LOW-COMPLEXITY PERCEPTUAL JPEG2000 ENCODER FOR AERIAL IMAGES

Graduate Students: Han Oh[1] and Yookyung Kim[1]
Faculty Advisors: Michael W. Marcellin[1], Ali Bilgin[2,1], and Francesc Aulí-Llinàs[3]
[1]Department of Electrical and Computer Engineering, [2]Biomedical Engineering, The University of Arizona, Tucson, AZ, USA
[3]Department of Information and Communications Engineering, Universitat Autònoma de Barcelona, Barcelona, Spain

ABSTRACT

A highly compressed image inevitably has visible compression artifacts. To minimize these artifacts, many compression algorithms exploit the varying sensitivity of the human visual system (HVS) to different frequencies. However, this sensitivity has typically been measured at the near-threshold level where distortion is just noticeable. Thus, it is unclear that the same sensitivity applies at the supra-threshold level where distortion is highly visible. In this paper, we measure the sensitivity of the HVS for several supra-threshold distortion levels based on our JPEG2000 distortion model. Then, a low-complexity JPEG2000 encoder using the measured sensitivity is described. For aerial images, the proposed encoder significantly reduces encoding time while maintaining superior visual quality compared with a conventional JPEG2000 encoder.

Keywords: JPEG2000, human visual system, low-complexity.

1. INTRODUCTION

Transmission of an aerial image of a large size via a bandwidth-constrained channel often requires a high compression ratio. An image encoded with such a high compression ratio inevitably has visible quality degradation induced by compression. Many image compression experts have worked to minimize this quality degradation. JPEG2000, a wavelet-based compression standard, provides more powerful compression performance at low bitrates compared to JPEG, but the image quality degradation from blurring or ringing artifacts that occurs in JPEG2000 still hinders the interpretation of aerial images that contain a particularly large number of high-frequency details.

Many conventional JPEG2000 implementations optimize the codestream to minimize mean-squared error (MSE) at a given bitrate [1]. However, this optimization may involve high computational complexity as well as high memory usage. Furthermore, because the MSE relies only on the numerical difference between the original image and the reconstructed images rather than on the perceptual difference of human observers, lower MSEs do not always mean higher image quality [2]. To achieve better visual quality at the same bitrates, the encoder should take into account the properties of the human visual system (HVS).

The human eye has different sensitivities to different spatial frequencies. The varying sensitivity to different frequencies, represented by the contrast sensitivity function (CSF), is exploited in many image quality assessment metrics and encoders and contributes to improving image quality. However, the sensitivity is typically measured at the near-threshold level where distortion is just noticeable. Thus, it is unclear that the same sensitivity applies at the supra-threshold level where distortion is highly visible [3].

In this paper, we measure the sensitivities for several supra-threshold distortion levels through psychophysical experiments based on a JPEG2000 quantization distortion model. Then, using the sensitivities, we propose a JPEG2000 encoder that produces superior visual quality compared with conventional JPEG2000 encoders at the same bitrate. The proposed encoder encodes images faster than conventional encoders since it can skip the computationally intensive rate-distortion optimization stage. All resulting codestreams are JPEG2000 Part-I compliant, so any JPEG2000 decoder can decode the codestream without any modification.

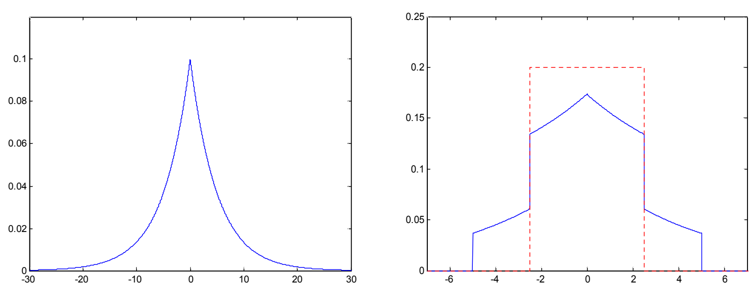
2. SUPRA-THRESHOLD QUANTIZATION DISTORTION OF JPEG2000

