Distortion Estimators for Bitplane Image Coding

Francesc Aulí-Llinàs, *Member, IEEE,* and Michael W. Marcellin, *Fellow, IEEE*

***Abstract*—Bitplane coding is a common strategy used in current image coding systems to perform lossy, or lossy-to-lossless, compression. There exist several studies and applications employing bitplane coding that require estimators to approximate the distortion produced when data are successively coded and transmitted. Such estimators usually assume that coefficients are uniformly distributed in the quantization interval. Even though this assumption simplifies estimation, it does not exactly correspond with the nature of the signal. This work introduces new estimators to approximate the distortion produced by the successive coding of transform coefficients in bitplane image coders, which have been determined through a precise approximation of the coefficients' distribution within the quantization intervals. Experimental results obtained in three applications suggest that the proposed estimators are able to approximate distortion with very high accuracy, providing a significant improvement over state-of-the-art results.**

***Index Terms*—Distortion estimation, bitplane coding, image coding, rate-distortion optimization, JPEG2000.**

I. Introduction

SINCE the seminal work of Mallat [1] introducing the use of the wavelet transform for the coding of images, and of Shapiro [2] introducing the bitplane coding strategy to successively refine image distortion in wavelet image coding, most of the proposed lossy image coding systems encode wavelet coefficients in a bitplane by bitplane basis. Let $[t_{K-1}, t_{K-2}, ..., t_1, t_0]$ be the binary representation for an integer v which, in the discussion below, denotes the magnitude of the index obtained by quantizing a wavelet coefficient y, and with K denoting a sufficient number of bits to represent all coefficients. Bitplane coding strategies generally define a bitplane j as the same bit t_j from all coefficients, and encode the image from the most significant bitplane $K-1$ to the lowest bitplane 0. The first non-null bit of a coefficient, i.e., that $t_s = 1$ such that $\nexists\ s' > s$ with $t_{s'} = 1$, is called the *significant bit* of the coefficient. The remaining bits $t_r, r < s$ are called *refinement bits*.

A valuable advantage of bitplane image coding is that it can generate a quality progressive bitstream that can be successively transmitted and decoded at increasing bitrates. When the bitstream is truncated, wavelet coefficients may not be fully transmitted, and thus the decoder carries out a dequantization operation consisting of assigning a reconstruction value $\hat{y}$ that lies somewhere in the corresponding quantization interval. If t_P denotes the last transmitted bit of v, in the case of a deadzone scalar quantization with step size Φ, the reconstruction procedure is expressed as

$$\hat{y} = \begin{cases} 0 & \text{if } P > s \\ (\hat{v} + \delta) \cdot \Phi 2^P & \text{if } P \leq s \end{cases}, \quad (1)$$

where $\hat{v} = [t_{K-1}, t_{K-2}, ..., t_P]$, and $\delta \in [0, 1)$ adjusts the reconstruction value $\hat{y}$ within the quantization interval $[\Phi 2^P, \Phi 2^{P+1})$.

The most common approach to carry out the reconstruction procedure is to set coefficient $\hat{y}$ to the center of the quantization interval, for both significance and refinement coding, which corresponds to setting $\delta = 1/2$. Mid-point reconstruction satisfies the minimax criterion and minimizes mean squared error if the wavelet coefficients are uniformly distributed within the quantization interval. It has been suggested [3, Ch. 10.5.1], [4] that other δ parameters might better represent the signal, and that the distribution of wavelet coefficients can be modeled by a generalized Laplacian distribution [5]. However, this issue has not been carefully addressed in the context of bitplane coding, and it has implications that go beyond the reconstruction procedure carried out in the decoder.

Some encoders, for instance, employ approximations of the image distortion to optimize the construction of the codestream [6], while other applications need to estimate the distortion of the image after coding [7]. When the distortion metric is Mean Squared Error (MSE), image distortion is approximated as $D = \sum_k [G_b \cdot (y[k] - \hat{y}[k])^2]$, with $y[k]$ and $\hat{y}[k]$ respectively denoting transform coefficients and their quantized representation after transmission, and with G_b denoting the energy gain factor of subband b to which the coefficients belong. When the original coefficients $y[k]$ are not available, or when computational resources are restricted, distortion may be estimated rather than actually computed. One strategy is to estimate the initial squared error and the squared error decreases that can be expected from coding significance and refinement bits. Such an approach was proposed by Li and Lei [8] to determine a rate-distortion optimized scanning order that could be implicitly followed by both the encoder and the decoder. In [9], an estimator was used to adapt the Set Partitioning In Hierarchical Trees (SPIHT) [10] to memory constrained environments, and in [4] image distortion was estimated to protect real-time image and video transmissions. Within the framework of JPEG2000, the a priori computation of distortion-rate slopes using expected distortion decreases helps to reduce the computational load of the encoder [11], and to model the rate-distortion characteristics of codeblocks

The authors are with the Department of Electrical and Computer Engineering, University of Arizona, Tucson, AZ, USA. (phone: +34 935813571; fax: +34 935814477; e-mail: fauli@deic.uab.es, marcellin@ece.arizona.edu).

A preliminary version of this work was presented in the IEEE Data Compression Conference 2009. This work has been partially supported by the Spanish Government (MICINN), by FEDER, and by the Catalan Government, under Grants MEC_BPD-2007-1040, TSI2006-14005-C02-01, and SGR2005-00319.

for transcoding purposes [12]. Several other studies and applications have employed mid-point reconstruction when decoding the image, estimating distortion, or optimizing coding procedures.

The main purpose of the research described here is to accurately determine the squared error and squared error decrease expected from significance and refinement bits considering a probability distribution model that better captures the nature of the signal. We show that, through the proposed estimators, distortion can be approximated with very high accuracy. This has several practical applications: rate-distortion models can be better adjusted, the computational complexity of implementations can be reduced, and transcoding procedures can obtain essentially the same performance whether or not the original image is available. We restrict our attention to three scenarios in which the accuracy of estimators is evaluated for the 9/7 discrete wavelet transform (DWT), and the 5/3 integer wavelet transform (IWT). Our conclusions can be easily extended to other applications and wavelet-based approaches.

This paper is structured as follows: Section II formulates the expected squared error and squared error decrease per coded bit in a general manner, and proposes the use of an accurate probability distribution conceived from experimental evidence; Section III describes three applications in which the estimators can be employed to approximate image distortion; and Section IV assesses the performance of the proposed approach through experimental results. The last section draws conclusions.

II. Expected squared error per coded bit

A. General formulation

For notational simplicity, we assume that coefficients are normalized by the quantization step size Φ in what follows. Let us first consider the squared error produced by those coefficients that become significant at bitplane P^*. If x denotes the magnitude of such a coefficient, i.e., $2^{P^*} \leq x < 2^{P^*+1}$, and $p(x)$ denotes the conditional probability density function for x, the initial squared error $\mathcal{D}^{sig}_{P^*}$ of such coefficients can be determined as

$$\mathcal{D}^{sig}_{P^*} = \int_{2^{P^*}}^{2^{P^*+1}} p(x)\ x^2\ dx. \tag{2}$$

Following the same notation, we determine the squared error decrease that can be expected when such coefficients are encoded to bitplane P^* as

$$\triangle\mathcal{D}^{sig}_{P^*} = \int_{2^{P^*}}^{2^{P^*+1}} p(x) \left[x^2 - \left(x - (2^{P^*} + \delta_{P^*} \cdot 2^{P^*})\right)^2\right]\ dx, \tag{3}$$

where δ_{P^*} stands for the reconstruction factor δ used at bitplane P^*. The distortion decrease in this expression is determined as the squared error before coding bit P^*, minus the squared error after coding bit P^*. Since coefficients that become significant in bitplane P^* are recovered as zero for previous bitplanes, $\triangle\mathcal{D}^{sig}_{P} = 0, P > P^*$.

