Stationary Probability Model for Bitplane Image Coding through Local Average of Wavelet Coefficients

Francesc Aulí-Llinàs, *Member, IEEE*

***Abstract*—This paper introduces a probability model for symbols emitted by bitplane image coding engines, which is conceived from a precise characterization of the signal produced by a wavelet transform. Main insights behind the proposed model are the estimation of the magnitude of wavelet coefficients as the arithmetic mean of its neighbors' magnitude (the so-called local average), and the assumption that emitted bits are under-complete representations of the underlying signal. The local average-based probability model is introduced in the framework of JPEG2000. While the resulting system is not JPEG2000 compatible, it preserves all features of the standard. Practical benefits of our model are enhanced coding efficiency, more opportunities for parallelism, and improved spatial scalability.**

***Index Terms*—Bitplane coding, context-adaptive coding, JPEG2000.**

I. Introduction

ARITHMETIC coding [1], [2] is a data compression technique that utilizes the probabilities of input symbols to represent the original message by an interval of real numbers in the range $[0, 1)$. Briefly described, the arithmetic coder segments the interval $[0, 1)$ in as many subintervals as the size of the original alphabet, with subinterval sizes according to the probabilities of input symbols, and chooses the subinterval corresponding to the first symbol. This procedure is repeated for following symbols within the selected intervals. The transmission of any number within the range given by the final interval guarantees that the reverse procedure decodes the original message losslessly.

Key to the compression efficiency is the probability model used for input symbols. The more skewed the probabilities, the larger most of the chosen subintervals, thus the easier to find a number within the final interval with a short representation. When the probability mass function of input symbols is not known a priori, or when it is too costly to compute or transmit it, probabilities can be adaptively adjusted as data are fed to the coder by means of heuristics, finite-state machines, or other techniques aimed to adjust symbol probabilities to the incoming message. The incorporation of such mechanism is referred to as *adaptive* arithmetic coding.

Another valuable mechanism introduced to arithmetic coders is the use of contexts. In addition to the input symbol, this mechanism provides to the coder the context in which the symbol is found. Contexts are devised to capture high-order statistics of data, so more redundancy can be removed. Combined with adaptivity, this mechanism is referred to as *context-adaptive* arithmetic coding.

Context-adaptive arithmetic coding has been spread in the field of image and video compression since mid-nineties. Standards of previous generations, such as JPEG [3] or MPEG-2 [4], employed variable length coding due to the lack of efficient arithmetic coding implementations, and to reduce computational costs. The introduction of highly efficient and low-cost implementations, like the Q-coder [5] or the M-coder [6], enabled the use of arithmetic coding in most modern image and video coding standards such as JBIG [7], JPEG2000 [8], or H.264/AVC [9].

This work is concerned with the probability model deployed in image coding systems that utilize bitplane coding together with (context-adaptive) arithmetic coding. Bitplane coding is a technique used for lossy, or lossy-to-lossless, compression that refines image distortion by means of progressively transmit the binary representation of image coefficients from the most significant bit to the least significant bit. Emitted bits are fed to an arithmetic coder that, if context-adaptive, captures statistics of data through some context formation approach.

Rather than to base the probability model on high-order statistics of symbols emitted by the bitplane coding engine, this work introduces a model that is conceived from a precise characterization of the signal that is coded. The proposed approach assumes stationary statistical behavior for emitted symbols. Probabilities are determined using the immediate neighbors of the currently coded coefficient. Our model can thus be considered as contextual, but non-adaptive. The proposed model is assessed in the framework of JPEG2000, which is a representative image coding system employing bitplane coding together with context-adaptive arithmetic coding. Experimental results indicate 2% increase on coding efficiency at medium and high bitrates, with no degradation at low bitrates. Furthermore, more options for parallelism, and enhanced coding efficiency for images that are coded with a high degree of spatial scalability are provided.

This paper is organized as follows. Section II reviews state-

The author is with the Department of Information and Communications Engineering, Universitat Autònoma de Barcelona, Bellaterra, Spain. (phone: +34 935813571; fax: +34 935814477; e-mail: fauli@deic.uab.es). This work has been partially supported by the Spanish Government (MICINN), by FEDER, and by the Catalan Government, under Grants RYC-2010-05671, 2008-BPB-0010, TIN2009-14426-C02-01, and 2009-SGR-1224.

of-the-art of lossy and lossless image and video coding to emphasize the consolidation of context-adaptive coding in modern image compression. Section III introduces the mathematical framework from which our probability model has been conceived, and Section IV describes forms to put in practice that model, proposing a general approach. Section V assesses the performance of the suggested implementation. The last section summarizes this work and remarks some points.

II. Review of entropy coding in image compression

In general, image and video coding systems are structured in three main coding stages. The first stage is aimed to remove inter-pixel redundancy, producing an –ideally– non-redundant signal. The second stage is commonly embodied as a coding engine that codes the outcome of the first stage in a lossy, lossless, or lossy-to-lossless regime. To diminish the statistical redundancy of symbols emitted by the coding engine, the third stage employs some entropy coding technique such as Huffman coding [10], Golomb-Rice coding [11], [12], or arithmetic coding, to produce the final codestream.

In the case of lossy image coding, the first inter-pixel redundancy removal stage is commonly implemented as a transform, like the discrete wavelet transform [13], [14], that captures high and low frequencies within subbands of a multi-resolution representation of the image. Since the introduction of bitplane coding [15], [16], most lossy and lossy-to-lossless image coding engines employ this strategy to successively refine image distortion. Let $[t_{K-1}, t_{K-2}, ..., t_1, t_0], t_i = \{0, 1\}$ be the binary representation for an integer υ that corresponds to the magnitude of the index obtained by quantizing a wavelet coefficient χ, with K denoting a sufficient number of bits to represent all coefficients. Bitplane coding strategies generally define bitplane j as the same bit t_j from all coefficients, and encode the image from the most significant bitplane $K-1$ to the least significant bitplane 0. The first non-zero 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. The significance state of χ in bitplane j is defined as

$$\Phi(\chi, j) = \begin{cases} 0 & \text{if } j > s \\ 1 & \text{otherwise} \end{cases} . \quad (1)$$

The most popular approach to remove high-order statistical redundancy of bits emitted by bitplane coders is context-adaptive arithmetic coding. To maximize the potential of arithmetic coding the approach to form contexts is aimed to skew the probabilities of emitted symbols. First bitplane coders employed primitive context formation approaches based on the coding tree [15], [16], which subsequently were enhanced by more elaborated strategies [17]–[19]. Nowadays, most approaches use the significance state of the neighbors of the currently coded coefficient. More precisely, let $\chi^n, 1 \leq n \leq N$ denote the n-th neighbor of χ. Commonly, contexts are selected as some function of these neighbors. Often, this function considers only $\{\Phi(\chi^n, j)\}$ and (possibly) indicates the number and the position of the neighbors that are significant in the current or previous bitplanes. The context helps to determine the probability of the currently emitted binary symbol t_j, expressed as $P_{sig}(t_j), j \geq s$ for significance coding, and as $P_{ref}(t_j), j < s$ for refinement coding.

Earliest image coding systems already encountered that the coding efficiency of context-adaptive arithmetic coding may drop substantially when the arithmetic coder is not fed with enough data to adjust the probabilities reliably [20], [21]. This problem is known as context dilution or sparsity, and has been tackled from different points of view. The approach used in JPEG2000, for example, employs an heuristic based on the image features captured in each wavelet subband [22], [23, Ch. 8.3.2], though heuristics pointed to other aspects also achieve competitive performance [24]–[27]. Other approaches to tackle the context dilution problem are statistical analysis [28], [29], delayed-decision algorithms [30], or M-order finite-context models [31]. From a theoretic point of view, the use of entropy-based concepts such as mutual information provides optimal performance [32]–[36].

Although there exist many variations, lossless image compression commonly employs predictive techniques to estimate the magnitude of the current sample in the first redundancy removal stage. Predictions use different strategies such as weighted averages of the magnitude of previously coded samples [37], [38], adaptive neural predictors [39], or neighbor configurations aimed to capture image features [21], [40]. The difference between the predicted magnitude and the actual one, called residual, is coded in the second stage by a coding engine that selects appropriate contexts for each sample. Context selection approaches are implemented using tree-models [20], fuzzy modeling [39], or context-matching with pre-computed lookup tables [41], among other techniques. Both residual and context are fed to an entropy coder, which may employ variable-length coding, such as in JPEG-LS standard [40], [42], or context-adaptive arithmetic coding [31], [37], [39], [43]–[45], which excels for its superior coding efficiency.

In the case of video coding, spatial and temporal redundancy among frames of a video sequence is typically removed through spatial predictors and motion compensation mechanisms. Residuals might then be decorrelated through some discrete cosine-like transform. Again, context-adaptive coding is employed extensively to diminish the statistical redundancy of symbols. First approaches are found as early as in 1994 [46], and context modeling has been approached from different points of view, such as Markov process [47], or heuristics [6].

Context-adaptive arithmetic coding is a consolidated technology in image compression to remove statistical redundancy of symbols emitted by coding engines. Probabilities of emitted symbols are determined through the context formation approach, which must be carefully selected. Even so, experimental evidence [48] suggests that non-elaborated approaches also achieve competitive performance. Furthermore, other coding strategies using *non-adaptive* arithmetic coding, such as those based on Tarp-filter [49]–[51], may also achieve competitive compression efficiency.