The dyadic wavelet transform employed in JPEG2000 has several advantages in terms of perceptual image compression. As shown in Fig. 1, a highly compressed image using the wavelet transform yields perceptually more pleasing quality without blocking artifacts compared to an image compressed using the common block-based discrete cosine transform (DCT) at the same bitrate. The wavelet transform decomposes an image into logarithmically spaced spatial frequencies and four orientations of 0°, 90°, 45° and 135°. This decomposition is similar to the decomposition process performed by the primary visual cortex (V1) of the HVS. Compared with other cortical transforms designed to mimic the V1, the dyadic wavelet transform is computationally much more efficient at applying the varying sensitivity values to the subband. In addition, because the wavelet transform retains spatial information as well as frequency information, it facilitates the application of masking models calculated from background images [4]. Also, the wavelet transform inherently supports multi-resolution decoding, so it is possible to design an encoder that encodes images adaptively depending on the display resolution [5].

In an error-free communication channel, visual quality degradation of an image encoded using JPEG2000 is due to information loss that occurs by quantizing wavelet coefficients. Wavelet coefficients that statistically have a generalized Gaussian distribution are first quantized with the deadzone quantizer of an initial quantization step size. Then, the quantization indices

(a) (b)

Figure 1. Images encoded with (a) JPEG2000 and (b) JPEG at the same bitrate (0.350 bits-per-pixel (bbp)). The JPEG image has the perceptually more distracting blocking artifacts.



(a) (b)

Figure 2. Models for probability density functions for: (a) wavelet coefficients in HL, LH, and HH subbands ($\sigma^2 = 50$); (b) quantization distortions in HL, LH, and HH subbands ($\sigma^2 = 50$, $\Delta = 5$). Dashed lines represent the commonly assumed uniform distribution.

are bit-plane coded. The effective quantization step size in subband $\mathbf{b}$, denoted as $\Delta_{\mathbf{b}}$, is determined by the initial quantization step size and the number of bit-planes to be included in the final codestream. The effective quantization step size produces quantization distortion with the distribution shown in Fig. 2 (b). When the quantization distortion is below the visibility threshold (VT) measured through psychophysical experiments, the quantization distortion is referred to as sub-threshold distortion, and the resulting image has visually lossless quality [6].

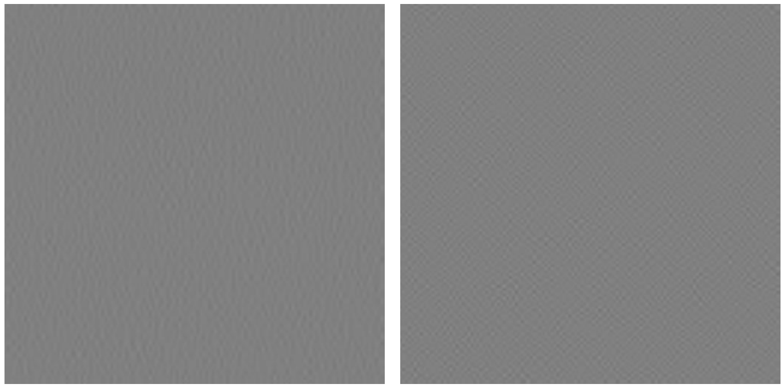
For supra-threshold quantization distortion in which the distortion is larger than the VT, many perceptual compression algorithms apply the contrast sensitivity function (CSF) obtained from the near-threshold level, assuming that the CSF still holds at the supra-threshold distortion level (e.g., [7]). Although these algorithms are reported to improve compression performance, it is unclear whether quantization step sizes based on the CSF are optimal at the supra-threshold distortion level. Contrary to the assumption made about the near-threshold level CSF, [8] claims, through a contrast-matching experiment using a sinusoidal grating, that the perceived contrast of supra-threshold distortion is less dependent on spatial frequency, and that the sensitivity profile is much flatter than that obtained at the near-threshold distortion level. This is called contrast constancy and it has been further supported by other experiments using both Gabor patches and broadband noise [9]. However, [10] uses wavelet data to show that images quantized based on contrast constancy are worse than images quantized with the contrast sensitivity of the near-threshold distortion level. This result was attributed to the global precedence effect (i.e., visual integration of structural information in a coarse-to-fine-scale fashion). A CSF was suggested for the supra-threshold distortion level that is similar to the CSF of the near-threshold distortion level. However, this CSF is obtained using a uniform quantizer, so it is also unclear whether this CSF is accurate for JPEG2000 and other codecs that use a deadzone quantizer. In the following section, we describe a method of measuring the sensitivity for supra-threshold distortion levels using a JPEG2000 quantization distortion model.