When assuming a uniform probability distribution within the significance interval, $p(x) = 1/2^{P^*}$ and the δ_{P^*} value that maximizes the average distortion decrease is at the center of the interval, i.e., $\delta_{P^*} = 1/2$. Substituting $\delta_{P^*} = 1/2$ and $p(x) = 1/2^{P^*}$ in (2) and (3) results in $\mathcal{D}^{sig}_{P^*} = 7/3 \cdot 2^{2P^*}$ and $\triangle\mathcal{D}^{sig}_{P^*} = 9/4 \cdot 2^{2P^*}$, which corresponds to [8], [9], and [4].

The squared error after the significance bit is transmitted is

$$\mathcal{D}^{ref}_{P^*-1} = \int_{2^{P^*}}^{2^{P^*+1}} p(x) \left(x - (2^{P^*} + \delta_{P^*} \cdot 2^{P^*})\right)^2\ dx, \tag{4}$$

and the squared error decrease expected for refinement coding at bitplane P^*-1 is determined according to

$$\begin{aligned}\triangle\mathcal{D}^{ref}_{P^*-1} = \int_{2^{P^*}}^{1.5\cdot 2^{P^*}} p(x) \Big[\left(x - (2^{P^*} + \delta_{P^*} \cdot 2^{P^*})\right)^2 - \\ \left(x - (2^{P^*} + \delta_{P^*-1} \cdot 2^{P^*-1})\right)^2\Big]\ dx\ + \\ \int_{1.5\cdot 2^{P^*}}^{2^{P^*+1}} p(x) \Big[\left(x - (2^{P^*} + \delta_{P^*} \cdot 2^{P^*})\right)^2 - \\ \left(x - (2^{P^*} + 2^{P^*-1} + \delta_{P^*-1} \cdot 2^{P^*-1})\right)^2\Big]\ dx\ ,\end{aligned} \tag{5}$$

where the first and the second integrals denote the average distortion decrease when the refinement bit is 0 and 1, respectively. Note that with some abuse of notation, the densities may be different in the two integrals. In each integral, the distortion decrease is determined as the squared error after coding the significant bit minus the squared error once the refinement bit is transmitted.

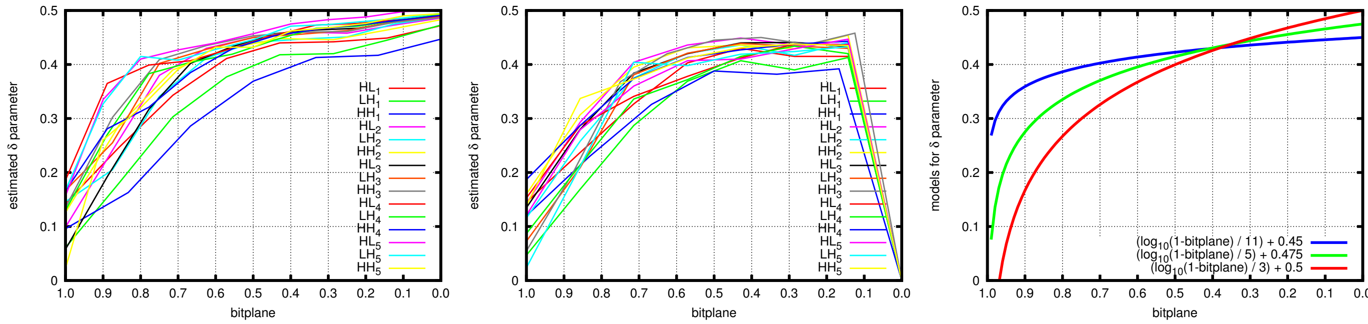
When mid-point reconstruction is used and a uniform distribution is assumed, expressions (4) and (5) simplify to $\mathcal{D}^{ref}_{P^*-1} = 1/12 \cdot 2^{2P^*}$ and $\triangle\mathcal{D}^{ref}_{P^*-1} = 1/16 \cdot 2^{2P^*}$. More generally, $\mathcal{D}^{ref}_{P} = 1/3 \cdot 2^{2P}$ and $\triangle\mathcal{D}^{ref}_{P} = 1/4 \cdot 2^{2P}$, $P < P^*$, which also corresponds to [8], [9], and [4].

In addition to considering $\mathcal{D}^{sig}_{P}$ and $\mathcal{D}^{ref}_{P}$ when estimating the image distortion (see below), it is worth considering the distortion produced by coefficients with null contributions for the given quantizer step size, i.e., the distortion caused by coefficients $x \in [0, 1)$. Such distortion is denoted as $\mathcal{D}^{null}$ in this work, and is determined according to $\mathcal{D}^{null} = \int_0^1 p(x)\ x^2 dx$. In the case of the uniform distribution, this simplifies to $\mathcal{D}^{null} = 1/3$.

B. Determination of the probability distribution

Assuming a uniform probability distribution within the quantization intervals enormously simplifies the expressions above. However, the parameters δ_P and $p(x)$ should ideally be chosen to reflect the actual nature of the signal.

The primary objective when reconstructing coefficients is to minimize the average distortion between the dequantized coefficients and the original ones. This is ideally achieved



(a) Experimental results (9/7 DWT) (b) Experimental results (5/3 IWT) (c) Models

Fig. 1: Experimental results in (a),(b) depict the estimated $\delta_{P,b}$ parameter for the "Portrait" image (gray scale, size 2048×2560, 5 wavelet decomposition levels) when using the DWT and IWT respectively. (c) depicts the resemblance to exponential functions. Horizontal axes represents the bitplane number normalized to $[1, 0]$.

when, for instance, coefficients that have to be recovered within the quantization interval $[2^P, 2^{P+1})$ are reconstructed as the centroid of this interval. If $\#C$ stands for the number of coefficients that have non-null contributions just after bitplane P is transmitted, i.e., those $x \geq 2^P$, the centroid $\mathcal{I}_{P,b}$ of subband b at bitplane P can be estimated as

$$\mathcal{I}_{P,b} = \frac{1}{\#C} \sum_{x \in b} \left(x - 2^P \cdot \left\lfloor \frac{x}{2^P} \right\rfloor \right) \;\forall\; x \geq 2^P, \tag{6}$$

where $\lfloor \cdot \rfloor$ denotes the floor operation. Note that no distinction between significant and refinement coefficients is made since experimental evidence suggests that no substantial gains are achieved in doing so.