For simplicity, throughout this section χ denoted the magnitude of a wavelet coefficient, without considering its sign.

This work is not concerned with sign coding, which is another field of study [28], [45], [46], [52]–[54].

III. Signal's characterization through local average of wavelet coefficients

Our final goal is to determine $P_{sig}(t_j)$ and $P_{ref}(t_j)$, which are the probabilities of symbols emitted by a bitplane coding engine that encodes the signal produced by a wavelet transform. Examples in this section are produced with the irreversible CDF 9/7 wavelet transform [14], [55]. The generalized algorithm introduced in Section IV can be extended to any other wavelet transform.

The characterization of the signal produced by wavelet transforms is a topic explored in the literature since the nineties. Studies with objectives similar to those pursued in this work are [56], [57]. Main differences between these studies and this work are the use of different conditional probability density functions, and the implementation of the proposed approach in the framework of the advanced image coding standard JPEG2000 without restraining any one of its features. Of especial interest is that our approach allows the independent coding of small sets of wavelet coefficients, which is necessary to provide spatial scalability in JPEG2000. This forces to use only intrascale dependencies when defining the probability model, which may not be supported by other non-adaptive approaches such as [50], [56], [57].

The marginal probability density function (pdf) for coefficients within wavelet subbands is commonly modeled as a generalized Gaussian distribution [56]–[58]. Figure 1(a) depicts this Gaussian-like pdf, referred to as $p(\chi)$, when only the magnitude of coefficients is considered. The determination of probabilities for bits emitted for χ considers information regarding its neighborhood (i.e., χ^n). There exist many strategies exploring different neighbors configurations at different spatial positions and subbands. Nonetheless, the most thorough work studying the correlation between χ and χ^n seems to conclude that intrasubband dependencies are enough to achieve high efficiency [59]. More precisely, that work indicates that only the *immediate* neighbors of χ might be sufficient to determine probabilities reliably. This harmonizes with our goal to use only localized information of χ.

Our experience points to the consideration of the 8 or 4 immediate neighbors of χ to model probabilities. Coding efficiency is similar for both options, so for computational simplicity we consider that χ's neighborhood includes only the 4 immediate neighbors that are above, below, to the right, and left of χ. With some abuse of notation, let $\chi^n, 1 \leq n \leq N$ denote these neighbors, with $N = 4$. The main insight of our model is to assume that the magnitude of the currently encoded coefficient can be estimated through the magnitude of its neighbors, which is an assumption also used in other works (e.g., see [18], [19], [57], [59] for lossy compression, and [37], [38] for lossless compression). Let us define the *local average*, denoted as φ, of wavelet coefficients as the arithmetic mean of the neighbors' magnitudes according to

$$\varphi = \frac{1}{N} \sum_{n=1}^{N} \chi^n \ . \tag{2}$$

Though weighted averages [18], [19] have also been considered, [59] indicates that to equally weight neighbors achieves best performance. To validate that χ is correlated with φ, our first pertinent step is to compute the marginal pdf for $\chi - \varphi$, which is referred to as $g(\chi - \varphi)$ and is depicted in Figure 1(b). As shown in this figure, $g(\chi - \varphi)$ is nearly centered on 0, suggesting that most coefficients have a similar magnitude to its local average. This experimental evidence encourages the development of probability models based on $p(\chi)$ and $g(\chi - \varphi)$. Unfortunately, such an approach does not work in practice. Our previous work presented in [60], for example, relates $p(\chi)$, $g(\chi-\varphi)$ to $P_{sig}(t_j)$, $P_{ref}(t_j)$ through a mathematical framework, but practical implementations do not achieve competitive coding performance. This is caused due to $g(\chi - \varphi)$ masquerades an important feature of the signal. Let us explain further. Note in Figure 1(a) that most coefficients have magnitudes around 0. This implies that $g(\chi - \varphi)$ mostly characterizes these zero-magnitude coefficients, masquerading the less frequent coefficients that have larger magnitudes. This is a flaw for the probability model since it prevents the determination of reliable probabilities for symbols emitted in all bitplanes other than the lowest one.

The flaw of $g(\chi - \varphi)$ can be overcame considering the density of coefficients within subbands *jointly* with the fact that changes on the spatial dimension occur smoothly [56], [57]. To illustrate this point, let us assume that the neighbors of a coefficient with magnitude $\chi = x$ have magnitudes within the range $\chi^n \in (x - m, x + m)$, and that the density of the neighbors magnitude have a linear decay (i.e., the probability that a neighbor's magnitude is x' is inversely proportional to $|x - x'|$). Through this assumption, the conditional pdf for χ^n given χ may be determined as

$$f(\chi^n \mid \chi) = \begin{cases} \dfrac{\chi^n - \chi + m}{m^2} & \text{if } \chi - m < \chi^n \leq \chi \\ \\ \dfrac{|\chi^n - \chi - m|}{m^2} & \text{if } \chi < \chi^n < \chi + m \end{cases} \ , \tag{3}$$

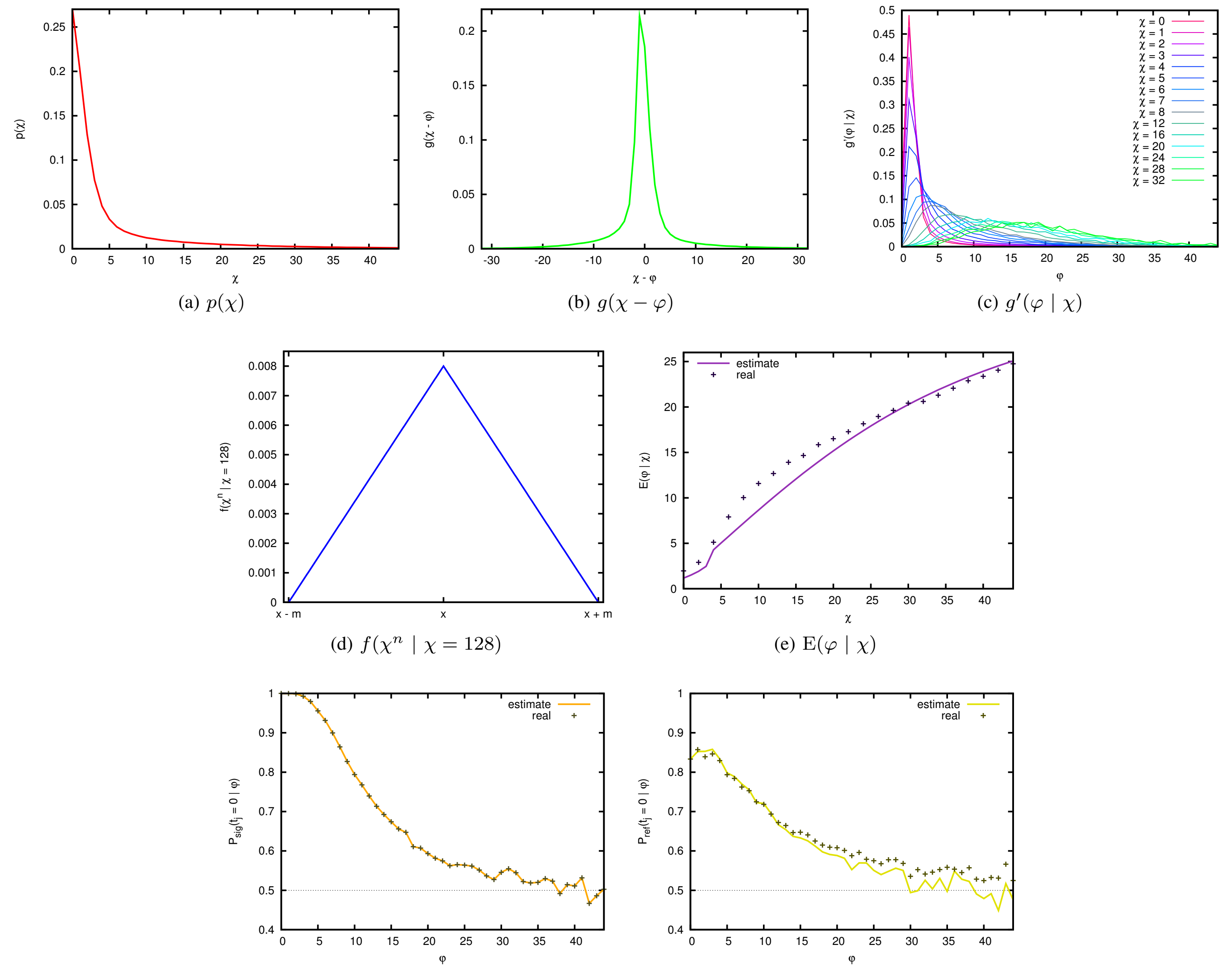
with parameter m being

$$m = \begin{cases} 4 & \text{if } \chi < 4 \\ \chi - 2 & \text{if } \chi \geq 4 \end{cases} \ . \tag{4}$$

Mathematics behind Equation (3) are described in Appendix A, and parameter m is determined empirically; the sole purpose of Equations (3) and (4) is illustrative[1]. Figure 1(d) depicts $f(\chi^n \mid \chi)$ when $\chi = 128$.

