

Entropy-based Evaluation of Context Models for Wavelet-transformed Images

Francesc Aulí-Llinàs, *Senior Member, IEEE*

Abstract—Entropy is a measure of a message uncertainty. Among others aspects, it serves to determine the minimum coding rate that practical systems may attain. This work defines an entropy-based measure to evaluate context models employed in wavelet-based image coding. The proposed measure is defined considering the mechanisms utilized by modern coding systems. It establishes the maximum performance achievable with each context model. This helps to determine the adequateness of the model under different coding conditions, and serves to predict with high precision the coding rate achieved by practical systems. Experimental results evaluate four well-known context models using different types of images, coding rates, and transform strategies. They reveal that, under specific coding conditions, some widely-spread context models may not be as adequate as it is generally thought. The hints provided by this analysis may help to design simpler and more efficient wavelet-based image codecs.

Index Terms—Context models, image entropy, wavelet transform, bitplane image coding, JPEG2000.

I. INTRODUCTION

ENTROPY is an information-theory measure that defines the uncertainty, or randomness, of a message. Since its inception in 1948 by Claude E. Shannon [1], the entropy has been employed in myriad fields such as data compression and communication, cryptography, or statistical inference, among others. In data compression, the entropy of a message establishes a limit for its lossless compression. Practical coding systems may generate a compacted representation of the message with a length that may approach –but never be shorter– than its entropy. The entropy may serve to multiple purposes such as the evaluation of a coding system efficiency, the characterization of different types of signals, or the prediction of the rate required to transmit a message.

In the field of image and video coding, the entropy and entropy-related measures have been used extensively. Without aiming to be exhaustive, the coefficients of a wavelet-transformed image are modeled through relative entropy (i.e., Kullback–Leibler divergence) in [2], and through conditional and relative entropy in [3]–[5]. Video sequences decorrelated with the wavelet transform are modeled using the entropy and the conditional entropy in [6], [7]. [8] and [9] describe rate-distortion optimization methods for image and video

codecs employing entropy-based models, and [10] utilizes the entropy to estimate the performance of Wyner-Ziv coding. More recently, the entropy of wavelet coefficients has been employed to evaluate scanning order strategies [11] and embedded quantization schemes [12].

This work defines an entropy-based measure to evaluate context models employed to code wavelet-transformed images. The context model is a fundamental aspect of an image coding scheme that greatly influences its performance [2], [11], [13]–[17]. In general, the context models are devised for a specific coding engine, transform strategy, or image type. To the best of our knowledge, there is no work evaluating context models in a general manner. Our aim is to define a framework in which context models can be compared under different coding conditions, including various types of wavelet filters, transform strategies, coding rates, and images belonging to different fields. An entropy-based measure is ideal to do so since it establishes the minimum coding rate achievable with each model, ascertaining its adequateness for particular coding conditions.

The proposed measure is defined mimicking the mechanisms employed by wavelet-based coding schemes. Four context models conceived from well-known approaches in the literature are evaluated. The experimental results reveal interesting aspects that may serve to improve the performance and to simplify the complexity of future image codecs. They also show that the proposed measure predicts with precision the coding rate produced by practical systems.

Objectives similar to ours have been previously pursued in the literature in different forms. The purpose of [18], for example, is to ascertain whether the efficiency of wavelet-based codecs can be enhanced by means of context models that exploit intra- or inter-scale dependencies. [19] uses mutual information to determine contexts that achieve minimum coding rate. [20] describes a rate model for wavelet-based codecs that achieves high accuracy without needing to encode the image. The main difference between this research and previous work is that the proposed measure is applied with different context models that exploit various types of dependencies of wavelet data, is computed for different levels of quality to obtain an accurate rate-distortion analysis of the image, and it allows direct comparison with the coding performance achieved by current codecs.

The paper is structured as follows. Section II briefly overviews wavelet-based coding schemes. Section III defines the entropy-based measure and describes four context models that can be used with the proposed measure. Section IV evaluates these context models when coding three different types of images under different conditions. Section V concludes this

paper providing some remarks.

II. OVERVIEW OF WAVELET-BASED CODING SCHEMES

Typically, wavelet-based coding schemes are structured in three main stages. The first stage decorrelates spatial redundancies by means of a wavelet transform that decomposes the image in different subbands, forming a dyadic structure [21]. Such a decomposition produces three subbands per decomposition level except at the highest level, which contains four subbands. Subbands capture the low or high vertical and horizontal frequencies of the image. They are denoted by LL, HL, LH , or HH with the first and second L/H referring to low, or high, vertical and horizontal frequencies captured in the subband, respectively. After the wavelet transform, the coefficients are quantized with a uniform deadzone scalar quantizer with step size Δ . The quantization operation is expressed as $x = \left\lfloor \frac{|w|}{\Delta} \right\rfloor$, with w denoting the wavelet coefficient and $\lfloor \cdot \rfloor$ denoting the floor operation. For simplicity, in this paper Δ is assumed to be the same for all wavelet subbands, though other quantization schemes might also be employed.

The second stage of the coding scheme utilizes an engine that exploits psycho-visual redundancies to progressively reduce the image distortion. To do so, most systems employ a bitplane coding strategy [11]. Let $[b_{M-1}, b_{M-2}, \dots, b_1, b_0]$, $b_i \in \{0, 1\}$ be the binary representation for the quantization index x , with M denoting a sufficient number of bits to represent all coefficients. Let $d \in \{+, -\}$ be the sign of w . Bitplane coding strategies define bitplane j as the collection of bits b_j from all coefficients, and encode coefficients from the most significant bitplane $M-1$ to the least significant bitplane 0. The first non-zero bit of a coefficient, i.e., that $b_s = 1$ such that $\nexists s' > s$ with $b_{s'} = 1$, is called the significant bit of the coefficient. The sign of the coefficient is coded immediately after its significant bit. The remaining bits $b_r, r < s$ are called refinement bits.

The third stage of a typical wavelet-based coding scheme is aimed at the reduction of statistical redundancies produced by the bitplane coding engine. Arithmetic coding [22] is among the most efficient techniques to do so. Key to achieve competitive performance is to employ a model of probabilities that predicts with accuracy the symbols emitted by the coding engine. Generally, such models determine probabilities employing the so-called context of the coefficient, which is obtained considering the significance state (see below), the magnitude, the sign, or other relevant features of some neighbors of the coefficient coded. The context helps determining the probability of the index to become significant in the current bitplane, its refinement bits, and its sign.

III. PROPOSED MEASURE

A. Definition

First, we define the order-zero entropy of the coefficients in wavelet subband v . Let us denote the marginal probability mass function (pmf) of the quantization indices and the coefficients signs in v as $P_{X_v}(x)$ and $P_{D_v}(d)$, respectively.

The order-zero entropy of the data in v is computed as the entropy of the indices plus the weighted entropy of the signs, i.e.,

$$H(v) = H(X_v) + H(D_v), \quad (1)$$

with

$$H(X_v) = - \sum_x P_{X_v}(x) \log_2 P_{X_v}(x), \quad (2)$$

and

$$H(D_v) = - \left(\sum_d P_{D_v}(d) \log_2 P_{D_v}(d) \right) \cdot \frac{\|x \neq 0 \in v\|}{\|x \in v\|}. \quad (3)$$

$\|x \in v\|$ and $\|x \neq 0 \in v\|$ in (3) denote the number of indices (i.e., all quantized coefficients), and the number of indices different from 0, in subband v , respectively. (3) multiplies the entropy of the signs by the percentage of indices different from 0 because the sign needs only to be transmitted for non-zeros. The order-zero entropy of the transformed image is then computed as the weighted sum of the subbands' entropy as

$$H(V) = \sum_v H(v) \cdot \frac{\|x \in v\|}{\|x\|}, \quad (4)$$

with $\|x\|$ denoting the number of quantization indices of the whole image. This definition of the zero-order entropy of the image is similarly derived in [4].