3. SENSITIVITY TO THE SUPRA-THRESHOLD QUANTIZATION DISTORTION IN JPEG2000

3.1. PERCEPTUAL DISTORTION MATCHING EXPERIMENT

To compare the perceptual differences in the supra-threshold quantization distortion between different subbands, a pair of images is displayed simultaneously, as shown in Fig. 3. Each image (stimulus) is a gray image obtained by applying the inverse wavelet transform to wavelet data containing quantization distortions. In this work, stimulus generation begins with wavelet subband data corresponding to a 512×512 gray image. These initial data have all coefficients of all subbands set to 0. Quantization distortion is randomly generated in a subband of interest according to the quantization distortion model of JPEG2000, as shown in Fig. 2 (b) [6]. This distortion is added to the entire spatial extent in the selected subband. The left image is the reference image, and its quantization step size is explicitly specified and is unchanged during the experiment, while the quantization step size of the right image begins with an arbitrary value and changes based on the responses of a human viewer (subject). In this work, the variances of the wavelet subbands for HL, LH and HH is assumed to be 50, and the variance for LL is assumed to be 2000.

The comparison of the two distorted images is done using a two-alternative forced-choice (2AFC) method. The subject is asked to select the perceptually more distorted image. The two images are displayed for an unlimited amount of time until the subject selects an image. For each trial, the quantization step size of the right image is adaptively adjusted by the QUEST staircase procedure in the Psychophysics Toolbox [11] so that the two images have similar perceptual distortion levels. Through 24 repeated trials, the quantization step size of the right image that perceptually matches the quantization step size of the left image is determined.



(a) (b)

Figure 3. An example of perceptual distortion matching experiment. The left image and right image are obtained by applying the inverse wavelet transform to wavelet data containing quantization distortions. The quantization distortions of the left and right image are generated with (a) a quantization step size of 3.5 in (HL,3) subband and (b) a quantization step size of 4.5 in (HH,3) subband, respectively.

The determined quantization step size is then used to generate the reference image (left image) in the next experiment. To facilitate comparison between two images, HH subbands are compared with HL subbands at the same wavelet transform level. The HL subband of wavelet transform level k is compared with the HL subband of wavelet transform level $k+1$ or $k-1$. LH subbands are assumed to require the same thresholds as HL subbands, so only HL subbands are measured. Also, to minimize the propagation of measurement errors, the experiment begins with the center frequency subband, (HL,3). The quantization step size of the (HL,3) subband determines the supra-threshold distortion level. In this work, the quantization step size of the (HL,3) subband is sequentially set to different multiples of VT: q times the VT. For example, the VT (the quantization step size where distortion remains invisible) is approximately 0.5 when the variance of the wavelet coefficients is 50. In this supra-threshold distortion matching experiment, the quantization step size of the (HL,3) subband is respectively set to 1.0 ($q = 2$), 1.5 ($q = 3$), 2.0 ($q = 4$), and so on, to find a set of quantization step sizes. The set of quantization step sizes that provide similar perceptual distortion at a given q is denoted here as q-VT.

The experimental environment is arranged similarly to a typical office environment. Stimuli are displayed on a Dell U2410 24-in In-Plane Switching (IPS) panel LCD monitor in ambient light. The monitor has a dot pitch of 0.27 mm, a resolution of 1920 $\times$ 1200, an image brightness of approximately 70 cd/m^2 at a mid-gray level, a contrast ratio of 1000:1, and is connected to the PC through a Digital Visual Interface (DVI) cable. The viewing distance is 60 cm (23.6 inches) with a resulting visual resolution of 38.72 pixels/degree. The center spatial frequencies for the five wavelet transform levels are 19.36, 9.68, 4.84, 2.42, and 1.21 cycles/degree. Two subjects, who are familiar with wavelet quantization distortion and who have normal visual acuity, conducted the experiments.