We have estimated the centroids for the "Portrait" image of the ISO 12640-1 corpus using 5 levels of the 9/7 DWT. To ease the visual interpretation, instead of depicting $\mathcal{I}_{P,b}$ for each subband and bitplane, Figure 1(a) depicts the corresponding values of the reconstruction factor, calculated as $\delta_{P,b} = \mathcal{I}_{P,b}/2^P$, with the bitplane number P normalized to the nominal range, i.e., we plot P/K_b on the horizontal axis, K_b denoting the number of bitplanes needed to represent the coefficients in subband b. Except for the lowest frequency subband (not depicted in the graphs), all subbands have similar statistical centroids, which are significantly different from $1/2$, especially for the higher bitplanes. Similar results are obtained for other images of the corpus.

Taking $\delta_{P,b}$ as an indicator of the actual probability distribution within the quantization interval, it is clear that different probability distributions rather than the uniform one should be considered. The objective is to unbalance the probability of coefficients within the quantization interval, especially for the highest bitplanes. We assume that the probability density function (pdf) takes the lineal form $p'(x) = c + \alpha \cdot x$, though others functions are valid and achieve similar results. We also assume that the centroid occurs at a point within the quantization interval at which the accumulated probability of coefficients is $1/2$. More precisely $\int_{2^P}^{2^P+\delta \cdot 2^P} p'(x)\ dx = 0.5$. The solution to the linear system

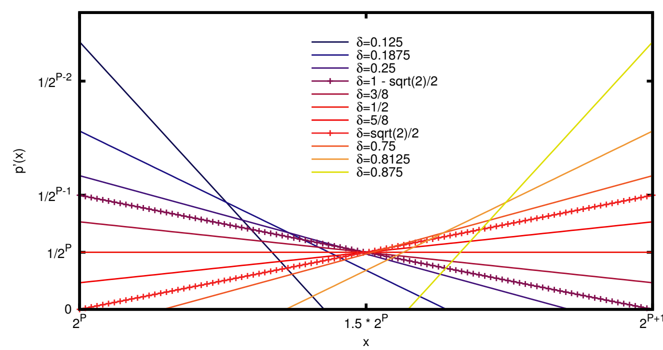


Fig. 2: Probability density functions calculated for different values of the parameter δ_P.

$$\left\{ \begin{array}{l} \displaystyle\int_{2^P}^{2^P+\delta_P \cdot 2^P} (c + \alpha \cdot x)\ dx = 0.5 \\ \displaystyle\int_{2^P}^{2^{P+1}} (c + \alpha \cdot x)\ dx = 1 \end{array} \right. , \tag{7}$$

with respect to variables c and α, allows us to express $p'(x)$ as a function of bitplane P and parameter δ_P as

$$p'(x) = \begin{array}{l} \dfrac{\delta_P \cdot 2^{1-P} - 3 \cdot 2^{-1-P} + (\delta_P)^2 \cdot 2^{-P}}{(\delta_P)^2 - \delta_P} - \\ \dfrac{2^{-2P}(2 \cdot \delta_P - 1)}{(\delta_P)^2 - \delta_P} \cdot x \end{array} . \tag{8}$$

When $\delta_P = 1/2$, expression (8) simplifies to $p'(x) = 1/2^P$. As shown in Figure 2, when $\delta_P \neq 1/2$, $p'(x)$ slopes down or up balancing the probabilities of coefficients within the quantization interval. As seen in Figure 2, when $\delta_P < 1 - \frac{\sqrt{2}}{2}$ or $\delta_P > \frac{\sqrt{2}}{2}$, $p'(x)$ becomes the max between the lineal form and 0.

Preferably, the values for $\mathcal{D}_P^{sig}$, $\triangle\mathcal{D}_P^{sig}$, $\mathcal{D}_P^{ref}$, and $\triangle\mathcal{D}_P^{ref}$ are determined through the estimated centroids from each individual image, as described above. However, this may not be practical. A simple yet effective strategy is to approximate the centroids through some function, since our experience suggests that a coarse estimation, roughly approximating low reconstruction factors for the highest bitplanes and high reconstruction factors for the lowest bitplanes, is sufficient to approximate distortion with high accuracy (see Section IV). Figure 1(c) depicts, for example, some exponential functions that resemble the experimentally calculated values for $\delta_{P,b}$. Since all subbands have similar statistical centroids, a single function can be employed for all of them. Experiments in Section IV approximate $\delta_{P,b}$ through

$$\delta'_{P,b} = \frac{log_{10}\left(\frac{K_b - P}{K_b}\right)}{5} + 0.475 \ , \tag{9}$$

except for the lowest frequency subband, which has mid-point centroids. Variations on this function, or even the consideration of different functions for each subband, do not change results significantly.

C. Other considerations

A practical advantage derived from using a function to approximate the values of $\delta_{P,b}$ is that the distortion and distortion decrease expected per coded bit can be pre-computed for different values of K_b, and used through lookup tables in implementations. Table I depicts, for example, the lookup table of the squared error and the squared error decrease for significance and refinement coding when $K_b = 8$. For comparison purposes, this table[1] also shows the estimators for $\delta = 1/2$. Another practical advantage derived from using a function to approximate the values of $\delta_{P,b}$ is that the decoder can use pre-computed lookup tables containing the estimated value for the non-transmitted bits of the coefficient, thus the reconstruction procedure can be carried out through a bitwise operation joining transmitted and non-transmitted bits. This operation has negligible computational costs.

Regarding the use of the IWT, the decoder must take into account that when integer transformations are employed, wavelet coefficients are represented in a discrete space, and the reconstruction procedure must carry out a rounding operation [13]. Furthermore, estimators need to consider the fact that the quantization interval is $[2^P, 2^{P+1} - 1]$ rather than $[2^P, 2^{P+1})$. To the best of our knowledge, this issue has not previously been studied in the context of distortion estimation. As we will see in Section IV, it may have a significant impact. Even though the approach described above still holds, the expressions need to be re-written to consider the discrete space as

$$\hat{\mathcal{D}}^{sig}_{P^*} = \sum_{\hat{x}=2^{P^*}}^{2^{P^*+1}-1} \hat{p}(\hat{x}) \ \hat{x}^2 \ , \tag{10}$$

$$\triangle\hat{\mathcal{D}}^{sig}_{P^*} = \sum_{\hat{x}=2^{P^*}}^{2^{P^*+1}-1} \hat{p}(\hat{x}) \left[\hat{x}^2 - \left(\hat{x} - \mathrm{R}(2^{P^*} + \delta_{P^*} \cdot 2^{P^*})\right)^2\right] \ , \tag{11}$$

[1]Extended lookup tables can be found at http://www.deic.uab.es/∼francesc

TABLE I: Squared error and squared error decrease expected at each bitplane from significance and refinement bits.