As previously stated, bitplane coding strategies do *not* code wavelet coefficients at once. They progressively refine the quantization indices in each bitplane. It is of interest to analyze whether the bit-by-bit transmission of the indices affects their entropy. Instead of computing $H(X_v)$ as in (2), the entropy obtained when the indices are transmitted bit by bit is computed as

$$\begin{aligned} \hat{H}(X_v) = & - \sum_x P_{X_v}(x) \cdot [\log_2 P_{B_v}(b_{M-1}) + \\ & \log_2 P_{B_v}(b_{M-2} \mid b_{M-1}) + \dots + \\ & \log_2 P_{B_v}(b_0 \mid b_1, b_2, \dots, b_{M-1})] . \end{aligned} \quad (5)$$

$P_{B_v}(b_k)$ above is the marginal pmf of bits b_k from all quantization indices in subband v . The sum of logarithms in (5) accounts for the number of bits necessary to transmit index x when its binary representation is transmitted bit by bit. By expressing $P_{B_v}(b_k)$ through $P_{X_v}(x)$, this sum of logarithms is expressed as

$$\begin{aligned}
& \log_2 P_{B_v}(b_{M-1}) + \log_2 P_{B_v}(b_{M-2} | b_{M-1}) + \dots + \\
& \log_2 P_{B_v}(b_0 | b_1, b_2, \dots, b_{M-1}) = \\
& \log_2 P_{B_v}(b_{M-1}) + \log_2 \frac{P_{B_v}(b_{M-2}, b_{M-1})}{P_{B_v}(b_{M-1})} + \dots + \\
& \log_2 \frac{P_{B_v}(b_0, b_1, b_2, \dots, b_{M-1})}{P_{B_v}(b_1, b_2, \dots, b_{M-1})} = \\
& \log_2 \frac{\sum_{x=0}^{2^{M-1}-1} P_{X_v}(x)}{\sum_{x=0}^{2^{M-1}-1} P_{X_v}(x)} + \log_2 \frac{\sum_{x=0}^{2^{M-2}-1} P_{X_v}(x)}{\sum_{x=0}^{2^{M-1}-1} P_{X_v}(x)} + \dots + \\
& \log_2 \frac{\sum_{x=0}^{2^0-1} P_{X_v}(x)}{\sum_{x=0}^{2^1-1} P_{X_v}(x)} = \log_2 \frac{\sum_{x=0}^{2^0-1} P_{X_v}(x)}{\sum_{x=0}^{2^{M-1}-1} P_{X_v}(x)} = \\
& \log_2 P_{B_v}(b_0, b_1, b_2, \dots, b_{M-1}) = \log_2 P_{X_v}(x=0)
\end{aligned} \tag{6}$$

when $x = 0$, and in $\log_2 P_{X_v}(x)$ in general. Although the bit-by-bit transmission of the indices have practical advantages, (6) shows that, in terms of entropy, the transmission of the indices bit-by-bit or at once is equivalent. So (2) and (5) are considered indistinctly herein.

The order-zero entropy of the transformed image does not consider the context-modeling mechanisms employed by current coding systems. Instead of using marginal pmfs, the context model is introduced into the measure through the conditional pmfs of the indices and signs, which are referred to as $P'_{X_v}(x | \Theta_x)$ and $P'_{D_v}(d | \Theta_d)$, respectively. Θ_x and Θ_d denote the context employed for the indices and for the signs, respectively. The conditional entropy of the data in subband v is expressed as

$$\begin{aligned}
H'(v) = & - \sum_x P'_{X_v}(x | \Theta_x) \log_2 P'_{X_v}(x | \Theta_x) \\
& - \left(\sum_d P'_{D_v}(d | \Theta_d) \log_2 P'_{D_v}(d | \Theta_d) \right) \cdot \frac{\|x \neq 0 \in v\|}{\|x \in v\|}
\end{aligned} \tag{7}$$

The conditional entropy of the image, denoted by $H'(V)$, is derived as in (4) but employing $H'(v)$ instead of $H(v)$. The conditional bit-by-bit entropy is denoted by $\hat{H}'(X_v)$, and is derived as in (5) but replacing $P_{B_v}(b_i)$ by $P'_{B_v}(b_i | \Theta_x)$. Again, $\hat{H}'(X_v) = H'(X_v)$.

$H'(V)$ employs contexts Θ_x and Θ_d . As seen below, these contexts intentionally require the complete quantized representation of the neighbors of the coefficient. In a practical system,

the neighbors are not fully reconstructed at the decoder until the transmission ends, so contexts are commonly computed employing partial representations of the coefficients [13], [17]. Contrarily to practical systems, Θ_x and Θ_d employ the complete representation of the coefficients because the proposed measure is aimed at the prediction of the maximum compression performance attainable in practice. This is, at most, real codecs can obtain estimates of the contexts as good as Θ_x , Θ_d . The better the estimates, the higher the coding performance, with $H'(V)$ representing the maximum efficiency.

$H'(V)$ is computed using a wavelet-transformed image in which the energy gain factor of all subbands is set to 1. This is achieved by multiplying wavelet data by the L_2 -norm of the synthesis basis vectors of the subband's filter-bank. Then, the quality of the image that is obtained when reversing the wavelet transform is controlled through the quantization step size Δ , which is the same for all subbands. Large Δ s achieve low quality images due to rough quantization, while small Δ s achieve high quality images. The rate-distortion analysis for an image is obtained computing its entropy using different Δ s. In the experiments of Section IV, the irreversible 9/7 wavelet filter-bank is employed with $\Delta \in [1, 1000]$, which achieves Peak Signal to Noise Ratios (PSNR) ranging from 20 to 55 dB for natural images, approximately. The proposed entropy-based measure can also be employed with quantization schemes that employ a different Δ for each subband, though then the subbands' energy gain factors must be considered accordingly.

To control the image quality, or the coding rate, through Δ achieves results similar to those achieved with more sophisticated methods of rate-distortion optimization like, for instance, that described in [13]. This is due to the rate-distortion properties of wavelet data and bitplane coding. As appraised in [11], [23], the truncation of all quantization indices at the same bitplane or, equivalently, the transmission of all quantized coefficients using the same Δ , achieves near-optimal performance for that rate, or quality.

B. Context models

The efficiency of an image coding system greatly depends on the context model. Effective context models produce conditional pmfs with highly skewed probabilities, resulting in low coding rates. Four different context models summoning most of the approaches employed in the literature are described herein. In what follows, the symbol Θ is substituted by σ , ϕ , φ , or ψ depending on the context model that it refers to. For each context model, three different types of dependencies of wavelet data are evaluated. Superindices INTRA, INTER, and 3D are added to the context symbol to denote the type of dependency. INTRA considers intra-scale dependencies, so only coefficients within the same subband are utilized to determine the context. INTER considers inter-scale dependencies, employing coefficients of the subband that are at the immediately higher decomposition level. 3D considers 3D dependencies in images that have more than one component, utilizing coefficients at the same subband but at the previous,