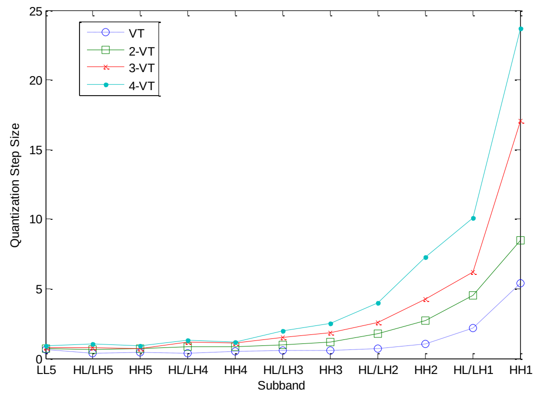


Figure 4. Perceptually equivalent quantization step sizes measured through the psychophysical experiments.

orientation	$a_{1,\theta}$	$a_{2,\theta}$	$a_{3,\theta}$	$b_{1,\theta}$	$b_{2,\theta}$	$b_{3,\theta}$
HL/LH	-0.188	3.920	-2.562	0.010	-0.117	-1.107
HH	-0.919	13.200	-14.270	0.010	-0.166	-1.380

Table 1. Fitted quadratic parameters.

3.2. EXPERIMENTAL RESULTS

Figure 4 shows perceptually equivalent quantization step sizes for three supra-threshold quantization distortion levels, together with the VT. Similar to the VT, the quantization step sizes become larger as the spatial frequency of the subband increases. Nevertheless, when the quantization step size for the (HL,3) subband increases by a multiple of q, the quantization step sizes of the other subbands do not increase by the same multiple. The quantization step sizes for low-frequency subbands slightly increase, whereas the quantization step sizes for high-frequency subbands significantly increase.

Estimates of the quantization step sizes for supra-threshold distortion levels for values of the multiple q other than those tested can be obtained with the following power functions fitted to the measured quantization step sizes using the least-squares method.

$$\Delta_{\theta,k}(q) = (a_{1,\theta}q^2 + a_{2,\theta}q + a_{3,\theta})k^{(b_{1,\theta}q^2 + b_{2,\theta}q + b_{3,\theta})} \quad \text{for } q > 1 \tag{1}$$

where $k \in \{1, 2, ..., 5\}$ is the wavelet transform level, and $\theta \in \{HL/LH, HH\}$ is the orientation of the subband. The fitted quadratic parameters are listed in Table 1. Figure 5 shows the quantization step sizes measured during the experiments and the fitted function. For the (LL,5) subband, $\Delta_{LL,5}(q) = 0.0236q^2 - 0.0117q + 0.6196$.

4. LOW-COMPLEXITY PERCEPTUAL JPEG2000 ENCODER AND ENCODING RESULTS

In a typical JPEG2000 implementation, the wavelet coefficients in subband $\mathbf{b} = (\theta, k)$ are initially quantized with a small quantization step size $\Delta_{\mathbf{b}}$. The subband is partitioned

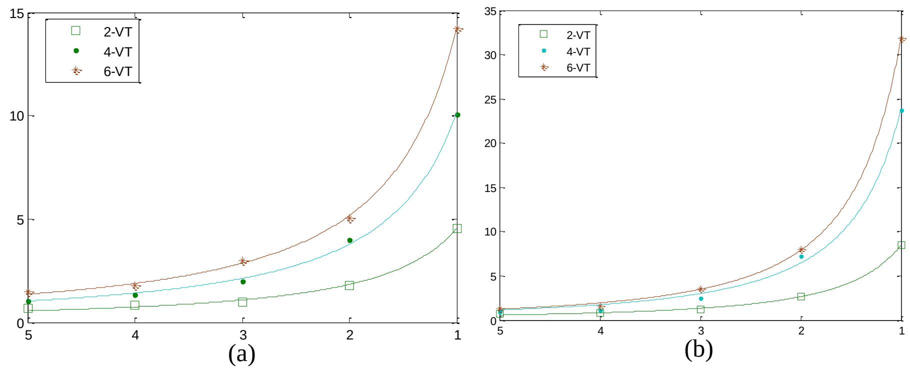


Figure 5. Quantization step sizes measured during the experiments (symbols) and the fitted function (lines) for (a) HL/LH subbands and (b) HH subbands.