Considering the range of $\varphi \in [0, 2^K)$, the expected value of φ is computed considering $p(\chi)$ and $f(\chi^n \mid \chi)$ as

[1]Note that, generally, χ^n may be out of the range $(x - m, x + m)$. Our assumption is solely used to illustrate the nature of the wavelet signal (see below).



(a) $p(\chi)$ (b) $g(\chi - \varphi)$ (c) $g'(\varphi \mid \chi)$

(d) $f(\chi^n \mid \chi = 128)$ (e) $\mathrm{E}(\varphi \mid \chi)$

(f) $P_{sig}(t_j = 0 \mid \varphi), j = 4$ (g) $P_{ref}(t_j = 0 \mid \varphi), j^* = 4, j = 3$

Fig. 1: Statistical analysis of data within a wavelet subband. Data correspond to the high vertical-, low horizontal-frequency subband (HL) of the first decomposition level produced by an irreversible CDF 9/7 wavelet transform applied to the "Portrait" image (2048×2560, 8 bit, gray-scale) of the ISO 12640-1 corpus.

$$\mathrm{E}(\varphi \mid \chi) = \int_{max(\chi-m,0)}^{min(\chi+m,2^K)} \chi^n \cdot \frac{p(\chi^n) \cdot f(\chi^n \mid \chi)}{\int_{max(\chi-m,0)}^{min(\chi+m,2^K)} p(\chi'^n) \cdot f(\chi'^n \mid \chi)\ d\chi'^n}\ d\chi^n\ , \tag{5}$$

which is depicted in Figure 1(e) as the plot labeled "estimate". This plot is computed using the real distribution of coefficients (i.e., that data employed to plot $p(\chi)$ in Figure 1(a)), and Equations (3), (4), and (5). Note that –contrarily to what was expected when studying Figure 1(b)– the expected value of φ does *not* have a linear relation with χ, but it grows roughly logarithmically. This is intuitively explained considering that the combination of $f(\chi^n \mid \chi)$ with $p(\chi)$ results in that a coefficient with magnitude $\chi = x$ will have more coefficients with magnitudes smaller than x than coefficients with magnitudes larger than x. This causes φ to be smaller than χ except when $\chi = 0$. Graphically explained, this concept is seen placing the top of the pyramid represented in Figure 1(d) on one value of Figure 1(a). On the right side of the pyramid are less coefficients than on the left side.

The deficiency of $g(\chi - \varphi)$ to properly capture the nature of the signal is overcame with the conditional pdf for φ given χ, which is referred to as $g'(\varphi \mid \chi)$. Figure 1(c) depicts $g'(\varphi \mid \chi)$ for values of χ using real data. As opposed to $g(\chi - \varphi)$, in this case $g'(\varphi \mid \chi)$ evaluates the value of φ without subtracting χ, for the sake of clarity. When $\chi = 32$, for instance, Figure 1(c) indicates that most coefficients have a local average φ around 20, instead of 32 as might be inferred from Figure 1(b). With $g'(\varphi \mid \chi)$ the actual expected value for φ is determined as

$$\mathrm{E}'(\varphi \mid \chi) = \int_0^{2^K} \varphi \cdot g'(\varphi \mid \chi)\ d\varphi\ , \quad (6)$$

which is depicted in Figure 1(e) as the plot labeled "real". The expected value for φ using real data (i.e., $\mathrm{E}'(\varphi \mid \chi)$) and the expected value estimated through Equation (5) are very similar, which seems to validate our assumptions embodied in Equations (3) and (4).

The union of $g'(\varphi \mid \chi)$ and $p(\chi)$ in the joint pdf $h(\chi, \varphi) = p(\chi) \cdot g'(\varphi \mid \chi)$ is a suitable indicator of the signal's nature that has also been used in other works [59]. Our initial goal can now be accomplished considering this joint pdf and the local average of wavelet coefficients. Assuming that φ is known, probabilities for significance coding at bitplane j are determined as the probability of insignificant coefficients coded at bitplane j divided by all coefficients coded in that bitplane according to

$$\begin{aligned} &P_{sig}(t_j = 0 \mid \varphi) = P(\chi < 2^j \mid \chi < 2^{j+1}, \varphi) = \\ &\frac{P(\chi < 2^j \mid \varphi)}{P(\chi < 2^{j+1} \mid \varphi)} = \frac{\displaystyle\int_0^{2^j} h(\chi,\varphi)\ d\chi}{\displaystyle\int_0^{2^{j+1}} h(\chi,\varphi)\ d\chi} = \\ &\frac{\displaystyle\int_0^{2^j} p(\chi) \cdot g'(\varphi \mid \chi)\ d\chi}{\displaystyle\int_0^{2^{j+1}} p(\chi) \cdot g'(\varphi \mid \chi)\ d\chi}\ . \end{aligned} \quad (7)$$

In our practical implementation, pdfs of Equation (7) are estimated (see below), so that the coder only needs to compute φ to determine probabilities.

Similarly, probabilities for refinement bits that are emitted for coefficients that became significant at bitplane j^* and that are refined at bitplane $j^* - 1$ (i.e., the first refinement bit of coefficients in the range $[2^{j^*}, 2^{j^*+1})$) are determined according to

$$\begin{aligned} &P_{ref}(t_{j^*-1} = 0 \mid \varphi) = \\ &P(2^{j^*} \le \chi < 2^{j^*} + 2^{j^*-1} \mid 2^{j^*} \le \chi < 2^{j^*+1}, \varphi) = \\ &\frac{\displaystyle\int_{2^{j^*}}^{2^{j^*}+2^{j^*-1}} p(\chi) \cdot g'(\varphi \mid \chi)\ d\chi}{\displaystyle\int_{2^{j^*}}^{2^{j^*+1}} p(\chi) \cdot g'(\varphi \mid \chi)\ d\chi}\ . \end{aligned} \quad (8)$$

In a more general form, probabilities for refinement bits emitted at bitplane j for coefficients that became significant at bitplane j^* are determined as

$$\begin{aligned} &P_{ref}(t_j = 0 \mid \varphi) = \\ &\frac{\displaystyle\sum_{i=0}^{2^{j^*-j-1}-1} \int_{2^{j^*}+i\cdot 2^{j+1}}^{2^{j^*}+i\cdot 2^{j+1}+2^j} p(\chi) \cdot g'(\varphi \mid \chi)\ d\chi}{\displaystyle\int_{2^{j^*}}^{2^{j^*+1}} p(\chi) \cdot g'(\varphi \mid \chi)\ d\chi}\ . \end{aligned} \quad (9)$$

Figures 1(f) and 1(g) respectively depict $P_{sig}(t_j = 0 \mid \varphi)$ and $P_{ref}(t_j = 0 \mid \varphi)$ using real data of a wavelet subband. Plots labeled "estimate" report results obtained using pdfs $p(\chi)$ and $g'(\varphi \mid \chi)$ (i.e., that data employed to respectively depict Figures 1(a) and 1(c)), and computing probabilities using Equations (7) and (9). To validate that the developed framework is sound, plots labeled "real" report probabilities using the actual probability mass function of emitted symbols, which is computed empirically. Probabilities achieved by the proposed model are very similar to the actual ones. Results hold for other bitplanes, wavelet subbands, and images.

As the reader may notice, the real magnitude of coefficients has been used throughout this section. In practice, the real magnitude of coefficients is not available at the decoder until all bitplanes are transmitted. As seen in the next section, the use of partially transmitted coefficients does not degrade the performance of the proposed approach significantly.

IV. Implementation of the Local Average-Based Probability Model

A. Forms of implementation

Practical implementations may deploy the framework introduced in the previous section in different forms, such as:

1) **Ad hoc algorithm:** each wavelet transform produces a signal that, though being similar in general, it has its own particularities. The study and application of the local average approach to particular filter-banks may lead to algorithms devised for one type of transform.
2) **pdf modeling:** there are many works [57], [61], [62] that model the pdf of wavelet coefficients through a generalized Gaussian, or Gamma, distribution. The local average framework could be applied likewise modeling $p(\chi)$ and $g'(\varphi \mid \chi)$.
3) **Statistical analysis:** the extraction of statistics from a representative set of images can lead to computationally simple yet efficient implementations through, for example, the use of lookup tables.

We explored the first and second approaches in our previous works [48] and [60], respectively. The former work introduces an ad hoc algorithm for the irreversible CDF 9/7 wavelet transform that excels for its simplicity. Unfortunately, it can not be generalized to other wavelet transforms such as the reversible LeGall 5/3 wavelet transform [63], which is also included in the JPEG2000 core coding system. The latter work uses a generalized Gaussian distribution to model pdfs. Our experience seems to indicate that the modeling of $g'(\varphi \mid \chi)$ may be troublesome.

Herein, the implementation of the local average-based probability model uses the third approach. The main idea is to extract statistics from wavelet subbands belonging to images transformed using some particular filter-bank. These statistics are averaged to generate lookup tables (LUT) that capture the essence of the signal. This procedure can be used for any type of wavelet transform. LUTs are computed beforehand and they are known by coder and decoder without need to explicitly transmit them. Similar techniques are used in other scenarios [41], [56], [57].