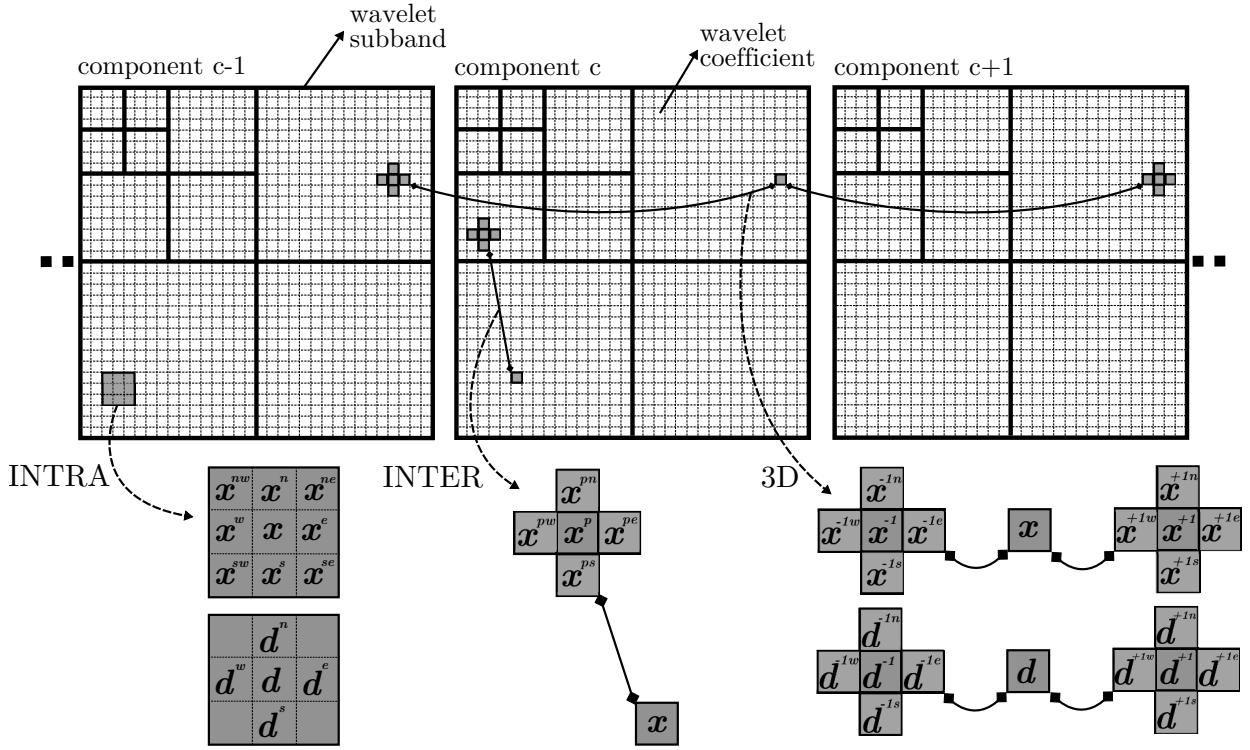


Fig. 1: Coefficients considered by the three types of dependencies evaluated when coding index x and sign d . INTRA employs adjacent coefficients within the same subband, INTER utilizes coefficients of the same subband but at the immediately higher decomposition level, and 3D uses coefficients at the same spatial position but of the previous, or next, component. The sign is not considered by INTER since little correlation exists between signs of different subbands [24]–[29].

or at the next, component. Fig. 1 depicts the coefficients utilized by these three types of dependencies. The figure also provides the notation employed to refer them.

The first context model is the simplest. It is denoted by σ and considers the index and the sign of only one coefficient, more precisely, the left/parent/previous coefficient for the intra/inter/3D dependency type. σ_x and σ_d are equal to the corresponding index or sign, respectively. For example, $\sigma_x^{INTRA} = x^w$, and $\sigma_d^{INTRA} = d^w$. Table I summarizes the coefficients employed to compute this and following contexts.

The second context model, denoted by ϕ , is borrowed from the JPEG2000 standard [30]. The main idea behind the JPEG2000 context model is to use the significance state of the indices and the signs of the coefficients. The significance state of coefficient x at bitplane j is 1 if its significance bit is already coded, and 0 otherwise, i.e.,

$$S(x) = \begin{cases} 1 & \text{if } j \leq s \\ 0 & \text{otherwise} \end{cases}. \quad (8)$$

Function $f_x(\cdot)$ then employs the significance state of the eight adjacent neighbors of the coefficient to determine nine different contexts. These nine contexts are devoted to significance coding. A function of the signs of the four adjacent neighbors, denoted by $f_d(\cdot)$, determines five different contexts devoted to sign coding. More details on the JPEG2000 context model can be found in [30]. Herein, ϕ_x^{INTRA} employs the same function

$f_x(\cdot)$ as that defined in JPEG2000 but slightly modifying the definition of the significance state to

$$S'(x) = \begin{cases} 1 & \text{if } x \neq 0 \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

so that it can be used in the conditional pmf $P'_{X_v}(x | \phi_x^{INTRA})$. For sign coding, ϕ_d^{INTRA} employs $f_d(\cdot)$ as defined in JPEG2000. The two remaining contexts models φ and ψ also employ the sign coding directives of JPEG2000 since, to the best of our knowledge, they achieve some of the best results reported in the literature. Although JPEG2000 defines a different context model for the coding of the refinement bits, its effectiveness is low [17], so it is not used herein. The INTER and 3D dependencies use functions similar to $f_x(\cdot)$ and $f_d(\cdot)$ but employing the coefficients defined in Table I.

The third context model, denoted by φ , is based on the so-called local average of wavelet coefficients [17]. It assumes that the magnitude of an index can be approximated as the arithmetic mean of its neighbors' magnitude, and that symbols emitted by bitplane coding engines are under-complete representations of the signal. In [17], a mathematical framework that characterizes the signal produced by wavelet transforms aids the development of this context model. Herein, the use of the complete quantized representation of the quantized coefficients simplifies its application. It can be implemented through an arithmetic mean as detailed in Table I.

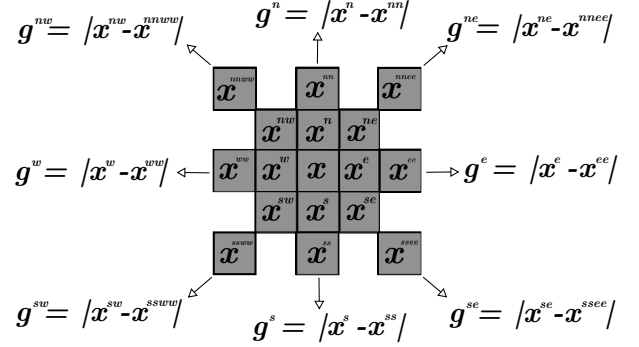
TABLE I: Formation of the context models evaluated. $R(\cdot)$ denotes the rounding operation.

		INTRA	INTER	3D
simple	$\sigma_x =$	x^w	x^p	x^{-1}
	$\sigma_d =$	d^w	d^p	d^{-1}
JPEG2000	$\phi_x =$	$f_x(S'(x^n, x^{ne}, x^e, x^{se}, x^s, x^{sw}, x^w, x^{nw}))$	$f'_x(S'(x^p, x^{pn}, x^{pe}, x^{ps}, x^{pw}))$	$f''_x(S'(x^{-1}, x^{-1n}, x^{-1e}, x^{-1s}, x^{-1w}, x^{+1}, x^{+1n}, x^{+1e}, x^{+1s}, x^{+1w}))$
	$\phi_d =$	$f_d(d^n, d^e, d^s, d^w)$	—	$f'_d(d^{-1}, d^{-1n}, d^{-1e}, d^{-1s}, d^{-1w}, d^{+1}, d^{+1n}, d^{+1e}, d^{+1s}, d^{+1w})$
local av.	$\varphi_x =$	$R((x^n + x^{ne} + x^e + x^{se} + x^s + x^{sw} + x^w + x^{nw})/8)$	$R((x^p + x^{pn} + x^{pe} + x^{ps} + x^{pw})/5)$	$R((x^{-1} + x^{+1})/2)$
	$\varphi_d =$	—	—	—
Gradient-based	$\psi_x =$	$\begin{cases} R((x^n + x^s)/2) & \text{if } x^n, x^s \geq T_x \text{ and } g^n, g^s < T_g \\ R((x^w + x^e)/2) & \text{if } x^w, x^e \geq T_x \text{ and } g^w, g^e < T_g \\ R((x^{nw} + x^{se})/2) & \text{if } x^{nw}, x^{se} \geq T_x \text{ and } g^{nw}, g^{se} < T_g \\ R((x^{sw} + x^{ne})/2) & \text{if } x^{sw}, x^{ne} \geq T_x \text{ and } g^{sw}, g^{ne} < T_g \\ R((x^n + x^e + x^s + x^w)/4) & \text{otherwise} \end{cases}$	—	$\begin{cases} \text{similar as in INTRA but} \\ \text{considering 5 directions that involve} \\ \text{coefficients from other components} \\ R((x^{-1} + x^{+1})/2) & \text{otherwise} \end{cases}$
	$\psi_d =$	—	—	—