into codeblocks, and the quantization indices of each codeblock are bit-plane coded. Each bit-plane is coded in three coding passes, except for the most significant bit-plane (MSB), which is coded in one coding pass. Thus, a codeblock with M bit-planes has $3M - 2$ coding passes. Common rate-distortion optimization methods compute the rate increase and distortion reduction for each coding pass, and select coding passes included in the final codestream as those with highest distortion-length slopes as formulated in [1]. The main penalty imposed by conventional rate-distortion optimization methods is that they require the encoding of coding passes that will not be included in the final codestream. This can be partially addressed by estimation techniques [12], though they may require to encode some more coding passes than those strictly required, or they may produce a degradation on quality. This paper proposes the use of Eq. (1) to determine the quantization step sizes so that all coding passes are included. Image quality in this coding method is controlled by choosing a target quality q rather than a target bitrate. Because no extra coding passes are computed and the PCRD-opt stage is skipped, this coding method can dramatically reduce the encoding time.

Table 2 compares the encoding times for an 8-bit grayscale aerial image of size 14580×14565 at several bitrates for the proposed method, for a conventional JPEG2000 implementation that encodes all coding passes, and for a JPEG2000 implementation that uses heuristics based on already encoded codeblocks to reduce the amount of the aforementioned over-coding [13]. The encoding times were measured using Kakadu v.5.1 [13] on a PC equipped with an Intel Core 2 Duo E7400 2.8 GHz CPU and 4 GB of RAM. The target bitrates for the conventional and heuristic JPEG2000 encoders were chosen to match those from the bitrates of the images encoded by the proposed method. The conventional JPEG2000 encoder requires 21.71 seconds on average to encode the image with little variation according to the target bitrate. The use of heuristics reduces the encoding time by approximately a factor of 2 on average. The encoding time of the proposed method reduces the encoding time by an additional factor of 2 on average. The savings of the heuristic and proposed methods increase with decreasing bitrate.

q	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0
bitrate (bpp)	0.477	0.363	0.303	0.263	0.235	0.214	0.198	0.184	0.173
proposed (sec)	7.401	6.145	5.742	5.561	5.438	5.084	4.999	4.894	4.829
conventional (sec)	23.474	22.515	22.972	21.231	21.752	21.450	21.151	20.298	20.545
heuristic (sec)	13.979	12.184	11.754	10.521	10.487	10.178	9.948	9.175	9.142

Table 2. Encoding times for an 8-bit grayscale aerial image of size 14580×14565 at several bitrates for the proposed, conventional, and heuristic methods.

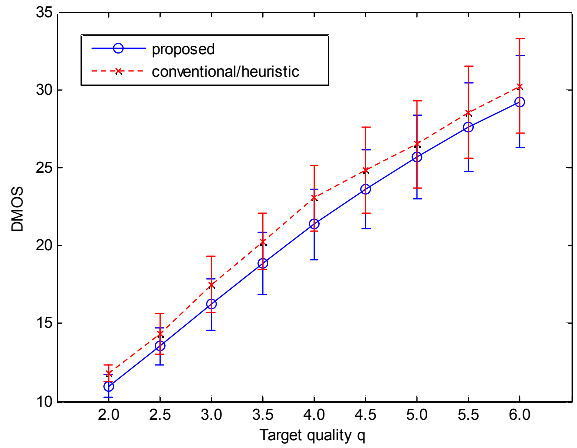


Figure 6. Average DMOSs converted from VIF scores. The points are the average DMOSs and the error bars represent $\pm$ 1 standard deviation. Smaller values indicate higher visual quality.