Our approach is as follows. The pdf for wavelet coefficients (i.e., $p(\chi)$), and the conditional pdf for the local average (i.e., $g'(\varphi \mid \chi)$) are extracted for all wavelet subbands of some images, an one $p(\chi)$ and one $g'(\varphi \mid \chi)$ are generated per subband as the average of all images. For each subband, two LUTs containing probabilities $P_{sig}(t_j = 0 \mid \varphi)$ and $P_{ref}(t_j = 0 \mid \varphi)$ are computed using the averaged pdfs and Equations (7) and (9). In coding time, only φ needs to be computed to determine probabilities. The LUT for significance coding is referred to as $\mathcal{L}_{sig,b}$, with b standing for the wavelet subband to which belongs. $\mathcal{L}_{sig,b}$ contains as many rows as bitplanes are needed to represent all coefficients in that subband (i.e., K rows). For each row, there are 2^K columns representing all possible values of φ. Cells contain pre-computed probabilities, so that $P_{sig}(t_j = 0 \mid \varphi)$ is found at position $\mathcal{L}_{sig,b}[j][R(\varphi)]$, where $R(\cdot)$ is the rounding operation. The LUT for refinement coding has the same structure with an extra dimension that accounts for the bitplane at which the refined coefficient became significant. The refinement LUT is referred to as $\mathcal{L}_{ref,b}$, and it is accessed as $\mathcal{L}_{ref,b}[j^*][j][R(\varphi)]$, with j^* denoting the bitplane at which the coefficient became significant, and j denoting the current refinement bitplane.

In practice, probabilities in $\mathcal{L}_{sig,b}$ and $\mathcal{L}_{ref,b}$ can be estimated using relative frequencies conditioned on φ, avoiding the need for numerical integration. LUTs in the experimental results of Section V are generated using the eight images of the ISO/IEC 12640-1 corpus, which is compounded of natural images having different characteristics. Wavelet transforms evaluated in Section V are the irreversible CDF 9/7 wavelet transform, and the reversible LeGall 5/3 wavelet transform. The use of an extended corpus of images to generate $\mathcal{L}_{sig,b}$ and $\mathcal{L}_{ref,b}$ does not improve results significantly. To compute LUTs for each image and explicitly transmit them neither improves results significantly. Nonetheless, our experience seems to indicate that the type of image, or the sensor with which images are acquired, may produce slightly different statistical behaviors.

B. Implementation in JPEG2000 framework

The local average-based probability model is implemented in the core coding system of JPEG2000 [8]. Briefly described, this coding system is wavelet based with a two-tiered coding strategy built on the Embedded Block Coding with Optimized Truncation (EBCOT) [64]. The tier-1 stage independently codes small sets of wavelet coefficients, called codeblocks, producing an embedded bitstream for each. The tier-2 stage constructs the final codestream selecting appropriate bitstream segments, and coding auxiliary information.

The implementation of our model in JPEG2000 requires modifications in the tier-1 coding stage. This stage uses a fractional bitplane coder and the MQ coder, which is a context-adaptive arithmetic coder. The fractional bitplane coder of JPEG2000 encodes each bitplane using three coding passes: Significance Propagation Pass (SPP), Magnitude Refinement Pass (MRP), and Cleanup Pass (CP). SPP and CP are devoted to significance coding, whereas MRP refines the magnitude of significant coefficients. The difference between SPP and CP is that SPP scans those coefficients that have at least one significant neighbor, thus they are more likely to become significant than the coefficients scanned by CP, which have none significant neighbor. Tier-1 employs 19 different contexts to code symbols [23, Ch. 8.3.2]: 9 for significance coding, 3 for refinement coding, 5 for sign coding, and 2 for the run mode[2]. More information regarding JPEG2000 is found in [23].

The implementation of the local average approach in JPEG2000 is rather simple. Instead of reckon contexts, the SPP coding pass computes the local average $\hat{\varphi}$ (as defined below) for each coefficient, and seeks $P_{sig}(t_j = 0 \mid \hat{\varphi})$ in $\mathcal{L}_{sig,b}$. The MRP coding pass is modified likewise. The most important point of this practical implementation is that $\hat{\varphi}$ is computed with the magnitude of partially transmitted coefficients. Quantized neighbors at the decoder are denoted as $\hat{\chi}^n$, whereas the magnitude of χ recovered at the decoder is denoted as $\hat{\chi}$.

The use of $\hat{\chi}^n$ instead of χ^n does not impact coding efficiency except at the highest bitplanes, when little information has been transmitted and most coefficients are still quantized as 0. This penalization can be mostly avoided putting in practice assumptions embodied in Equations (3), (4), and (5), which generally predict that the magnitude of χ^n should be less than that of χ. In our implementation, when one coefficient becomes significant, or when its magnitude is refined, its 8 immediate *in*significant neighbors are considered as $\hat{\chi}'^n = \hat{\chi} \cdot \beta$ when computing the local average $\hat{\varphi}$. More precisely, $\hat{\varphi}$ is computed as

$$\hat{\varphi} = \frac{1}{N} \sum_{n=1}^{N} \begin{cases} \hat{\chi}^n & \text{if } \hat{\chi}^n \neq 0 \\ \hat{\chi}'^n & \text{otherwise} \end{cases} . \tag{10}$$

Evidently, $\hat{\chi}'^n$ is only a rough approximation of the real magnitude of the neighbor. A more precise estimate would use $\mathrm{E}(\varphi \mid \chi)$ as defined in Equation (5) since it is the expected value for the neighbors, and $\hat{\chi}'^n$ would not be limited to the 8 immediate neighbors but be expanded to farther neighbors. Unfortunately, such an approach would increase computational complexity excessively. As seen in the next section, our simplified approach achieves near-optimal performance. Experience indicates that $\beta \in [0.2, 0.6]$. It is set to $\beta = 0.4$ in our experiments.

The CP coding pass of JPEG2000 can not use a local average-based approach. CP is devised to scan large areas of

[2]The run mode is a coding primitive used by CP that helps skipping multiple insignificant coefficients with a single binary symbol.

insignificant coefficients, so for most coefficients $\hat{\varphi} = 0$. Instead of using the 9 significance contexts defined in JPEG2000, our experiments employ the model introduced in [48], which defines only two contexts according to

$$c = \begin{cases} 0 & \text{if } \sum_{n=1}^{N} \Phi(\chi^n, j) = 0 \\ 1 & \text{otherwise} \end{cases} . \quad (11)$$

This context formation approach is aimed to illustrate that simple approaches achieve high efficiency in this framework. The sign and run mode coding primitives are left as formulated in JPEG2000.

V. Experimental results

The performance of the local average-based probability model is evaluated for all images of the corpora ISO 12640-1 and ISO 12640-2 (8 bits gray-scale, size 2560×2048). We recall that LUTs are generated using images of the ISO 12640-1. Except when indicated, JPEG2000 coding parameters are: 5 levels of irreversible 9/7, or reversible 5/3 wavelet transform, codeblock size of 64×64, single quality layer codestreams, no precincts. The base quantization step sizes corresponding to bitplane 0 when the irreversible 9/7 wavelet transform is used are chosen accordingly to the L_2-norm of the synthesis basis vectors of the subband [23, Ch. 10.5.1], which is a common practice in JPEG2000. The dequantization procedure is carried out as defined in [65]. Experiments depict results for images "Portrait" and "Musicians"(corpus ISO 12640-1), and "Fishing goods" and "Japanese goods" (corpus ISO 12640-2). Similar results hold for the other images of corpora.

In the first set of experiments the coding efficiency of JPEG2000 is evaluated using four strategies to model probabilities. The first one uses the same context for all bits emitted in each coding pass, i.e., 1 context for SPP, 1 context for MRP, 1 context for CP, 1 context for sign coding, and the 2 original contexts for the run mode. This strategy is labeled "ABAC" (Adaptive Binary Arithmetic Coding). The second strategy is the context selection formulated by JPEG2000, which is considered near-optimal as a context-adaptive approach [35]. It is labeled "CABAC - JPEG2000" (Context-Adaptive Binary Arithmetic Coding). The third strategy is that formulated in Section IV-B, which employs LUTs and the local average $\hat{\varphi}$. It is labeled "LAV - LUTs, quantized magnitude". The fourth strategy reports the *theoretical* optimal performance that a local average-based approach could achieve. It uses the real magnitude of coefficients to compute φ, and the real probability mass function of emitted symbols. This fourth model is not valid in practice since real magnitudes are not available at the decoder until the whole codestream is transmitted, and because to compute the probability mass function of symbols requires a preliminary coding step with high computational cost. Nevertheless, it is useful to appraise the performance achieved by the proposed algorithm. This fourth strategy is labeled "LAV - real statistics, real magnitude".

Figures 2 and 3 provide the results achieved for the irreversible 9/7 and reversible 5/3 wavelet transforms, respectively. For each image, figures depict three graphs reporting the bitstream lengths generated by SPP, MRP, and CP coding passes. In all figures, results are presented as the difference between the evaluated strategy and JPEG2000. In each graph, each triplet of columns depicts the result achieved when all codeblocks of the image are encoded from the highest bitplane of the image to the one indicated in the horizontal axes[3]. Labels above the straight line indicate the peak signal to noise ratio (PSNR) of the image at that bitplane, and the bitrate attained by the original JPEG2000 context selection. Columns below the straight line indicate that the bitstream lengths generated for the corresponding coding pass using that strategy are shorter than those generated by JPEG2000. Results suggest that the proposed approach achieves superior coding efficiency to that of JPEG2000 for SPP and MRP coding passes, especially at medium and high bitrates and for refinement coding. The performance of "LAV" is especially good at low bitplanes since at these bitplanes most coefficients are reconstructed with high accuracy, so the model becomes very reliable. The performance achieved by the proposed algorithm is not far from the theoretically optimal one for most images. For CP coding passes, the performance achieved by the proposed model and JPEG2000 is virtually the same.