The fourth context model, denoted by ψ , is conceived from the gradient adjusted predictor employed in CALIC [31], and the local average. ψ combines both approaches so that the gradient in the surroundings of the coefficient determines the neighbors that are employed to compute the local average. Such a model has never been proposed in the framework of wavelet-based image coding, so it is described with more detail. First, the gradients are computed as the absolute difference between two consecutive neighbors in the immediate surroundings of x . Fig. 2 depicts the coefficients involved to compute these gradients. Then, context ψ_x is computed as the local average of two adjacent neighbors that are on one of the four directions considered (vertical, horizontal, and two diagonals). The direction is chosen as that having the feeblest gradient. For intra-scale dependencies, for instance, the vertical direction is considered with a feeble gradient when indices x^n and x^s are both equal or greater than threshold T_x and gradients g^n and g^s are both lower than threshold T_g . As indicated in Table I, the other directions have equivalent considerations. When *none* direction has a feeble gradient, the context is computed as the local average of x^n, x^s, x^w, x^e . Thresholds T_x and T_g are selected to discard low intensity coefficients and strong gradients, respectively. They are empirically adjusted as $T_x = \max(\lfloor 2^{2+l+\alpha} \cdot \Delta \rfloor, 1)$, with l being the decomposition level (the first is 0), and

$$\alpha = \begin{cases} 2 & \text{if } v = \text{LL} \\ 1 & \text{if } v = \text{HL/LH} \\ 0 & \text{if } v = \text{HH} \end{cases} \quad (10)$$

$T_g = 4$ for all subbands. Due to the use of gradients that are adjacent to the coded coefficient, inter-scale dependencies can not be considered in this approach. ψ_x^{3D} considers the five basic directions that involve only coefficients from the immediately previous and following component. These 3D directions are not fully specified herein due to page constraints, though they can be found in the implementation provided below.

Fig. 2: Coefficients employed to compute eight gradients around x . Only intra-scale dependencies are illustrated.

We note that the four context models described above are devised for wavelet-transformed data. Other context models such as those employed in the latest video compression standard HEVC can not be directly applied herein since they are specifically devised for another type of transformed data [32].

IV. EXPERIMENTAL RESULTS

A. Natural images

The first set of experiments employ the eight natural images of the ISO 12640-1 corpus. These images are gray-scale, 8 bits per sample (bps), and have a size of 2560×2048 . Five levels of irreversible 9/7 wavelet decomposition are applied in all experiments except when indicated. The objective of the first test is to compare the entropy given by the proposed measure with the coding rate achieved by JPEG2000. Our JPEG2000 implementation BOI [33] is employed to carry out this test. JPEG2000 coding parameters are: lossy mode, codeblocks of 64×64 , and no precincts. Fig. 3 shows the results achieved when $H'(V)$ employs the context model borrowed from JPEG2000 with the use of intra-scale dependencies (i.e., ϕ^{INTRA}), which is mostly similar to the model of the standard. In this and following figures, the key provides the type of entropy reported ($H(V)$ or $H'(V)$), and the context model employed to compute it. The solid plots depicted in

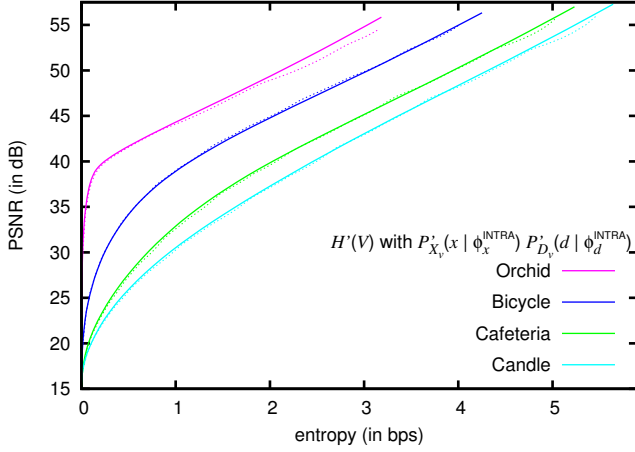


Fig. 3: Comparison between the entropy given by $H'(V)$ (solid plots) and the coding rate achieved with JPEG2000 (plots with dots), for four natural images of the ISO 12640-1 corpus.

Fig. 3 correspond to the entropy computed via $H'(V)$, whereas the plots with dots correspond to the results achieved with the actual JPEG2000 codec. The entropy is reported in bps, whereas the image quality is reported in PSNR. The results achieved for the four images depicted in the figure indicate that the proposed measure approximates with high accuracy the JPEG2000 coding rate, with only small variations that show a slightly larger rate for JPEG2000, in general. These variations may be caused due to the auxiliary information required in the JPEG2000 codestream. The differences between $H'(V)$ and JPEG2000 are negligible, validating the soundness of the proposed measure. These results hold for the other images of the corpus.

The second test evaluates the efficiency of the four context models when they exploit intra-scale dependencies. Fig. 4(a) depicts the results achieved for the “Bicycle” image. The figure also reports the entropy achieved when none context model is used (i.e., $H(V)$). The results suggest that the use of a context model significantly reduces the entropy. The context models ϕ (JPEG2000-like), φ (local-average approach), and ψ (gradient-based approach) achieve the lowest entropies for low image qualities. From medium to high PSNR values, the context models that achieve lowest entropies are φ and ψ . The JPEG2000-like context model achieves regular performance for these PSNRs when compared to the other models. Even the simplest context model σ (left-coefficient approach) achieves lower entropy than that of ϕ . At the highest image quality reported in the figure, the entropy of σ and φ is approximately 5% and 10% lower than that of ϕ , respectively.