To examine the visual quality of the three methods at the same bitrate, 10 test images of size 512×512 were used. The test images were created by cropping the 14580×14565 aerial image that had been down-scaled by a variety of factors. Figure 6 shows the average difference mean-opinion-scores (DMOSs) converted from the Visual Information Fidelity (VIF) scores [14]. The VIF metric has a very high correlation with human subject ratings and outperforms state-of-the-art objective metrics such as SSIM, Sarnoff, and VSNR [15]. In the figure, the points are the average DMOSs and the error bars represent $\pm$ 1 standard deviation. The DMOS is scaled from 0 to 100, and a lower DMOS represents image quality closer to the original image. The DMOS scores of the proposed method lie well below those of the conventional and heuristic JPEG2000 encoders at all bitrates, which suggests that the proposed method provides better quality despite the significantly reduced encoding time. Two encoded images are shown in Fig. 7. Though the conventional JPEG2000 encoder yields the optimal codestream in terms of the MSE, images encoded with the proposed method exhibit less blurring and superior visual quality than those encoded with the conventional JPEG2000 encoder. In particular, at low bitrates, the proposed method better preserves details compared with the conventional JPEG2000 encoder.

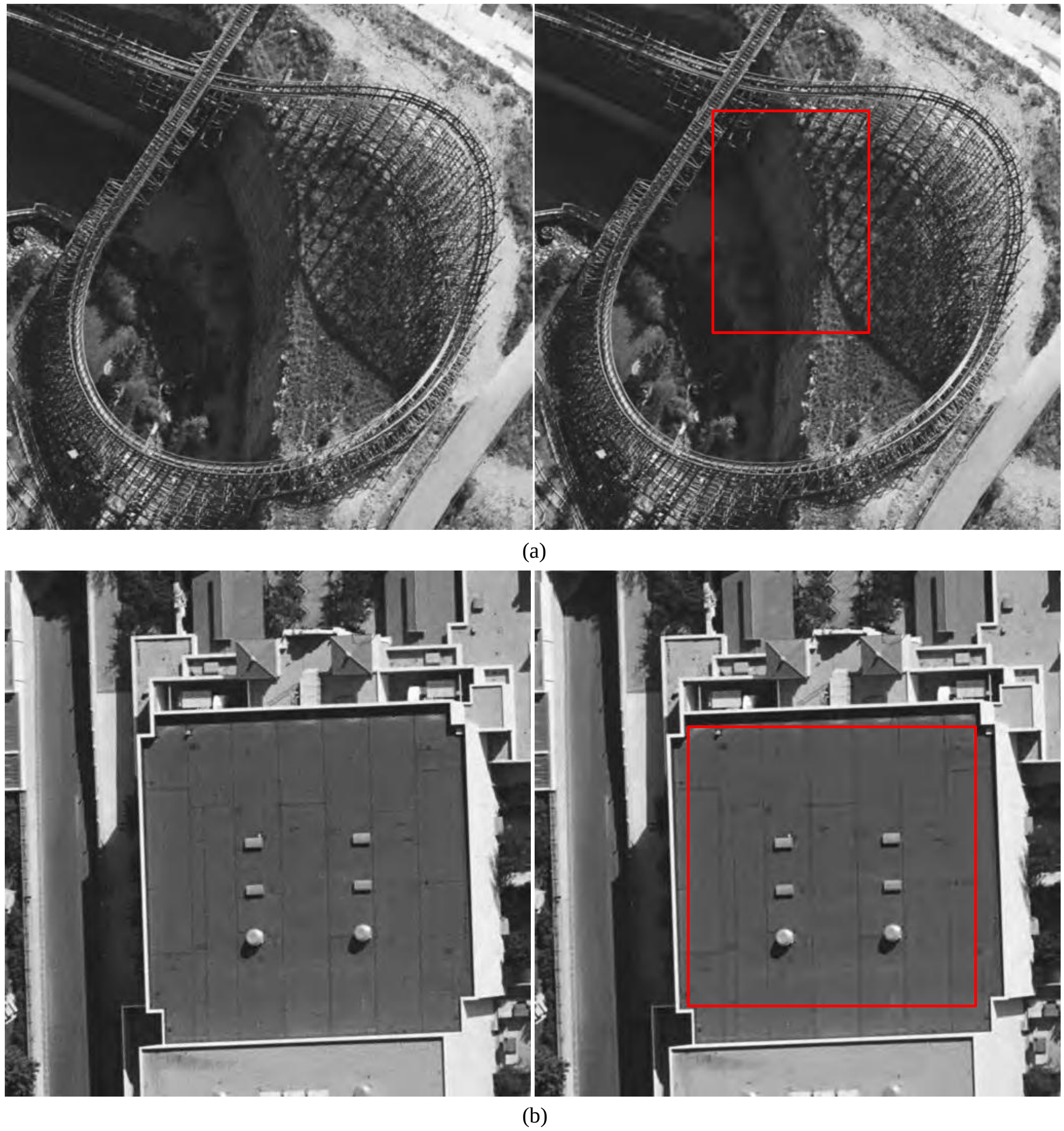


Figure 7. Images encoded with the proposed method (left images) and the conventional JPEG2000 encoder at the same bitrates (right images). The target quality for the proposed method is set to (a) $q = 3.5$ and (b) $q = 4.5$, respectively. The left images exhibit less blurring and superior visual quality than the right images as emphasized by the boxes.

5. CONCLUSIONS

This paper introduces a low-complexity perceptual JPEG2000 encoder for aerial images. The encoder is based on quantization step sizes obtained from perceptual distortion matching experiments. This quality-driven encoder significantly reduces the encoding time while maintaining higher visual quality at the same bitrate, as compared with a conventional JPEG2000 encoder. Furthermore, the resulting codestream is JPEG2000 Part-I compliant, so is decodable by any JPEG2000 decoder.

While the work presented in this paper uses only a fixed quantization step size for a subband, assuming that the variance of wavelet coefficients is fixed, an adaptive quantization scheme that varies with the variance of the wavelet coefficients is expected to provide consistent quality control and better visual quality, at the expense of slightly increased complexity. This is the subject of ongoing research.