The second set of experiments evaluates the coding performance of our approach when the image is coded at target bitrates. In this case, the post-compression rate distortion (PCRD) optimization method [64] is used to minimize the distortion of the image for the targeted bitrate. "LAV" is compared to strategies "ABAC", and "CABAC - JPEG2000" as defined earlier. To better appraise the impact of "LAV" on the significance and the refinement coding primitives, figures report the performance of our approach when it is applied in both directives, labeled "LAV (SPP, MRP) + CABAC 2 CONTEXTS (CP)", and when it is applied only for refinement coding, labeled "LAV (MRP) + CABAC JPEG2000 (SPP, CP)".

Figures 4 and 5 provide the results achieved for both types of wavelet transform. Each figure depicts the PSNR difference between the evaluated strategy and JPEG2000 at different bitrates. The proposed approach improves the coding performance of JPEG2000, especially at medium and high bitrates. At low bitrates, the performance is slightly degraded compared to JPEG2000, though when "LAV" is applied only for refinement coding, the performance is virtually the same one as that of JPEG2000. It is worth noting that when the full image is coded, these experimental results indicate that the performance gain between "ABAC" and "CABAC - JPEG2000" is about the same as the performance gain between "CABAC - JPEG2000" and the local average-based approach.

The third set of experiments is aimed to illustrate another benefit of our proposal. As it is reported, previous experiments use codeblock sizes of 64×64, which is the maximum size allowed by JPEG2000, and the one that achieves the best coding performance. When the codeblock size is reduced, more

[3]We note that bitplane boundaries are suitable points to evaluate the performance of different probabilities models since the coding performance achieved at those bitrates/qualities is equivalent to that achieved by more sophisticated techniques of rate-distortion optimization at the same bitrates/qualities [66].

TABLE I: Increase on the codestream length when coding variations used for parallelization are employed. First row of each method reports bps when no coding variations are used, whereas second row of each method reports the increase on the codestream length when the RESTART and the RESET coding variations are used.

	"Portrait"	"Musicians"	"Fishing goods"	"Japanese goods"
CABAC - JPEG2000	4.07 bps	5.44 bps	3.37 bps	3.63 bps
+ RESET,RESTART	+0.07 bps	+0.08 bps	+0.06 bps	+0.07 bps
LAV	4.00 bps	5.34 bps	3.33 bps	3.57 bps
+ RESET,RESTART	+0.04 bps	+0.05 bps	+0.04 bps	+0.04 bps

auxiliary information needs to be coded, which increases the length of the final codestream. Furthermore, the use of smaller codeblocks causes that less data are fed to the arithmetic coder, producing a noticeable degradation on coding efficiency for context-adaptive strategies. This drawback does not come out with the local average-based model due to the use of stationary probabilities. Figures 6 and 7 assess the coding performance of our approach compared to that of JPEG2000 when different codeblock sizes are used, for both types of wavelet transform. These figures report the coding performance of our approach as it is described in Section IV-B. As previously, results are presented as the difference between "LAV" and JPEG2000 at different target bitrates. Results suggest that "LAV" does not penalize coding performance as much as JPEG2000 when codeblock sizes are reduced. We remark that smaller codeblocks enhance the spatial scalability of the image, which is beneficial for interactive image transmission, for instance.

The fourth set of experiments points out another benefit of the local average-based probability model. JPEG2000 provides many opportunities for parallelism. The most popular one is inter-codeblock parallelization [67], which enables the coding of codeblocks in parallel. The standard also provides coding variations to allow intra-codeblock parallelization. More precisely, the RESET and RESTART coding variations force the arithmetic coder to reset probabilities of input symbols, and to terminate the bitstream segment at the end of each coding pass, respectively. Although this enables the parallel coding of coding passes, it decreases the coding performance since symbols' probabilities need to be adapted for each coding pass. Again, the local average approach does not suffer from the break of adaptivity thanks to the use of stationary probabilities. Table I reports the bitrate (in bits per sample (bps)) achieved when the whole codestream is transmitted using our probability model and JPEG2000, with and without using the RESET and RESTART coding variations. Codeblock sizes are 32×32, and the irreversible 9/7 wavelet transform is employed in these experiments. Results indicate that "LAV" may help to reduce the impact of the RESET coding variation when intra-codeblock parallelization is used. The increase on bitrate reported by "LAV" when RESET and RESTART are employed is due to RESTART requires the coding of auxiliary information.

The last set of experiments evaluates the computational costs of our probability model. We use our JPEG2000 Part 1 implementation BOI [68], which is programmed in Java. Tests are performed on an Intel Core 2 CPU at 2 GHz, and executed on a Java Virtual Machine v1.6 using GNU/Linux v2.6. Time results report CPU processing time. The JPEG2000 context selection uses many software optimizations as suggested in [23, Ch. 17.1.2], whereas the proposed method uses a similar degree of optimization. Table II reports the time required by the tier-1 stage when decoding the full image encoded using the original JPEG2000 context tables and "LAV", for the irreversible 9/7 wavelet transform. Results indicate that the computational costs of the JPEG2000 context selection are slightly lower than those required by "LAV" (third column of the table). This shortcoming might be overcame by strategies devised to minimize computational costs, such as the ad hoc algorithm introduced in [48]. See, for example, that the computational costs of "LAV" implemented as it is suggested in [48] (fourth column of Table II) are the same as those of JPEG2000. The coding performance of [48] is virtually the same as that reported in this paper.

TABLE II: Evaluation of the time spent by the tier-1 coding stage to decode all bitplanes of the image when different probability models are employed.

	JPEG2000	LAV	Ad hoc LAV [48]
"Portrait"	2.5 secs	3.3 secs	2.5 secs
"Musicians"	3.2 secs	4.2 secs	3.2 secs
"Fishing goods"	2.1 secs	3.0 secs	2.1 secs
"Japanese goods"	2.2 secs	3.1 secs	2.2 secs

VI. Conclusions

The principal contribution of this work is a mathematical model that relates the distribution of coefficients within wavelet subbands to the probabilities of bits emitted by bitplane coding engines. The proposed model characterizes the nature of the signal produced by wavelet transforms considering the density of coefficients: 1) *globally* for the whole subband, and; 2) *locally* for small neighborhoods of wavelet coefficients. Main insights behind the proposed model are the assumption that the magnitude of a wavelet coefficient can be estimated as the arithmetic mean of the magnitude of its neighbors (the so-called local average), and that symbols emitted by bitplane coding engines are under-complete representations of the underlying signal.

A second contribution of this work is the introduction of a simple algorithm employing the principles of the local average-based model to the core coding system of the advanced image compression standard JPEG2000. The proposed algorithm does not restrain any feature of JPEG2000, and increases coding efficiency.

An important point of the local average-based probability model is that it assumes stationary statistical behavior of emitted bits. This enables the parallelization of coding bitplanes with minimum penalization on coding efficiency, and enables the use of small codeblock sizes without impact on coding efficiency. Strategies similar to those introduced herein may extend this work to other transforms, to hyper-spectral and 3D image coding, and to video or lossless coding.

Acknowledgment

The author thanks Michael W. Marcellin for his comments and remarks, which helped to improve the quality of this manuscript.

Appendix A

Our assumption is that $f(\chi^n \mid \chi)$ is linearly increasing from $\chi - m$ to χ, and linearly decreasing from χ to $\chi + m$. Therefore it can be expressed as

$$f(\chi^n \mid \chi) = \begin{cases} \alpha \cdot (\chi^n - \chi + m) & \text{if } \chi - m < \chi^n \leq \chi \\ \\ \alpha \cdot (|\chi^n - \chi - m|) & \text{if } \chi < \chi^n < \chi + m \end{cases}, \tag{12}$$

where α is the slope. Since the pdf is symmetric on χ, we can use any part of the above equation to determine α through the integral between the pdf's boundary and χ, i.e.

$$\int_{\chi - m}^{\chi} \alpha \cdot (\chi^n - \chi + m) = \frac{1}{2}\ , \tag{13}$$

which results in $\alpha = \frac{1}{m^2}$. Substituting α for $\frac{1}{m^2}$ in Equation (12) results in Equation (3).

References

[1] J. Rissanen, "Generalized Kraft inequality and arithmetic coding," *IBM Journal of Research and Development*, vol. 20, no. 3, pp. 198–203, May 1976.

[2] I. Witten, R. Neal, and J. Cleary, "Arithmetic coding for data compression," *Communications of ACM*, vol. 30, no. 6, pp. 520–540, Jun. 1987.

[3] *Digital compression and coding for continuous-tone still images*, ISO/IEC Std. 10 918-1, 1992.

[4] *Generic Coding of Moving Pictures and Associated Audio Information? Part 2: Video*, ITU-T and ISO/IEC Std. ITU-T recommendation H.262 and ISO/IEC 13 818-2, 1994.