The solid plots of Fig. 4(b) report the same results as those of Fig. 4(a). To ease the comparison, the horizontal axis in Fig. 4(b) is the PSNR, whereas the vertical axis is the difference, in bps, between the evaluated model and $H(V)$. The local average approach experiences a slight degradation at low qualities. Note that in the quality range from 25 to 35 dB, φ achieves entropies that are approximately 0.03 bps higher than those of ϕ and ψ . The context model based on the gradient approach does not experience this degradation,

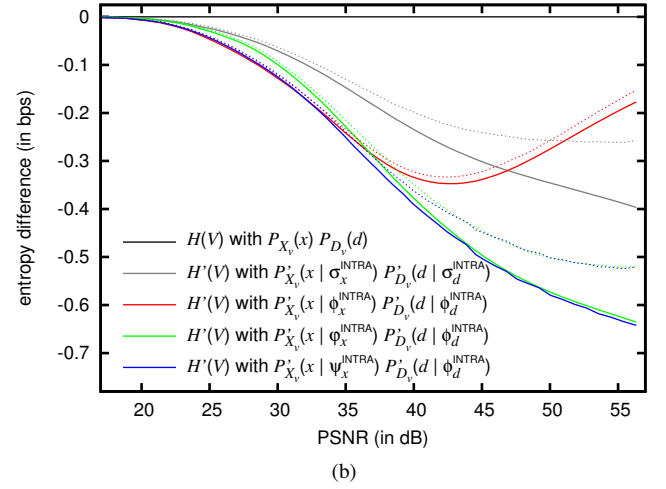
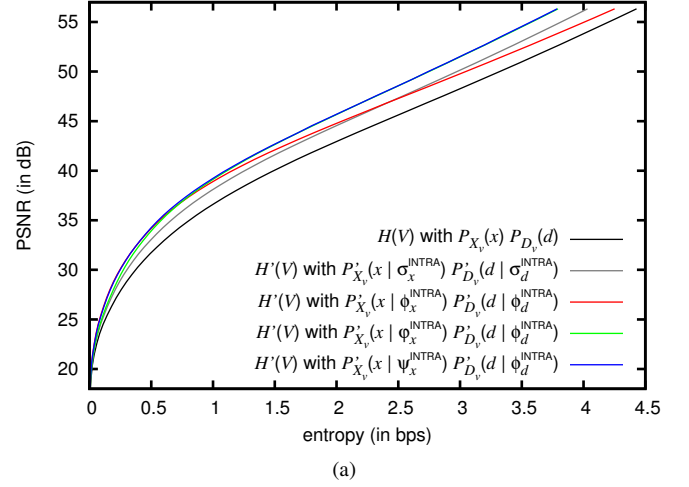


Fig. 4: Evaluation of the four context models, for the “Bicycle” image of the ISO 12640-1 corpus. (a) and (b) report the same results in different forms. The plots with dots of (b) indicate that the pmf are generated using data from all images of the corpus except “Bicycle”.

achieving the lowest entropies at all quality levels. Results hold for the other images of the corpus.

As stated before, the pmfs employed by $H'(V)$ are generated using the data of the image being evaluated. In practical systems, $P'_{X_v}(x | \Theta_x)$ is typically approximated with adaptive mechanisms [13]. However, $P'_{X_v}(x | \Theta_x)$ might also be approximated employing data from other images using a stationary probability model such that described in [17]. So, it is also of interest to appraise the entropy achieved when the pmfs are generated with a training set of images –that does not include the image evaluated. If $P''_{X_v}(x | \Theta_x)$ denotes the pmf generated with the training set, then such an approach replaces the first line of (7) by $-\sum_x P'_{X_v}(x | \Theta_x) \log_2 P''_{X_v}(x | \Theta_x)$. Equivalent changes are done in the second line of (7). The plots with dots depicted in Fig. 4(b) report the entropy achieved with this approach. Evidently, the entropy is higher than that achieved when using the conventional pmfs because $P''_{X_v}(x | \Theta_x)$ capture the statistical behavior of natural images

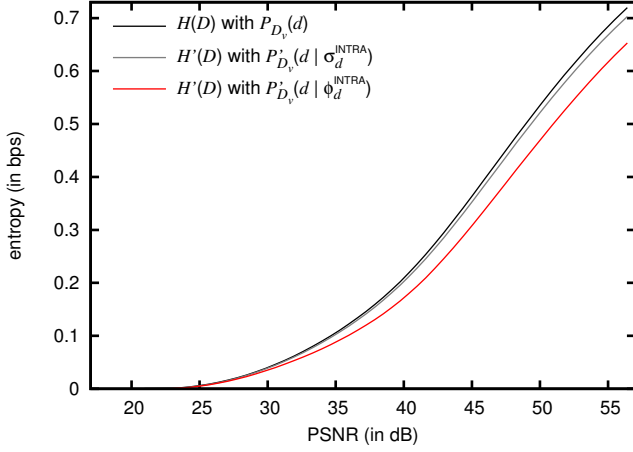


Fig. 5: Evaluation of the context models employed to code the sign, for the “Portrait” image of the ISO 12640-1 corpus. Only the entropy for the coefficients’ signs is reported.

in general rather than the specific behavior of the image evaluated. Nonetheless, it is worth noting that the increase in entropy produced due to the use of general pmfs does not modify significantly the behavior of the context models. In particular, the JPEG2000-like context model results almost unaffected by this variation, which suggest that the adaptive mechanisms employed in the standard might be removed without a significant penalization in coding performance. This insight has been used in [34] to enhance parallelism strategies in the JPEG2000 bitplane coding engine.

The next test evaluates the efficiency of context models σ_d^{INTRA} and ϕ_d^{INTRA} , employed for sign coding. Fig. 5 shows the entropy required to code the sign of the coefficients. For low quality levels, the entropy is low because most coefficients are quantized as 0. The higher the image quality, the higher the entropy. The context model that achieves the lowest entropy is that of JPEG2000 (i.e., ϕ_d^{INTRA}). The efficiency of σ_d^{INTRA} is similar to that achieved when none context model is employed, i.e., $H(D)$. At the highest quality reported in the figure, $H'(D)$ using ϕ_d^{INTRA} achieves an entropy 10% lower than that of $H(D)$.

The last test of this section evaluates the differences between the use of intra- and inter-scale dependencies. Fig. 6 reports the results achieved when using context models ϕ_x and φ_x . To enhance the comparison, the figure only reports the entropy of the indices, disregarding the entropy of the signs. For both models, the use of intra-scale dependencies achieves the lowest entropies, which corresponds with [2], [18]. The model that exploits the least the inter-scale dependencies is ϕ .

B. AVIRIS hyperspectral images

The next tests evaluate the entropy of three hyperspectral images that belong to the remote sensing field. The images are captured by the AVIRIS (Airbone Visible/Infrared Imaging Spectrometer) sensor. They are 16 bits per sample, have a size of 512×512 , contain 224 components, and are referred to as “cuprite”, “jasper”, and “lunarLake”. Two typical transform

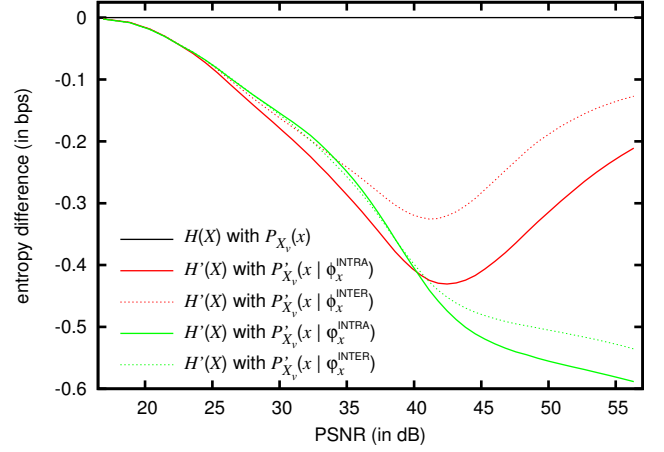


Fig. 6: Evaluation of the use of intra- and inter-scale dependencies with two context models, for the “Portrait” image of the ISO 12640-1 corpus. Only the entropy of the quantization indices is reported.

strategies are employed to decorrelate the redundancy of the image samples: “2D” and “1D+2D”. “2D” applies five levels of 2D wavelet decomposition in the image components. “1D+2D” first applies five levels of 1D wavelet decomposition along the depth axis of the image (i.e., on the spectrum), and then five levels of 2D wavelet decomposition on each component. The “1D+2D” transform strategy is called 3D hybrid transform and it is known to be among the most efficient transform strategies for this type of images [16].