	$\mathcal{D}^{sig}_P$			$\mathcal{D}^{ref}_P$	
P	$\delta = \frac{1}{2}$	$\delta'_{P,b}$ (according exp. (9))			$\delta = \frac{1}{2}$
7	38229.33	30120.28	$\delta' = 0.294$		
6	9557.33	8256.06	$\delta' = 0.355$	946.45	1365.33
5	2389.33	2152.14	$\delta' = 0.389$	301.79	341.33
4	597.33	552.40	$\delta' = 0.415$	80.34	85.33
3	149.33	140.76	$\delta' = 0.434$	20.64	21.33
2	37.33	35.72	$\delta' = 0.450$	5.23	5.33
1	9.33	9.04	$\delta' = 0.463$	1.32	1.33
0	2.33	2.28	$\delta' = 0.475$	0.33	0.33

	$\triangle\mathcal{D}^{sig}_P$			$\triangle\mathcal{D}^{ref}_P$	
P	$\delta = \frac{1}{2}$	$\delta'_{P,b}$ (according exp. (9))			$\delta = \frac{1}{2}$
7	36864	29173.83	$\delta' = 0.294$		
6	9216	7954.26	$\delta' = 0.355$	616.77	1024
5	2304	2071.79	$\delta' = 0.389$	215.98	256
4	576	531.76	$\delta' = 0.415$	58.83	64
3	144	135.53	$\delta' = 0.434$	15.27	16
2	36	34.39	$\delta' = 0.450$	3.89	4
1	9	8.71	$\delta' = 0.463$	0.98	1
0	2.25	2.20	$\delta' = 0.475$	0.25	0.25

$$\hat{\mathcal{D}}^{ref}_{P^*-1} = \sum_{\hat{x}=2^{P^*}}^{2^{P^*+1}-1} \hat{p}(\hat{x}) \left(\hat{x} - \mathrm{R}(2^{P^*} + \delta_{P^*} \cdot 2^{P^*})\right)^2 \ , \tag{12}$$

and

$$\begin{aligned}\triangle\hat{\mathcal{D}}^{ref}_{P^*-1} = & \sum_{\hat{x}=2^{P^*}}^{1.5 \cdot 2^{P^*}-1} \hat{p}(\hat{x}) \left[\left(\hat{x} - \mathrm{R}(2^{P^*} + \delta_{P^*} \cdot 2^{P^*})\right)^2 - \right. \\ & \left. \left(\hat{x} - \mathrm{R}(2^{P^*} + \delta_{P^*-1} \cdot 2^{P^*-1})\right)^2\right] + \\ & \sum_{\hat{x}=1.5 \cdot 2^{P^*}}^{2^{P^*+1}-1} \hat{p}(\hat{x}) \left[\left(\hat{x} - \mathrm{R}(2^{P^*} + \delta_{P^*} \cdot 2^{P^*})\right)^2 - \right. \\ & \left. \left(\hat{x} - \mathrm{R}(2^{P^*} + 2^{P^*-1} + \delta_{P^*-1} \cdot 2^{P^*-1})\right)^2\right] ,\end{aligned} \tag{13}$$

where $\mathrm{R}(\cdot)$ denotes the rounding operation, and $\hat{p}(\hat{x}) = \int_{\hat{x}}^{\hat{x}+1} p'(x) \ dx$ is an estimate of the probability that the coefficient takes the value $\hat{x}$. In the case of the IWT, D^{null} is obviously not relevant. As depicted in Figures 1(a) and 1(b), statistical centroids are similar for both the DWT and the IWT except for the lowest bitplane. Thus, experiments carried out in Section IV use equation (9) for centroids estimated in the case of both transforms. The one exception is for the lowest bitplane for the IWT, in which $\delta'_{0,b} = 0$.

For the sake of simplicity, throughout this section we have ignored the energy gain factor G_b for each subband, that arises when the transform is not orthonormal. When non-orthonormal filter-banks are employed, the energy gain factor G_b must multiply the values determined for the distortion estimators for each subband.

III. Application cases

A. Post compression rate-distortion optimization

A suitable application in which the proposed estimators may be employed is in the post-compression rate-distortion (PCRD) optimization process described in [6]. The PCRD optimization process is often used in JPEG2000 encoders to select those bitstream segments that minimize the image distortion at a given target bitrate. Such a process is useful since JPEG2000 encodes sets of wavelet coefficients (called codeblocks) independently, producing one embedded bitstream for each. These bitstreams can be potentially truncated at three points per bitplane, coinciding with the end of each coding pass as defined by the JPEG2000 standard. These coding passes are named the Significance Propagation Pass (SPP), the Magnitude Refinement Pass (MRP), and the Cleanup Pass (CP). The coding pass SPP encodes the significance state of those coefficients that are more likely to become significant in the current bitplane; MRP encodes the refinement bits of coefficients that have become significant in previous bitplanes; and CP encodes the significance state of coefficients not encoded in SPP or MRP.

The PCRD algorithm employs a generalized Lagrange multiplier formulation for a discrete set of points. In brief, PCRD first identifies the convex hull of the operational distortion-rate function for each codeblock bitstream. It then selects from the union of all codeblocks those segments with the highest distortion-rate slopes. When the distortion metric is MSE, the distortion of the bitstream segments corresponding to codeblock $\mathcal{B}_i$ are computed as $D_i^l = G_b \sum_{k \in \mathcal{B}_i} (y[k] - \hat{y}^l[k])^2$, where $\hat{y}^l[k]$ denotes the quantized representation of wavelet coefficients $y[k]$ when the bitstream for $\mathcal{B}_i$ is truncated after coding pass $l, 0 \leq l \leq l_{max}$. l identifies the coding pass as $l = P \cdot 3 + cp$, where P is the bitplane number, and $cp = \{2, 1, 0\}$ respectively for the SPP, MRP, and CP coding passes. The distortion-rate slope of coding pass l is computed as

$$S_i^l = \frac{\triangle D_i^l}{\triangle R_i^l} = \frac{D_i^{l-1} - D_i^l}{R_i^l - R_i^{l-1}}, \tag{14}$$

with R_i^l denoting the length (in bits) of the bitstream at truncation point l.

Since PCRD is used in many JPEG2000 implementations, this optimization process has been widely studied in the literature and several different methods have been proposed. In some cases, inexpensive estimations of the distortion decrease $\triangle D_i^l$ are employed in order to reduce the computational load of the encoder (see [7], [11], for instance). In other cases, the optimization of the distortion computation is a cause of concern, especially in software and hardware implementations. Kakadu[2], for example, which is well known for its excellent compression performance, uses a lookup table considering the 5 most significant bits of the coefficients to reduce computational costs and, more recently, a novel strategy for hardware-based architectures has been presented in [14].

[2]See http://www.kakadusoftware.com

Through the estimators introduced in the previous section of this paper, the distortion decrease of bitstream segments can be estimated by considering only the number of significant and refinement bits encoded in each coding pass, referred to as $\#S^l$ and $\#R^l$, according to

$$\triangle D_i^l = G_b \cdot \left[\triangle \mathcal{D}_P^{sig} \cdot \#S^l + \triangle \mathcal{D}_P^{ref} \cdot \#R^l \right], \tag{15}$$

where P is the bitplane to which coding pass l belongs. To count the number of significant/refinement bits is a computationally inexpensive operation during the encoding process [4], and has a low complexity when it is carried out before coding [11]. In addition to speed-up of the bitplane coding engine, the use of this technique may also help to reduce the memory requirements of the block coder [3, Ch. 17.2.4].

B. Intrinsic distortion estimation

The second application in which the use of the proposed estimators can be beneficial is image distortion estimation. It may be desirable to target image quality rather than the bitrate [7], or to approximate image distortion without comparing to the original samples. This latter case may be useful to minimize computational resources, or to aid in transcoding procedures [12].