[5] M. Slattery and J. Mitchell, "The Qx-coder," *IBM Journal of Research and Development*, vol. 42, no. 6, pp. 767–784, Nov. 1998.

[6] D. Marpe, H. Schwarz, and T. Wiegand, "Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 620–636, Jul. 2003.

[7] *Information technology - Lossy/lossless coding of bi-level images*, ISO/IEC Std. 14 492, 2001.

[8] *Information technology - JPEG 2000 image coding system - Part 1: Core coding system*, ISO/IEC Std. 15 444-1, Dec. 2000.

[9] *Advanced video coding for generic audiovisual services*, International Telecommunication Union Std. H.264, 2005.

[10] D. Huffman, "A method for the construction of minimum redundancy codes," *Proc. IRE*, vol. 40, pp. 1098–1101, 1952.

[11] S. Golomb, "Run-length encodings," *IEEE Trans. Inf. Theory*, vol. 12, no. 3, pp. 399–401, 1966.

[12] R. Rice, "Run-length encodings," *IEEE Trans. Commun.*, vol. 16, no. 9, pp. 889–897, 1971.

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

[14] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Process.*, vol. 1, no. 2, pp. 205–220, Apr. 1992.

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

[16] 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.

[17] X. Wu and J. hua Chen, "Context modeling and entropy coding of wavelet coefficients for image compression," in *Proc. IEEE International Conference Acoustics, Speech, and Signal Processing*, vol. 4, Apr. 1997, pp. 3097–3100.

[18] C. Chrysafis and A. Ortega, "Efficient context-based entropy coding for lossy wavelet image compression," in *Proc. IEEE Data Compression Conference*, Mar. 1997, pp. 241–250.

[19] Y. Yoo, A. Ortega, and B. Yu, "Image subband coding using context-based classification and adaptive quantization," *IEEE Trans. Image Process.*, vol. 8, no. 12, pp. 1702–1715, Dec. 1999.

[20] M. J. Weinberger, J. J. Rissanen, and R. B. Arps, "Applications of universal context modeling to lossless compression of gray-scaled images," *IEEE Trans. Image Process.*, vol. 5, no. 4, pp. 575–586, Apr. 1996.

[21] X. Wu and N. Memon, "Context-based, adaptive, lossless image coding," *IEEE Trans. Commun.*, vol. 45, no. 4, pp. 437–444, Apr. 1997.

[22] D. Taubman, "Remote browsing of JPEG2000 images," in *Proc. IEEE International Conference on Image Processing*, Sep. 2002, pp. 229–232.

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

[24] S.-T. Hsiang, "Highly scalable subband/wavelet image and video coding," Ph.D. dissertation, Rensselaer Polytechnic Institue, Troy, NY, Jan. 2002.

[25] G. Menegaz and J.-P. Thiran, "Three-dimensional encoding/two-dimensional decoding of medical data," *IEEE Trans. Med. Imag.*, vol. 22, no. 3, pp. 424–440, Mar. 2003.

[26] Y. Yuan and M. Mandal, "Novel embedded image coding algorithms based on wavelet difference reduction," *IEE Proceedings- Vision, Image and Signal Processing*, vol. 152, no. 1, pp. 9–19, Feb. 2005.

[27] P. Lamsrichan and T. Sanguankotchakorn, "Embedded image coding using context-based adaptive wavelet difference reduction," in *Proc. IEEE International Conference on Image Processing*, Oct. 2006, pp. 1137–1140.

[28] X. Wu, "Context quantization with Fisher discriminant for adaptive embedded wavelet image coding," in *Proc. IEEE Data Compression Conference*, Mar. 1999, pp. 102–111.

[29] J. Chen, Y. Zhang, and X. Shi, "Image coding based on wavelet transform and uniform scalar dead zone quantizer," *ELSEVIER Signal Processing: Image Communication*, vol. 21, no. 7, pp. 562–572, Aug. 2006.

[30] R. Singh and A. Ortega, "Reduced-complexity delayed-decision algorithm for context-based image processing systems," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 1937–1945, Aug. 2007.

[31] A. J. R. Neves and A. J. Pinho, "Lossless compression of microarray images using image-dependent finite-context models," *IEEE Trans. Med. Imag.*, vol. 28, no. 2, pp. 194–201, Feb. 2009.

[32] X. Wu, P. A. Chou, and X. Xue, "Minimum conditional entropy context quantization," in *Proc. IEEE International Symposium on Information Theory*, Jun. 2000, pp. 43–43.

[33] S. Forchhammer, X. Wu, and J. D. Andersen, "Lossless image data sequence compression using optimal context quantization," in *Proc. IEEE Data Compression Conference*, Mar. 2001, pp. 53–62.

[34] J. Chen, "Context modeling based on context quantization with application in wavelet image coding," *IEEE Trans. Image Process.*, vol. 13, no. 1, pp. 26–32, Jan. 2004.

[35] Z. Liu and L. J. Karam, "Mutual information-based analysis of JPEG2000 contexts," *IEEE Trans. Image Process.*, vol. 14, no. 4, pp. 411–422, Apr. 2005.

[36] M. Cagnazzo, M. Antonini, and M. Barlaud, "Mutual information-based context quantization," *ELSEVIER Signal Processing: Image Communication*, vol. 25, no. 1, pp. 64–74, Jan. 2010.

[37] A. Ruedin and D. Acevedo, "A class-conditioned lossless wavelet-based predictive multispectral image compressor," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 1, pp. 166–170, Jan. 2010.

[38] H. Wang, S. D. Babacan, and K. Sayood, "Lossless hyperspectral-image compression using context-based conditional average," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 4187–4193, Dec. 2007.

[39] L.-J. Kau, Y.-P. Lin, and C.-T. Lin, "Lossless image coding using adaptive, switching algorithm with automatic fuzzy context modelling," *IEE Proceedings- Vision, Image and Signal Processing*, vol. 153, no. 5, pp. 684–694, Oct. 2006.

[40] M. J. Weinberger, G. Seroussi, and G. Sapiro, "The LOCO-I lossless image compression algorithm: Principles and standardization into JPEG-LS," *IEEE Trans. Image Process.*, vol. 9, no. 8, pp. 1309–1324, Aug. 2000.

[41] Y. Zhang and D. A. Adjeroh, "Prediction by partial approximate matching for lossless image compression," *IEEE Trans. Image Process.*, vol. 17, no. 6, pp. 924–935, Jun. 2008.

[42] *JPEG-LS lossless and near-lossless compression for continuous-tone still images*, ISO/IEC Std. 14 495-1, 1999.

[43] X. Wu and N. Memon, "Context-based lossless interband compression-extending CALIC," *IEEE Trans. Image Process.*, vol. 9, no. 6, pp. 994–1001, Jun. 2000.

[44] J. Zhang and G. Liu, "A novel lossless compression for hyperspectral images by context-based adaptive classified arithmetic coding in wavelet domain," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 3, pp. 461–465, Jul. 2007.

[45] V. Sanchez, R. Abugharbieh, and P. Nasiopoulos, "Symmetry-based scalable lossless compression of 3D medical image data," *IEEE Trans. Med. Imag.*, vol. 28, no. 7, pp. 1062–1072, Jul. 2009.

[46] D. Taubman and A. Zakhor, "Multirate 3D subband coding of video," *IEEE Trans. Image Process.*, vol. 3, no. 5, pp. 572–588, Sep. 1994.

[47] L. Zhang, X. Wu, N. Zhang, W. Gao, Q. Wang, , and D. Zhao, "Context-based arithmetic coding reexamined for DCT video compression," in *Proc. IEEE International Symposium on Circuits and Systems*, May 2007, pp. 3147–3150.

[48] F. Auli-Llinas, "Local average-based model of probabilities for JPEG2000 bitplane coder," in *Proc. IEEE Data Compression Conference*, Mar. 2010, pp. 59–68.

[49] P. Simard, D. Steinkraus, and H. Malvar, "On-line adaptation in image coding with a 2-D tarp filter," in *Proc. IEEE Data Compression Conference*, Apr. 2002, pp. 23–32.

[50] C. Tian and S. S. Hemami, "An embedded image coding system based on tarp filter with classification," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, May 2004, pp. 17–21.

[51] J. Zhang, J. E. Fowler, and G. Liu, "Lossy-to-lossless compression of hyperspectral imagery using three-dimensional TCE and an integer KLT," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 814–818, Oct. 2008.

[52] A. T. Deever and S. S. Hemami, "Efficient sign coding and estimation of zero-quantized coefficients in embedded wavelet image codecs," *IEEE Trans. Image Process.*, vol. 12, no. 4, pp. 420–430, Apr. 2003.

[53] O. Lopez, M. Martinez, P.Pinol, M.P.Malumbres, and J. Oliver, "E-LTW: an enhanced LTW encoder with sign coding and precise rate control," in *Proc. IEEE International Conference on Image Processing*, Nov. 2009, pp. 2821–2824.

[54] S. Kim, J. Jeong, Y.-G. Kim, Y. Choi, and Y. Choe, "Direction-adaptive context modeling for sign coding in embedded wavelet image coder," in *Proc. IEEE International Conference on Image Processing*, Nov. 2009, pp. 2809–2812.

[55] I. Daubechies, *Ten lectures on wavelets*. Philadelphia, PA: SIAM, 1992.