The first test evaluates the four context models. Fig. 7 reports the entropy achieved by both the “2D” and “1D+2D” transform strategies, for the “cuprite” image. The same plot color reporting one context model is employed for both transform strategies because they can be clearly discerned in the figure. The context models utilized by the “2D” transform strategy exploit intra-scale dependencies, whereas the context models utilized by the “1D+2D” strategy exploit 3D dependencies. Regardless of the context model, the “1D+2D” strategy achieves lower entropies than those achieved by “2D”. This correspond with results in the literature [16]. All context models achieve similar results for low image qualities. At medium and high qualities, all context models achieve similar entropies except for ϕ (JPEG2000-like). At the highest quality level reported in the figure, ϕ achieves almost the same entropy as that achieved with $H(V)$, i.e., when none context model is employed. This seems to indicate that, for these images and transform strategies, context models based on the significance state are not appropriate.

Fig. 8 reports with solid plots the same evaluation as that of the previous figure but for the “jasper” image and only for the “2D” transform strategy. As before, this figure reports the entropy difference to ease the comparison. The context models ϕ and φ are reported for both intra- and inter-scale dependencies. Inter-scale dependencies are depicted with the plots with dots. Similarly to the results achieved for the natural images in Fig. 6, the use of ϕ_x^{INTER} does not reduce the entropy. Contrarily to the results achieved for the natural

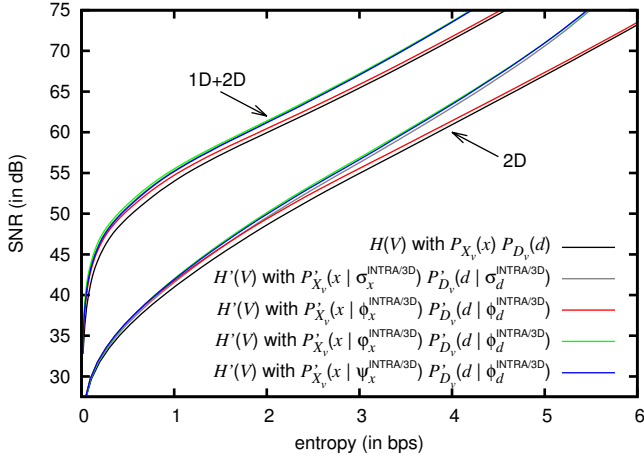


Fig. 7: Evaluation of the four context models when the “2D” and “1D+2D” transform strategies are applied on the AVIRIS image “cuprite”.

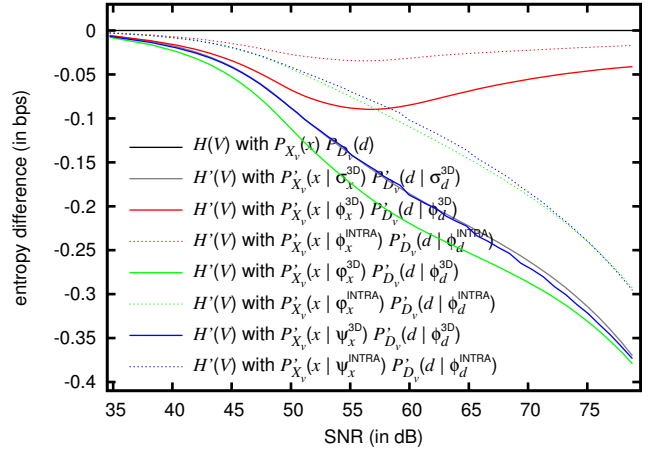


Fig. 9: Evaluation of intra-scale and 3D dependencies for the transform strategy “1D+2D”, for the AVIRIS “lunarLake” image.

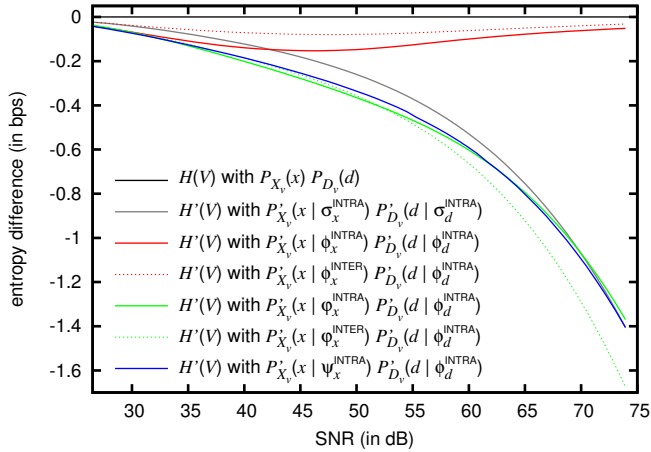


Fig. 8: Evaluation of intra- and inter-scale dependencies for the transform strategy “2D”, for the AVIRIS “jasper” image.

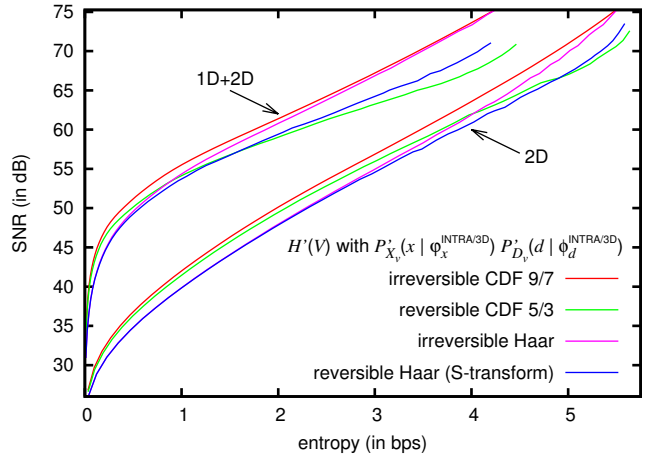


Fig. 10: Evaluation of four different wavelet transforms when the “2D” and “1D+2D” transform strategies are applied on the AVIRIS image “cuprite”.

images, the use of inter-scale dependencies with the context model φ_x^{INTER} reduces the entropy in approximately 20% at the highest quality level reported in the figure. This observation also holds for the other images.

Fig. 9 reports with solid plots the same evaluation as that of Fig. 7 but for the “lunarLake” image and only for the “1D+2D” strategy. Again, the context model that achieves lowest entropies is φ (local-average approach). The plots with dots depicted in the figure report the entropies achieved when the context models exploit intra-scale dependencies. As seen in the figure, the use of such dependencies significantly increases the entropy achieved.

The type of wavelet transform employed in a coding system has an important impact on the coding efficiency achieved. To illustrate this point, Fig. 10 reports the entropy obtained for the “cuprite” image when four different wavelet transforms are employed, namely, the irreversible 9/7, the reversible 5/3, the irreversible Haar, and the reversible Haar (or S-transform). In order not to clutter the figure, only the local average-based context model is depicted. The model employs intra-scale

dependencies for the “2D” strategy, and 3D dependencies for the “1D+2D” strategy. The lowest entropies are achieved by the irreversible 9/7, though at high quality levels the Haar transform achieves almost equivalent results, especially for the “1D+2D” transform strategy. This does not happen with the JPEG2000-like context model (not shown in the figure). Even though the reversible 5/3 wavelet achieves lower entropies than the S-transform at low quality levels, from medium to high qualities, the S-transform achieves lower entropies than the 5/3. This indicates that the transform employed to code an image is a choice that has to be taken considering the quality, or the rate, at which it will be transmitted.