6. ACKNOWLEDGEMENTS

This work has been partially supported by the European Union under Grant FP7-PEOPLE-2009-IIF FP7-250420, and by the Spanish Government under Grant RYC-2010-05671.

REFERENCES

1. Taubman, D. S. and Marcellin, M. W., JPEG2000: image compression fundamentals, standards, and practice, Kluwer Academic Publishers, Boston 2002.
2. Wu, H. R. and Rao, K. R., Digital Video Image Quality and Perceptual Coding, CRC 2005.
3. Chandler, D. M. and Hemami, S. S., "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," IEEE Transactions on Image Processing **16**(9), 2284–2298 2007.
4. Nadenau, M. J., Reichel, J., and Kunt, M., "Wavelet-based color image compression: exploiting the contrast sensitivity function," IEEE Transactions on Image Processing **12**(1), 58–70 2003.
5. Oh, H., Bilgin, A., and Marcellin, M. W., "Multi-resolution visually lossless image compression using JPEG2000," in Proceedings of International Conference on Image Processing, Sept 2010.
6. Oh, H., Bilgin, A., and Marcellin, M. W., "Visibility thresholds for quantization distortion in JPEG2000," in Proceedings of QoMex, July 2009.
7. Liu, Z., Karam, L. J., and Watson, A. B., "JPEG2000 encoding with perceptual distortion control," IEEE Transactions on Image Processing **15**, 1763–1778 2006.
8. Georgeson, M. A. and Sullivan, G. D., "Contrast constancy: Deblurring in human vision by spatial frequency channels," The Journal of Physiology **252**, 627–656 1975.
9. Brady, N. and Field, D. J., "What's constant in contrast constancy? the effects of scaling on the perceived contrast of bandpass patterns," Vision Research **35**(6), 739–756 1995.
10. Chandler, D. M. and Hemami, S. S., "Suprathreshold image compression based on contrast allocation and global precedence," in Proceedings of the SPIE, **5007**, 73–86 2003.
11. Brainard, D. H., "The psychophysics toolbox," Spatial Vision **10**(4), 433–436 1997.
12. Schelkens, P., Skodras, A., and Ebrahimi, T., The JPEG 2000 Suite, Wiley 2009.
13. Kakadu software. `http://www.kakadusoftware.com`.
14. Sheikh, H. R. and Bovik, A. C., "Image information and visual quality," IEEE Transactions on Image Processing **15**(2), 430–444 2006.
15. Sheikh, H. R., Sabir, M. F., and Bovik, A. C., "A statistical evaluation of recent full reference image quality assessment algorithms," IEEE Transactions on Image Processing **15**(11), 3441–3452 2006.