Still within the framework of JPEG2000, and considering D_i^l as an additive metric [6], the distortion of the image is determined as $D = \sum_i D_i^{\lambda(l)}$, where $\lambda(l)$ denotes the optimal set of truncation points given for a target bitrate, or quality. In this paragraph, we treat the case for the DWT. The IWT case is addressed in the paragraph that follows. The distortion produced by the quantized representation of the coefficients in $\mathcal{B}_i$ at coding pass l is estimated according to

$$\begin{aligned} {D'}_i^l = \quad & G_b \cdot \left[\sum_{L=0}^{l-1} \#S^L \cdot \mathcal{D}_{\lfloor L/3 \rfloor}^{sig} + \right. \\ & \left(\sum_{L=l}^{l_{max}} \#S^L \right) \cdot \mathcal{D}_{\lfloor l/3 \rfloor - 1}^{ref} + \\ & \left. \#R^{l-1} \cdot \ \triangle \mathcal{D}_{\lfloor l/3 \rfloor}^{ref} \ + \ \#N \cdot \mathcal{D}^{null} \ \right] , \end{aligned} \tag{16}$$

where the floor operation $\lfloor \cdot \rfloor$ is used here to determine a bitplane number from a coding pass. The first term of this expression accumulates the distortion produced by coefficients that become significant below the current coding pass. The second term accounts for the distortion produced by coefficients that become significant in the current or previous coding passes. This term is valid even when all bitplanes have been encoded (i.e., $l = 0$). In this case, the distortion is due to the initial quantization. In the case of $\delta = 1/2$, for example, this is $\mathcal{D}_{-1}^{ref} = 1/12$ according to expression (4). The third term accounts for the distortion decrease produced when the MRP pass of the current bitplane is coded, and therefore it is only non-null when the current coding pass is the SPP. The last term of the expression takes into account the distortion caused by coefficients with null contributions, $\#N$ denoting the number of such coefficients.

Due to the perfect reconstruction of the IWT from integer coefficients, the estimation of the distortion for such transforms must consider that null coefficients produce null distortion, and that when coefficients are completely transmitted there is no remaining distortion. On the other hand, it is worth considering the noise introduced by the nonlinear transform when coefficients are not perfectly reconstructed. This was first studied in [13] through a model that considered the degradation produced in the reconstructed coefficients due to the rounding operations of the IWT's lifting scheme. In the case of the 5/3 IWT with 3 levels of decomposition, the impact on the MSE was determined to be around $\beta = 0.5$. For the experiments carried out in the next Section using the 5/3 IWT with 5 levels of decomposition, our empirical results confirm that $\beta = 0.5$ is an appropriate choice. The distortion of coefficients at coding pass l for the IWT is estimated as

$$
\begin{aligned}
D''^{l}_{i} = \quad & G_b \cdot \Bigg[\sum_{L=0}^{l-1} \#S^L \cdot \hat{\mathcal{D}}^{sig}_{\lfloor L/3 \rfloor} \; + \\
& \left(\sum_{L=l}^{l_{max}} \#S^L \right) \cdot \hat{\mathcal{D}}^{ref}_{\lfloor l/3 \rfloor - 1} \; + \\
& \#R^{l-1} \cdot \triangle\hat{\mathcal{D}}^{ref}_{\lfloor l/3 \rfloor} \; + \; \#T \cdot \beta \Bigg]
\end{aligned} \quad , \qquad (17)
$$

with $\#T$ denoting the number of coefficients that have not been perfectly reconstructed.

C. Extrinsic distortion estimation

When the number of significant/refinement bits per coding pass is available or, in other words, when the *intrinsic* characteristics of the coding process are available, distortion can be determined through the previous approach. However, once the image has been already encoded, this information is not commonly available unless the codestream is fully decoded. Applications such as real-time video rendering [15], or interactive image and video transmission [16], require distortion estimation, but may have limited computational resources. Therefore, decoding the codestream to estimate distortion may not be an option.

An effective strategy to avoid decoding is to estimate the number of significant/refinement bits encoded in each coding pass. This can be carried out by considering only the length of the bitstream generated for coding passes, which can be obtained for JPEG2000 by decoding only packet headers[3]. This operation is computationally inexpensive. In other coding systems, the length of coding passes might also be extracted without needing to decode the full codestream [4].

With $\mathcal{L}_i^l$ standing for the length (in bits) of coding pass l, the number of significant/refinement bits can be simply estimated as $\#S = \gamma \cdot \mathcal{L}_i^l$, where γ represents the efficiency factor achieved by the coding engine, i.e., the number of significant/refinement bits that are encoded per emitted bit, on average. For the JPEG2000 coding engine, γ has been experimentally determined as 1.075 for MRP coding passes. For SPP coding passes (CP, respectively), γ shows a lineal increase from 0.25 (0.15, respectively) at the highest bitplane, to 0.3 (0.2, respectively) at the lowest bitplane.

[3]We assume that the restart coding variation is active [3, Chapter 12.4].

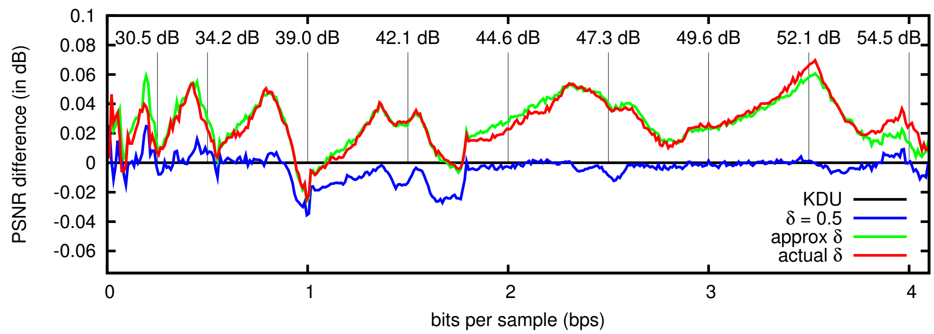
IV. Experimental results

We report the performance of the proposed estimators for the images "Portrait" and "Cafeteria" of the ISO 12640-1 corpus (8-bit gray-scale, size 2560×2048). Similar results are obtained for all images of the corpus. In all results[4], both the encoder and the decoder use the same values for the reconstruction factor δ, and the performance is reported when: 1) the actual centroids of the image are used (denoted as $actual\ \delta$); 2) the δ values are approximated through expression (9) (denoted as $approx\ \delta$); and 3) when mid-point reconstruction is employed (denoted as $\delta = 0.5$).