The last test reveals a peculiar feature of these hyperspectral images. Fig. 11 reports the entropy obtained with the “2D” strategy when the context models employ 3D dependencies. All context models except ϕ enormously reduce the entropy of the image. To illustrate this point, the figure also depicts, with the plot with dots, the entropy achieved by the “1D+2D” strategy using the context model that achieves lowest entropy

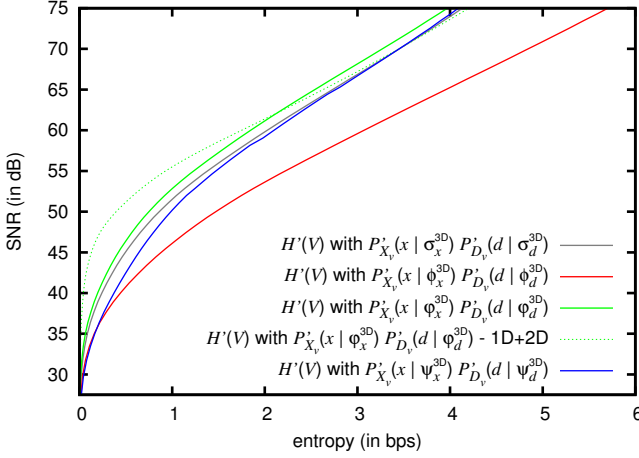


Fig. 11: Evaluation of 3D dependencies when employed with the “2D” transform strategy, for the AVIRIS “cuprite” image.

for that transform strategy (i.e., φ^{3D}). For low-to-medium quality levels, the “1D+2D” strategy results in the lowest entropies reported, however, from medium to high quality levels, the context models σ^{3D} , φ^{3D} , and ψ^{3D} employed with the “2D” strategy achieve the lowest entropies. These results indicate that these context models are able to decorrelate the entropy of the image as well as a 1D wavelet transform applied on the spectrum of the image. This insight is employed in [35] to reduce the computational costs of a coding scheme for hyperspectral images.

C. Computed Tomography images

The following tests evaluate the entropy of 3D images that belong to the medical field. The images are Computed Tomographies (CT) of a human chest. They are 12 bits per sample, have a size of 512×512 , contain 112 components, and are referred to as “A”, “B”, and “C”. The transform strategies “2D” and “1D+2D” are also applied for these images. Fig. 12 shows the entropy obtained by these two transform strategies. In order not to clutter the figure, only two context models are depicted. As before, the context models employed with the “1D+2D” strategy exploit 3D dependencies, whereas context models employed with the “2D” strategy exploit intra-scale dependencies. The differences between the “1D+2D” and “2D” strategies are not as significant as those found for the AVIRIS images. However, the context models show a similar behavior, suggesting that the best results are achieved by the gradient-based approach or the local average approach (not shown in the figure). At the highest quality level reported in the figure, the entropy difference between the gradient-based approach employing the “1D+2D” and “2D” strategies is almost negligible.

Figs. 13 and 14 report the same evaluations as those of Figs. 8 and 9 but for the CT images. The results achieved for the evaluation of the “2D” strategy in Fig. 13 are similar to those achieved for the AVIRIS images. The results achieved for the “1D+2D” evaluation in Fig. 14 are notably different from those obtained for the AVIRIS images. In this case,

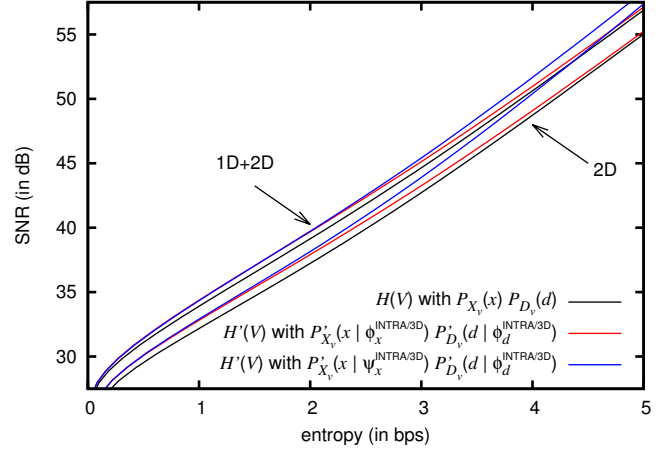


Fig. 12: Evaluation of the two context models when the “2D” and “1D+2D” transform strategies are applied on the CT image “A”.

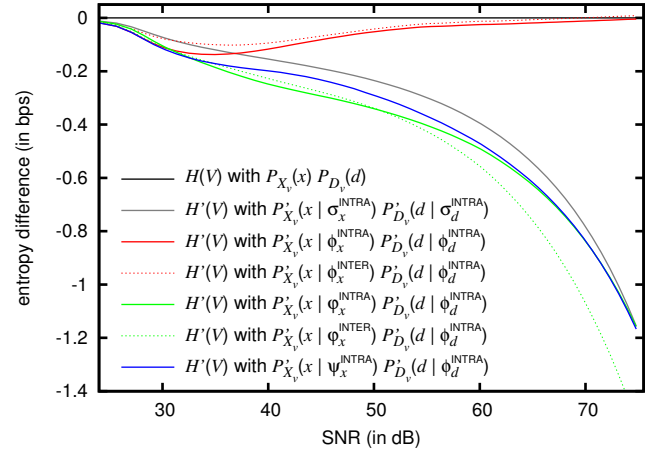


Fig. 13: Evaluation of intra- and inter-scale dependencies for the transform strategy “2D”, for the CT “B” image.

the use of intra-scale dependencies achieve equal or lower entropies than when using 3D dependencies. This suggest that the coding of different types of images should exploit different types of dependencies to achieve maximum coding efficiency.

V. DISCUSSION AND CONCLUSIONS

This paper introduces an entropy-based measure to evaluate different context models employed to code wavelet-transformed images. Four context models that can exploit three types of wavelet dependencies are employed with the proposed measure. The analysis obtained with this framework reveals interesting aspects of context models that may be employed to devise coding schemes that target particular types of images, transform strategies, or coding rates.

The main conclusions obtained for natural images suggest that: 1) context models based on the significance state are approximately 10% less efficient at high rates than the remaining models, 2) the context model that achieves lowest entropy is that that combines the gradient adjusted predictor with the

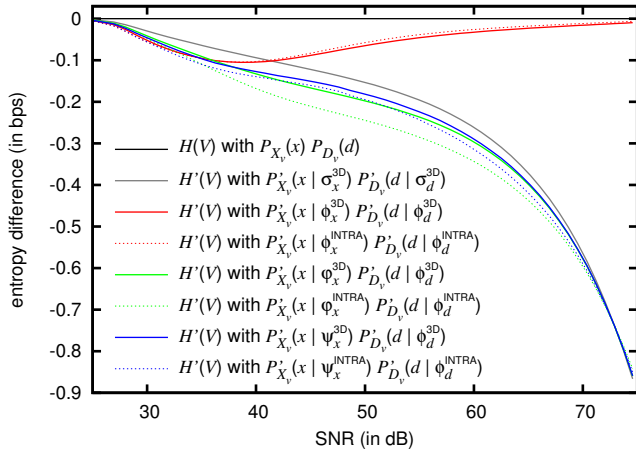


Fig. 14: Evaluation of intra-scale and 3D dependencies for the transform strategy “1D+2D”, for the CT “C” image.

local average approach, 3) the context model for signs that achieves lowest entropy is that defined in the JPEG2000 standard, 4) the context model of JPEG2000 may use stationary probabilities without a significant loss on coding performance, and 5) the use of intra-scale dependencies achieves the lowest entropy.

