, especially in the study of the human color visual system and the mathematical description of its perceptual processes using wavelets.

Maria Vanrell is an Associate Professor in the Computer Science Department of the Universitat Autònoma de Barcelona and is attached to the Computer Vision Center as a researcher. She received her Phd in Computer Science from the Unviversitat Autònoma de Barcelona in 1996. Her research interest is mainly focused in colour and texture in computer vision problems, including colour constancy, texture description and colour and texture grouping.