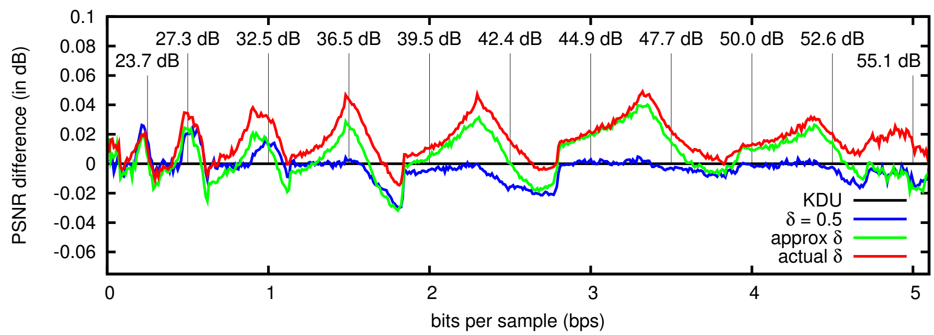
For the first application (PCRD minimization of distortion for a given target bitrate), images are encoded at 600 target bitrates uniformly distributed between 0.01 to 6 bits per sample (bps). The images are then decoded and compared with the originals. To fairly assess the proposed estimators, the images are also encoded at the same bitrates with Kakadu v6.1 employing its lookup table-based strategy to compute distortion (denoted as KDU), and decoded using mid-point reconstruction. Figures 3(a) and 3(b) depict the Peak Signal to Noise Ratio (PSNR) difference between the usual strategy of Kakadu (horizontal line) and the proposed estimators when the lossy mode of JPEG2000 is used. Labels in these plots report the actual PSNR at several points. Results suggest that the mid-point estimator achieves similar coding performance to that of Kakadu, and that pdf-based estimators improve it slightly, though the differences are minimal (no more than 0.07 dB). Figures 4(a) and 4(b) depict the results of the same test when using the JPEG2000 lossless mode. In this case, the best performance is achieved with pdf-based estimators, especially at high bitrates, where improvements around 1 dB are achieved. The figure also includes a plot labeled $\delta = 0.5$ (DWT) which depicts the coding performance achieved by the estimator determined for the DWT case (paragraph 2 of Section III-B) to stress the low performance of such an estimator in the IWT case.

The improvement in performance over that of Kakadu for the IWT case is due to both the encoder and decoder. As mentioned previously, the Kakadu encoder and decoder assume mid-point reconstruction when they respectively compute distortion decreases and dequantize coefficients, which is a common approach in most studies and implementations [17]. The maximum gains are achieved at high bit-rates, when bits belonging to the lowest bitplane are encoded. This suggests that the employed centroid has a significant displacement. To better appraise the improvement achieved with the pdf-based estimators when the JPEG2000 lossless mode is used, Figure 5 depicts the coding performance achieved when different assumptions are used in the encoder and the decoder. More precisely, we evaluate the coding performance achieved when the encoder computes the actual distortion decreases

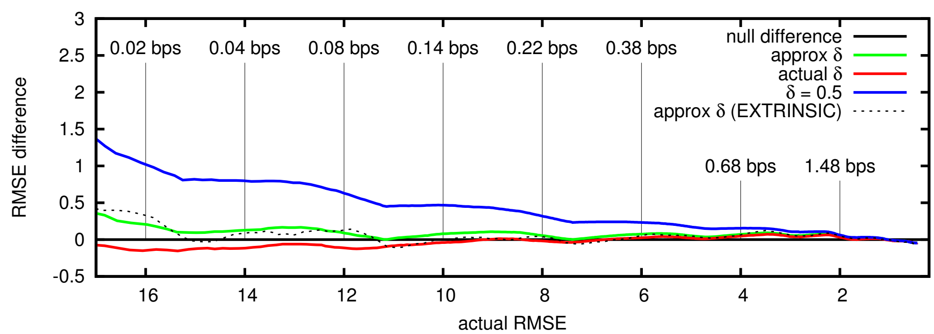
[4]Coding parameters are: 5 levels of WT, 64×64 codeblocks, no precincts, restart coding variation.



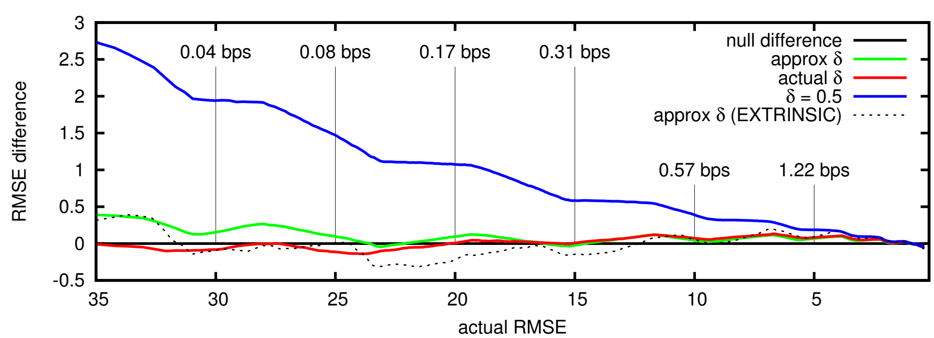
(a) Coding performance - "Portrait"



(b) Coding performance - "Cafeteria"



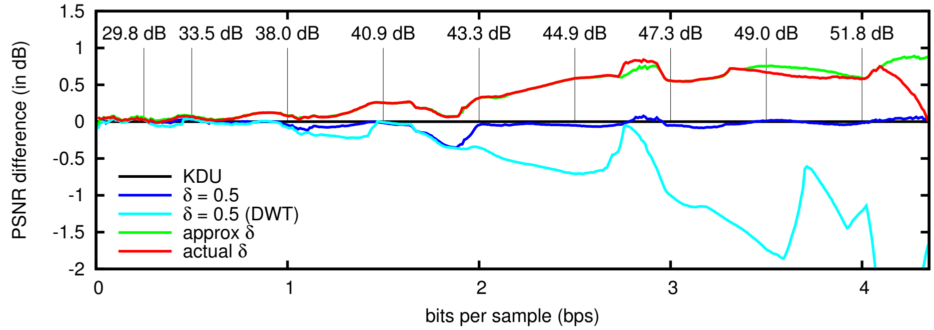
(c) Estimated distortion - "Portrait"



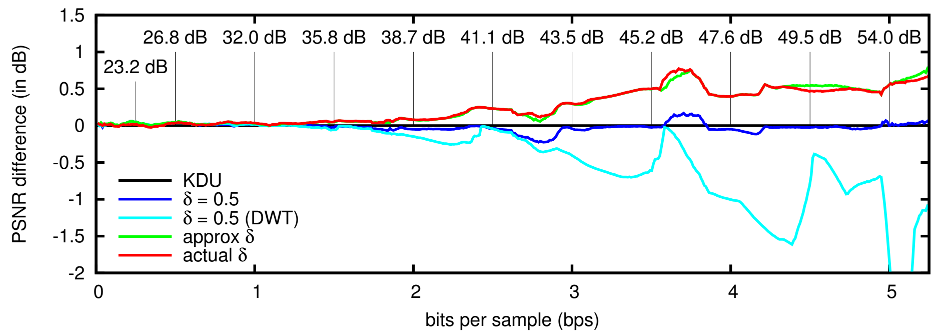
(d) Estimated distortion - "Cafeteria"

Fig. 3: Evaluation of the proposed estimators when using the 9/7 DWT (JPEG2000 lossy mode).