Results achieved for AVIRIS hyperspectral images suggest that: 1) the use of the significance state to define the context of a coefficient is not appropriate for these images, 2) the context model based on the local average can be effectively combined with the use of inter-scale dependencies to enhance compression efficiency, 3) 3D dependencies of wavelet data may significantly reduce the entropy when employing the “1D+2D” transform strategy, 4) the wavelet transform employed should be chosen together with the context model and the quality range at which the image is transmitted, and 5) 3D dependencies may be used with a “2D” transform strategy to enhance coding efficiency. For the CT images, similar conclusions as those drawn for the AVIRIS images are obtained except for the use of 3D dependencies with the “1D+2D” transform strategy, which does not achieve entropies as low as those achieved with intra-scale dependencies.

The experimental section of this paper reports only a small part of the experimental results generated to compile this summary. The proposed measure is especially useful to aid the design of new image codecs, and can be employed to test many other types of images, transforms, and context models. To ease its dissemination, the software tool developed in this work is left freely available in [36]. It is named ENTIMA and incorporates 10 different context models, 4 wavelet filterbanks, and several transform strategies for gray, color, and multi-component images.

REFERENCES

- [1] C. E. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.
- [2] 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.

- [3] 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.
- [4] S. Yea and W. A. Pearlman, “A wavelet-based two-stage near-lossless coder,” *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3488–3500, Nov. 2006.
- [5] 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.
- [6] M. Wang and M. van der Schaar, “Operational rate-distortion modeling for wavelet video coders,” *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3505–3517, Sep. 2006.
- [7] B. Foo, Y. Andreopoulos, and M. van der Schaar, “Analytical rate-distortion-complexity modeling of wavelet-based video coders,” *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 797–815, Feb. 2008.
- [8] M. D. Gaubatz and S. S. Hemami, “Robust rate-control for wavelet-based image coding via conditional probability models,” *IEEE Trans. Image Process.*, vol. 16, no. 3, pp. 649–663, Mar. 2007.
- [9] C.-Y. Tsai and H.-M. Hang, “One-sided p-GGD source modeling and rate-distortion optimization in scalable wavelet video coder,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 11, pp. 1557–1570, Nov. 2011.
- [10] V. Sheinin, A. Jagmohan, and D. ke He, “On the operational rate-distortion performance of uniform scalar quantization-based Wyner-Ziv coding of Laplace-Markov sources,” *IEEE Trans. Multimedia*, vol. 10, no. 7, pp. 1225–1236, Nov. 2008.
- [11] F. Auli-Llinas and M. W. Marcellin, “Scanning order strategies for bitplane image coding,” *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1920–1933, Apr. 2012.
- [12] F. Auli-Llinas, “General embedded quantization for wavelet-based lossy image coding,” *IEEE Trans. Signal Process.*, vol. 61, no. 6, pp. 1561–1574, Mar. 2013.
- [13] D. Taubman, “High performance scalable image compression with EBCOT,” *IEEE Trans. Image Process.*, vol. 9, no. 7, pp. 1158–1170, Jul. 2000.
- [14] P. Schelkens, A. Munteanu, J. Barbarien, M. Galca, X. Giro-Nieto, and J. Cornelis, “Wavelet coding of volumetric medical datasets,” *IEEE Trans. Med. Imag.*, vol. 22, no. 3, pp. 441–458, Mar. 2003.
- [15] W. A. Pearlman, A. Islam, N. Nagaraj, and A. Said, “Efficient, low-complexity image coding with a set-partitioning embedded block coder,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 11, pp. 1219–1235, Nov. 2004.
- [16] B. Penna, T. Tillo, E. Magli, and G. Olmo, “Transform coding techniques for lossy hyperspectral data compression,” *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1408–1421, May 2007.
- [17] F. Auli-Llinas, “Stationary probability model for bitplane image coding through local average of wavelet coefficients,” *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2153–2165, Aug. 2011.
- [18] 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.
- [19] 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.
- [20] M. D. Gaubatz and S. S. Hemami, “Efficient entropy estimation based on doubly stochastic models for quantized wavelet image data,” *IEEE Trans. Image Process.*, vol. 16, no. 4, pp. 967–981, Apr. 2007.
- [21] S. Mallat, “A theory of multiresolution signal decomposition: the wavelet representation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, pp. 674–693, Jul. 1989.
- [22] J. Rissanen, “Generalized Kraft inequality and arithmetic coding,” *IBM Journal of Research and Development*, vol. 20, no. 3, pp. 198–203, May 1976.
- [23] F. Auli-Llinas and J. Serra-Sagrasta, “JPEG2000 quality scalability without quality layers,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 7, pp. 923–936, Jul. 2008.
- [24] D. Taubman and A. Zakhor, “Multirate 3D subband coding of video,” *IEEE Trans. Image Process.*, vol. 3, no. 5, pp. 572–588, Sep. 1994.
- [25] X. Wu, “Context quantization with Fisher discriminant for adaptive embedded wavelet image coding,” in *Proc. IEEE Data Compression Conference*, Mar. 1999, pp. 102–111.
- [26] 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.
- [27] 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.

- [28] 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.
- [29] 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.
- [30] D. S. Taubman and M. W. Marcellin, *JPEG2000 Image compression fundamentals, standards and practice*. Norwell, Massachusetts 02061 USA: Kluwer Academic Publishers, 2002.
- [31] X. Wu and N. Memon, "CALIC - a context based adaptive lossless image codec," in *Proc. IEEE International Conference Acoustics, Speech, and Signal Processing*, vol. 4, Mar. 1996, pp. 1890–1893.
- [32] J. Sole, R. Joshi, N. Nguyen, T. Ji, M. Karczewicz, G. Clare, F. Henry, and A. Duenas, "Transform coefficient coding in HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1765–1777, Dec. 2012.
- [33] F. Auli-Llinas. (2014, Nov.) BOI codec. [Online]. Available: <http://www.deic.uab.cat/~francesc/software/boi>
- [34] F. Auli-Llinas and M. W. Marcellin, "Stationary probability model for microscopic parallelism in JPEG2000," *IEEE Trans. Multimedia*, vol. 16, no. 4, pp. 960–970, Jun. 2014.
- [35] F. Auli-Llinas, M. W. Marcellin, J. Serra-Sagrista, and J. Bartrina-Rapesta, "Lossy-to-lossless 3D image coding through prior coefficient lookup tables," *ELSEVIER Information Sciences*, vol. 239, no. 1, pp. 266–282, Aug. 2013.
- [36] F. Auli-Llinas. (2014, Nov.) ENTIMA - Entropy of a wavelet-transformed image. [Online]. Available: <http://www.deic.uab.cat/~francesc/software/entima>



Francesc Auli-Llinàs (S'06-M'08-SM'14) is a Ramón y Cajal Fellow with the Department of Information and Communications Engineering, Universitat Autònoma de Barcelona, where he received the B.Sc., B.E., M.Sc., and Ph.D. (cum laude) degrees in computer science in 2000, 2002, 2004, and 2006, respectively. Since 2002, he has been consecutively funded 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 D. Taubman at the University of New South Wales, and with the group of M. Marcellin at the University of Arizona. He develops and maintains BOI, a free-source JPEG2000 implementation. He was a recipient of two awards of bachelor given to the first students of the promotion in 2000 and 2002. In 2006, he was awarded with a free software mention from the Catalan Government for the development of BOI. In 2013, he was awarded with a distinguished R-Letter given by the IEEE Communications Society for a paper co-authored with M. Marcellin. He is reviewer for various magazines and symposiums and has authored numerous papers in the top journals and conferences of his field. His current research interests include a wide range of image coding topics, including highly scalable image and video coding systems, rate-distortion optimization techniques, massively parallel computing architectures, arithmetic coding, embedded quantization, and interactive transmission.