[56] S. M. LoPresto, K. Ramchandran, and M. T. Orchard, "Image coding based on mixture modeling of wavelet coefficients and a fast estimation-quantization framework," in *Proc. IEEE Data Compression Conference*, Mar. 1997, pp. 221–230.

[57] 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.

[58] Z. He and S. K. Mitra, "A unified rate-distortion analysis framework for transform coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 12, pp. 1221–1236, Dec. 2001.

[59] J. Liu and P. Moulin, "Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients," *IEEE Trans. Image Process.*, vol. 10, no. 11, pp. 1647–1658, Nov. 2001.

[60] F. Auli-Llinas, I. Blanes, J. Bartrina-Rapesta, and J. Serra-Sagrista, "Stationary model of probabilities for symbols emitted by bitplane image coders," in *Proc. IEEE International Conference on Image Processing*, pp. 497–500, Sep. 2010.

[61] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using Generalized Gaussian density and Kullback-Leibler distance," *IEEE Trans. Image Process.*, vol. 11, no. 2, pp. 146–158, Feb. 2002.

[62] S. K. Choy and C. S. Tong, "Statistical wavelet subband characterization based on generalized Gamma density and its application in texture retrieval," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 281–289, Feb. 2010.

[63] A. Calderbank, I. Daubechies, W. Sweldens, and B.-L. Yeo, "Lossless image compression using integer to integer wavelet transforms," in *Proc. IEEE International Conference on Image Processing*, vol. 1, Oct. 1997, pp. 596–599.

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

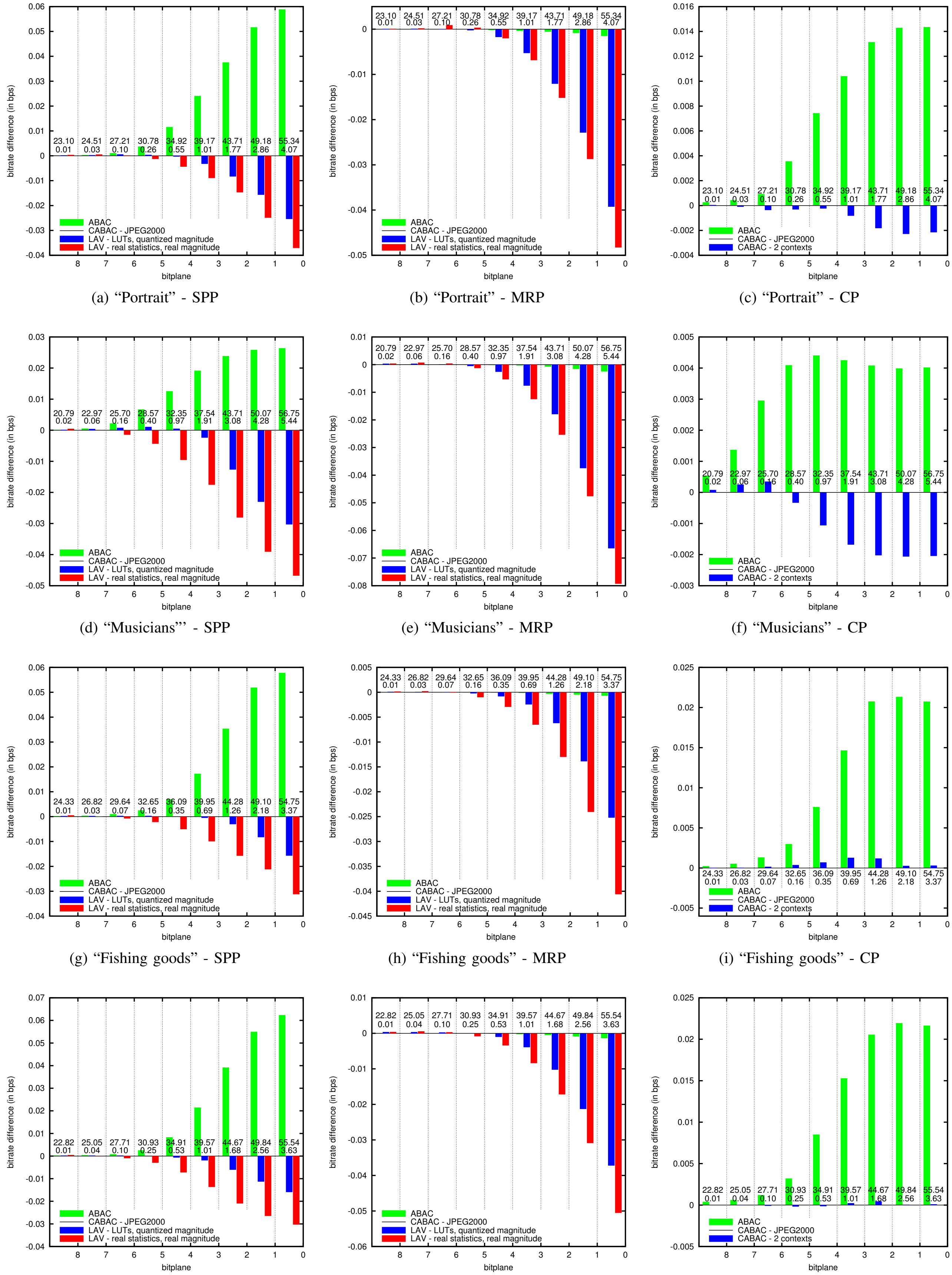
[65] F. Auli-Llinas and M. W. Marcellin, "Distortion estimators for bitplane image coding," *IEEE Trans. Image Process.*, vol. 18, no. 8, pp. 1772–1781, Aug. 2009.

[66] 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.

[67] M. Dyer, S. Nooshabadi, and D. Taubman, "Design and analysis of system on a chip encoder for JPEG2000," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 2, pp. 215–225, Feb. 2009.

[68] F. Auli-Llinas. (2010, Oct.) BOI software. [Online]. Available: http://www.deic.uab.es/~francesc

Francesc Aulí-Llinàs (S'2006-M'2008) is a researcher within the recruitment program Ramón y Cajal funded by the Spanish Government. He is currently working with the Group on Interactive Coding of Images of the Department of Information and Communications Engineering, at the Universitat Autònoma de Barcelona (Spain). He received the B.Sc. and B.E. degrees in Computer Management Engineering and Computer Engineering in 2000 and 2002, respectively, both with highest honors from the Universitat Autnoma de Barcelona (Spain), and for which he was granted with two extraordinary awards of Bachelor. In 2004 and 2006 he respectively received the M.S. degree and the Ph.D. degree (with honors), both in Computer Science from the Universitat Autònoma de Barcelona. Since 2002 he has been consecutively awarded with doctoral and postdoctoral fellowships in competitive calls. From 2007 to 2009 he carried out two research stages of one year each with the group of David Taubman, at the University of New South Wales (Australia), and with the group of Michael Marcellin, at the University of Arizona (USA). 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. He is involved in several research projects granted by the European Union, and the Spanish and Catalan Governments. His research interests include a wide range of image coding topics, including highly scalable image and video coding systems, rate-distortion optimization, distortion estimation, and interactive transmission, among others.



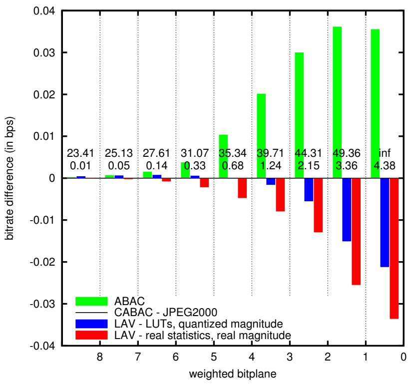
(a) “Portrait” - SPP (b) “Portrait” - MRP (c) “Portrait” - CP

(d) “Musicians”’ - SPP (e) “Musicians” - MRP (f) “Musicians” - CP

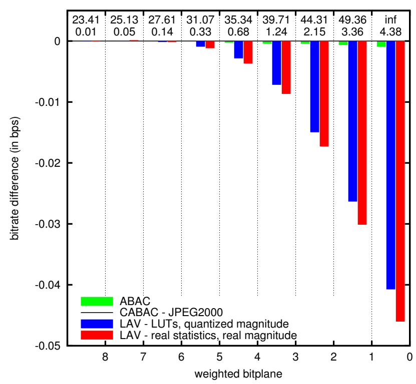
(g) “Fishing goods” - SPP (h) “Fishing goods” - MRP (i) “Fishing goods” - CP

(j) “Japanese goods” - SPP (k) “Japanese goods” - MRP (l) “Japanese goods” - CP

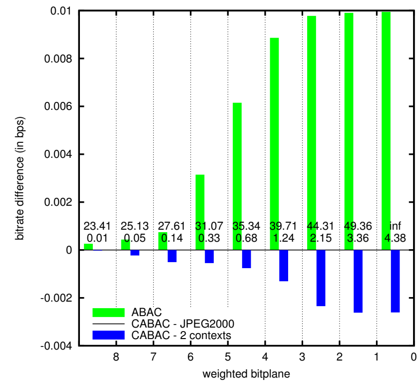
Fig. 2: Evaluation of the coding passes lengths generated by different probability models. Results are reported for the irreversible 9/7 wavelet transform. Results are cumulative from the highest bitplane to the one reported in the horizontal axis.