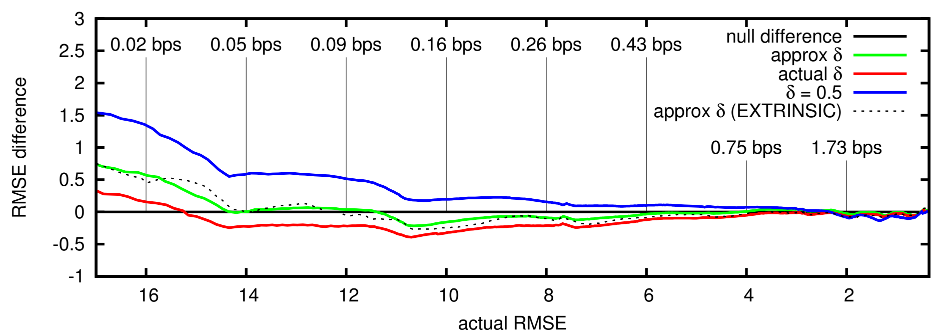
assuming mid-point reconstruction or when it uses the pdf-based estimators, respectively labeled as $encode\ KDU$ and $encode\ approx\ \delta$, and when the decoder uses mid-point or pdf-based reconstruction, respectively labeled as $decode\ \delta = 0.5$ and $decode\ approx\ \delta$. Results are reported as the PSNR difference between coding and decoding using actual distortion decreases and mid-point reconstruction – which is the usual strategy of Kakadu and the most common approach in implementations – against the other combinations. Note that the plot labeled $encode\ approx\ \delta$ / $decode\ approx\ \delta$ reports



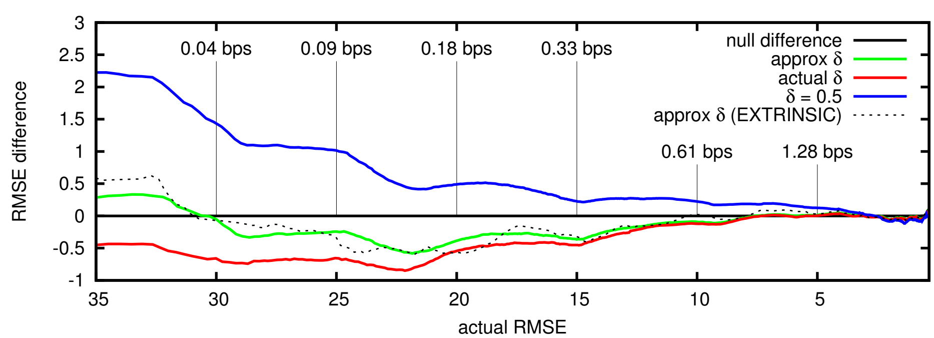
(a) Coding performance - “Portrait”



(b) Coding performance - “Cafeteria”



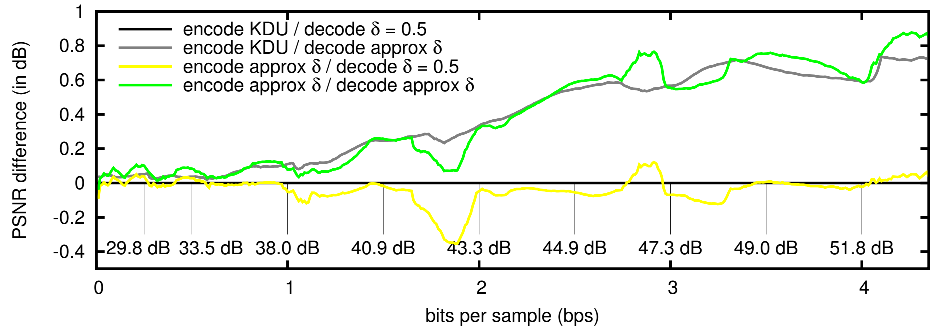
(c) Estimated distortion - “Portrait”



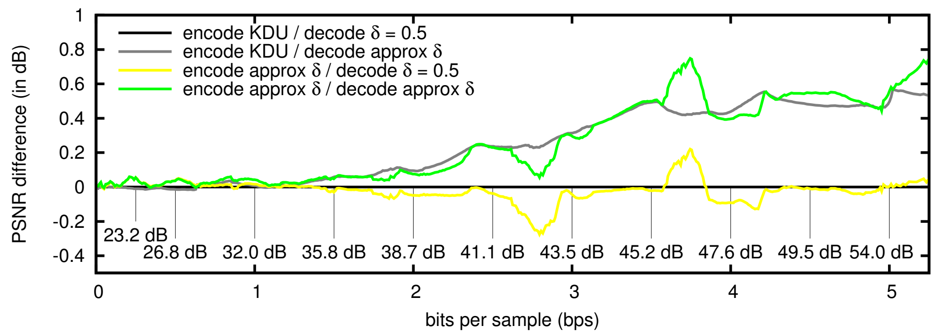
(d) Estimated distortion - “Cafeteria”

Fig. 4: Evaluation of the proposed estimators when using the 5/3 IWT (JPEG2000 lossless mode).

the same results as the plot labeled *approx* δ in Figures 4(a) and 4(b). Figure 5 suggests that, even though the encoder may compute actual distortion decreases in the usual way, assuming mid-point reconstruction, the decoder can enhance the quality of the image through pdf-based reconstruction. Furthermore, we remark that, in spite of recovering coefficients using mid-point reconstruction, decoders may still enhance the quality of decoded images for the IWT case by using the dequantization interval $[2^P, 2^{P+1}-1]$ instead of $[2^P, 2^{P+1})$. Though this enhancement should be considered when evaluating bitplane



(a) Coding performance - "Portrait"



(b) Coding performance - "Cafeteria"

Fig. 5: Evaluation of the coding performance achieved when different assumptions of mid-point and pdf-based reconstruction are used in the encoder and the decoder, with the 5/3 IWT (JPEG2000 lossless mode).

coding engines or rate-distortion optimization methods, the largest gains are achieved in a bitrate range that is typically visually lossless.

For the second application (intrinsic distortion estimation), the estimators can be either applied to target a quality for the final codestream, or to estimate the image distortion given a target bitrate. In both cases the results are the same, hence a single graph is provided. Figures 3(c) and 3(d) depict the difference between the estimated Root Mean Squared Error (RMSE) and its actual value. Labels in these plots report the bitrate at several points. For both the "Portrait" and "Cafeteria" images, the estimated distortion when using pdf-based estimation is close to the actual distortion, whereas the accuracy achieved with the mid-point approach is poor, especially at low bitrates. Similar results hold for the JPEG2000 lossless mode (Figures 4(c) and 4(d)). The graphs also depict the extrinsic distortion estimation (denoted as $approx\ \delta$ (EXTRINSIC)) as the third practical application for the estimators. Note that, even though the values for δ, and the number of significant/refinement coefficients are approximated (i.e., there is no reference to the original image), the difference between the estimated RMSE and the actual one is never higher than ± 0.5 for the lossy mode, and ± 0.75 for the lossless mode.

V. Conclusions

Distortion estimation is an important issue in studies and applications of lossy, and lossy-to-lossless, image compression. Most work estimates the distortion induced in wavelet coefficients assuming that they are uniformly distributed within the quantization interval, even though this does not correspond with the actual nature of the signal. The first contribution of this work is the presentation of a theoretic approach to determine distortion for non-uniform distributions. This has lead to estimators that can accurately approximate the distortion, and distortion decrease, due to bits emitted by bitplane coding engines. The second main contribution of this work is the introduction of the proposed estimators in three applications within the framework of JPEG2000. Experimental results suggest that, with the proposed approach, distortion estimators agree closely with actual distortion computations, significantly surpassing the performance of estimators using mid-point reconstruction. Our estimators can be employed in studies and applications to better adjust rate-distortion models, improve compression performance, reduce computational costs, determine distortion with high accuracy, and optimize transcoding procedures.

Acknowledgment

The authors thank the associate editor and the anonymous reviewers for their valuable remarks.

References

[1] S. Mallat, "A theory of multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, pp. 674–693, Jul. 1989.

[2] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Image Process.*, vol. 41, no. 12, pp. 3445–3462, Dec. 1993.

[3] D. S. Taubman and M. W. Marcellin, *JPEG2000 Image compression fundamentals, standards and practice*. Norwell, Massachusetts 02061 USA: Kluwer Academic Publishers, 2002.

[4] M. Farshchian, S. Cho, and W. A. Pearlman, "Optimal error protection for real-time image and video transmission," *IEEE Signal Process. Lett.*, vol. 11, no. 10, pp. 780–783, Oct. 2004.

[5] R. W. Buccigrossi and E. P. Simoncelli, "Image compression via joint statistical characterization in the wavelet domain," *IEEE Trans. Image Process.*, vol. 8, no. 12, pp. 1688–1701, Dec. 1999.

[6] D. Taubman, "High performance scalable image compression with EBCOT," *IEEE Trans. Image Process.*, vol. 9, no. 7, pp. 1158–1170, Jul. 2000.

[7] Y.-W. Chang, H.-C. Fang, C.-C. Cheng, C.-C. Chen, and L.-G. Chen, "Precompression quality-control algorithm for JPEG 2000," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3279–3293, Nov. 2006.

[8] J. Li and S. Lei, "An embedded still image coder with rate-distortion optimization," *IEEE Trans. Image Process.*, vol. 8, no. 7, pp. 913–924, Jul. 1999.

[9] F. W. Wheeler, "Trellis source coding and memory constrained image coding," Ph.D. dissertation, Rensselaer Polytechnic Institue, Troy, NY, Sep. 2000.

[10] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 243–250, Jun. 1996.

[11] K. Vikram, V. Vasudevan, and S. Srinivasan, "Rate-distortion estimation for fast JPEG2000 compression at low bit-rates," *IEE Electronics Letters*, vol. 41, no. 1, pp. 16–18, Jan. 2005.

[12] F. Auli-Llinas and J. Serra-Sagrista, "JPEG2000 quality scalability without quality layers," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 7, pp. 923–936, Jul. 2008.

[13] J. Reichel, G. Menegaz, M. J. Nadenau, and M. Kunt, "Integer wavelet transform for embedded lossy to lossless image compression," *IEEE Trans. Image Process.*, vol. 10, no. 3, pp. 383–392, Mar. 2001.

[14] A. K. Gupta, S. Nooshabadi, and D. Taubman, "Novel distortion estimation technique for hardware-based JPEG2000 encoder system," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 7, pp. 918–923, Jul. 2007.

[15] F. Auli-Llinas and D. Taubman, "JPEG2000 block-wise truncation of quality layers," in *Proc. IEEE/IET International Conference on Visual Information Engineering*, Aug. 2008, pp. 711–716.

[16] ——, "Optimal delivery of motion JPEG2000 over JPIP with block-wise truncation of quality layers," in *Proc. IEEE International Conference on Image Processing*, Oct. 2008, pp. 2856–2859.

[17] J. An and Z. Cai, "Efficient rate control for lossless mode of JPEG2000," *IEEE Signal Process. Lett.*, vol. 15, pp. 409–412, Apr. 2008.

Francesc Aulí-Llinàs (S'2006-M'2008) received the B.Sc. and B.E. degrees in Computer Management Engineering and Computer Engineering in 2000 and 2002 respectively, both with highest honours from the Universitat Autònoma de Barcelona (Spain), and for which he was granted with two extraordinary awards of Bachelor (awards given to the first students of each promotion). In 2002 he obtained a postgraduate fellowship from the Catalan Government to fund four years of doctoral studies, receiving the M.S. degree and the Ph.D. degree (with honors) in 2004 and 2006 respectively, both in Computer Science from the Universitat Autònoma de Barcelona.

He is the main developer of BOI, a JPEG2000 Part 1 implementation that was awarded with a free software mention from the Catalan Government in April 2006. In 2007, he was granted with a postdoctoral fellowship from the Spanish Government funding two years of postdoctoral research to carry out two research stages of one year each with the groups of David Taubman, at the University of New South Wales (Australia), and Michael Marcellin, at the University of Arizona (USA). He also collaborates with the Group on Interactive Coding of Images, from the Universitat Autònoma de Barcelona.

He is reviewer for several magazines and symposiums, has authored several papers in the most important journals and conferences in his field, and has co-directed one Ph.D. thesis. His research interests include a wide range of image coding topics, including highly scalable image and video coding systems, rate-distortion optimization techniques, distortion estimation, region-of-interest coding, and interactive image and video transmission among others.

Michael W. Marcellin (S'81-M'87-SM'93-F'02) was born in Bishop, California, on July 1, 1959. He graduated *summa cum laude* with the B.S. degree in Electrical Engineering from San Diego State University in 1983, where he was named the most outstanding student in the College of Engineering. He received the M.S. and Ph.D. degrees in Electrical Engineering from Texas A&M University in 1985 and 1987, respectively.

Since 1988, Dr. Marcellin has been with the University of Arizona, where he holds the title of Regents' Professor of Electrical and Computer Engineering, and of Optical Sciences. His research interests include digital communication and data storage systems, data compression, and signal processing. He has authored or coauthored more than two hundred publications in these areas.

Dr. Marcellin is a major contributor to JPEG2000, the emerging second-generation standard for image compression. Throughout the standardization process, he chaired the JPEG2000 Verification Model Ad Hoc Group, which was responsible for the software implementation and documentation of the JPEG2000 algorithm. He is coauthor of the book, *JPEG2000: Image compression fundamentals, standards and practice,* Kluwer Academic Publishers, 2002. This book is intended to serve as a graduate level textbook on image compression fundamentals, as well as the definitive reference on JPEG2000. Dr. Marcellin served as a consultant to Digital Cinema Initiatives (DCI), a consortium of Hollywood studios, on the development of the JPEG2000 profiles for digital cinema.

Professor Marcellin is a Fellow of the IEEE, and is a member of Tau Beta Pi, Eta Kappa Nu, and Phi Kappa Phi. He is a 1992 recipient of the National Science Foundation Young Investigator Award, and a corecipient of the 1993 IEEE Signal Processing Society Senior (Best Paper) Award. He has received teaching awards from NTU (1990, 2001), IEEE/Eta Kappa Nu student sections (1997), and the University of Arizona College of Engineering (2000). In 2003, he was named the San Diego State University Distinguished Engineering Alumnus. Professor Marcellin is the recipient of the 2006 University of Arizona Technology Innovation Award. From 2001 to 2006, Dr. Marcellin was the Litton Industries John M. Leonis Professor of Engineering. He is currently the International Foundation for Telemetering Professor of Electrical and Computer Engineering at the University of Arizona.