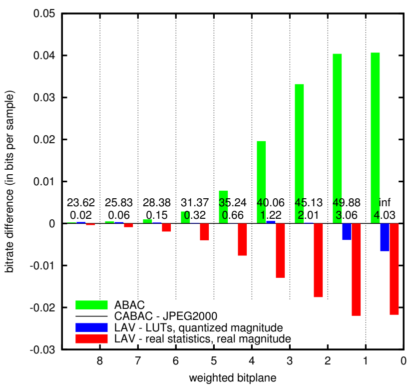
(a) “Portrait” - SPP



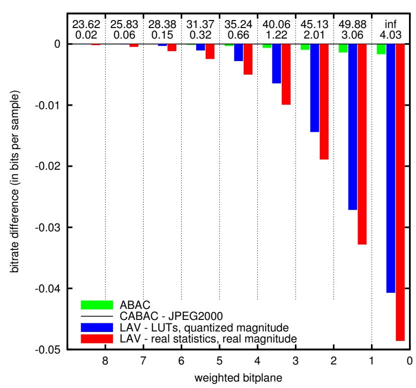
(b) “Portrait” - MRP



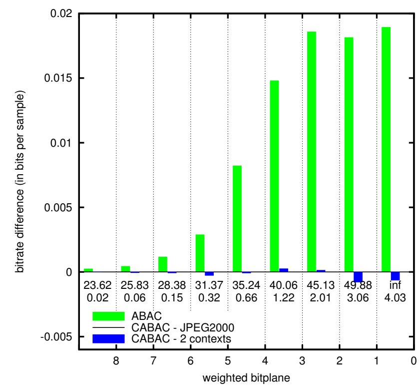
(c) “Portrait” - CP



(d) “Japanese goods” - SPP

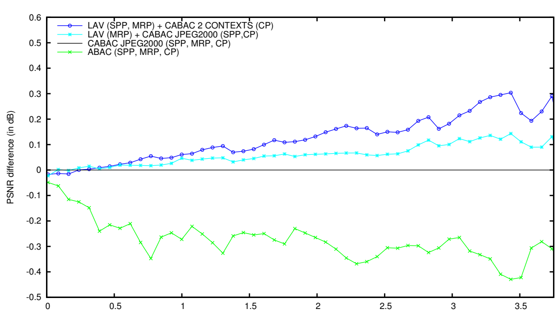


(e) “Japanese goods” - MRP

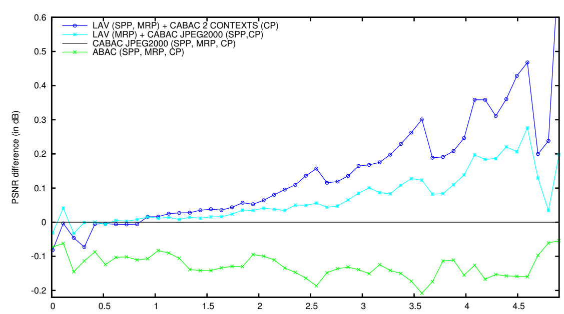


(f) “Japanese goods” - CP

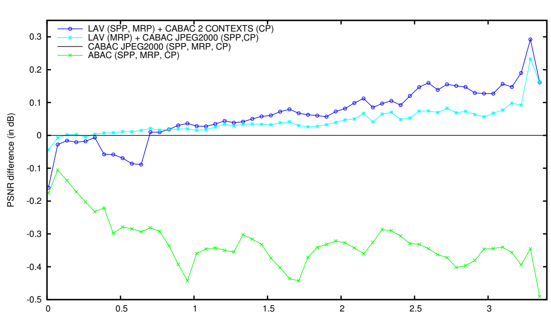
Fig. 3: Evaluation of the coding passes lengths generated by different probability models. Results are reported for the reversible 5/3 wavelet transform. Results are cumulative from the highest bitplane to the one reported in the horizontal axis.



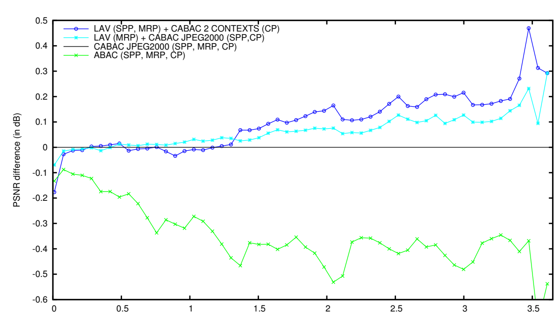
(a) “Portrait”



(b) “Musicians”

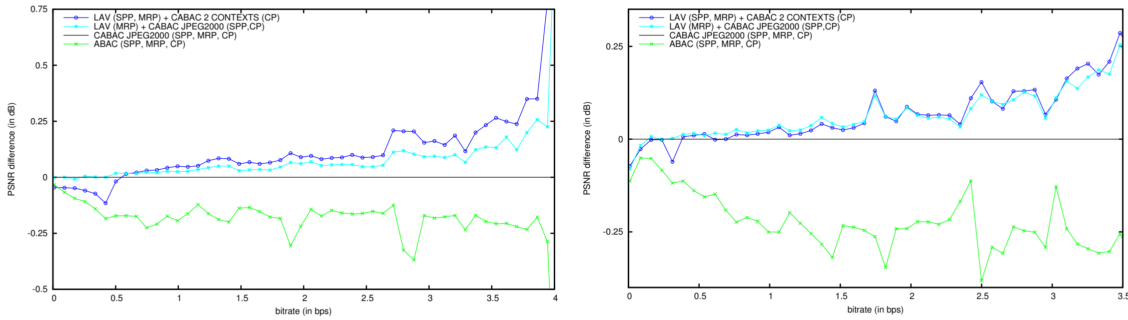


(c) “Fishing goods”



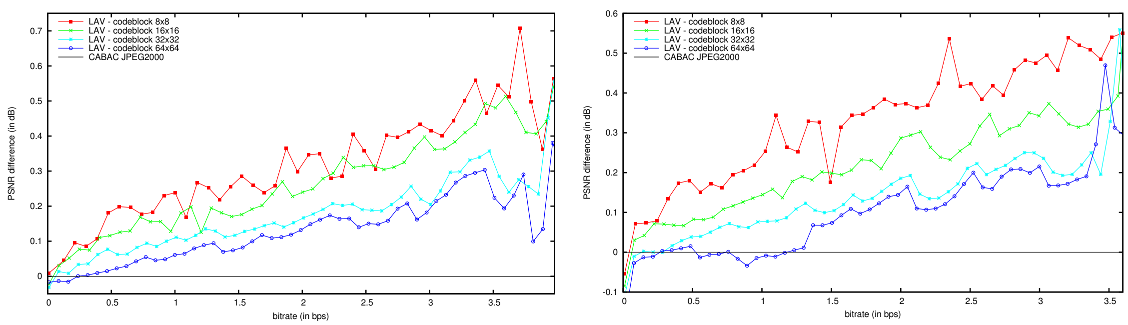
(d) “Japanese goods”

Fig. 4: Evaluation of the coding performance achieved by different probability models. Results are reported for the irreversible 9/7 wavelet transform.



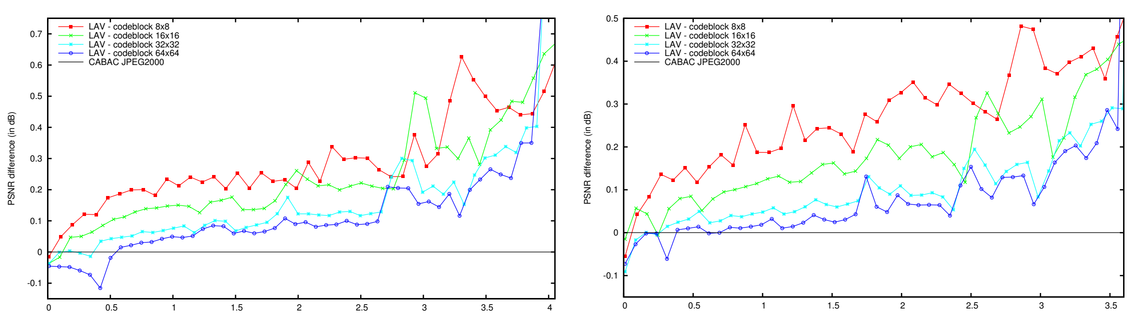
(a) "Portrait" (b) "Japanese goods"

Fig. 5: Evaluation of the coding performance achieved by different probability models. Results are reported for the reversible 5/3 wavelet transform.



(a) "Portrait" (b) "Japanese goods"

Fig. 6: Evaluation of the coding performance achieved with different codeblock sizes. Results are reported for the irreversible 9/7 wavelet transform. The PSNR difference between "LAV" and "CABAC - JPEG2000" is computed for each codeblock size independently, i.e., the straight horizontal line reports different performance of "CABAC - JPEG2000" for each codeblock size.



(a) "Portrait" (b) "Japanese goods"

Fig. 7: Evaluation of the coding performance achieved with different codeblock sizes. Results are reported for the reversible 5/3 wavelet transform. The PSNR difference between "LAV" and "CABAC - JPEG2000" is computed for each codeblock size independently, i.e., the straight horizontal line reports different performance of "CABAC - JPEG2000" for each codeblock size.