Entropy-based distortion measure and bit allocation for wavelet image compression

Thomas André*, Marc Antonini *Member, IEEE*, Michel Barlaud *Fellow, IEEE*, Robert M. Gray *Fellow, IEEE*

***Abstract*—Quality criteria for image coding are often based on mean square error. However, this is not always a relevant measure of visual quality at low bit-rate. We investigate here the properties of a distortion measure based on the conditional differential entropy of the input signal given its quantized value. The proposed measure appears to be a correct representation of the amount of information lost by quantization. An adaptive bit allocation algorithm is proposed in order to take advantage of this criterion. Experimental results illustrate the behavior of the proposed distortion measure and exhibit interesting visual properties for low bit-rate subband image coding.**

***Index Terms*—Entropy, distortion, image coding.**

EDICS Category: COD-OTHR

I. Introduction

Mean square error (MSE) has been widely used as a distortion measure for image and video coding. In most of standard image and video coders, optimal quantizers are determined in the sense of MSE, and bit allocation algorithms determine the best tradeoff between bit-rate and distortion in the sense of MSE. A major advantage of MSE is that analytical formula can be easily established for many problems under the assumption of Gaussian or generalized Gaussian probability density functions (PDF) [1], and under the hypothesis of high-resolution. In addition, MSE has good regularity properties and is measured with little computational effort.

It is well known, however, that these MSE estimators loose their efficiency when the data do not follow a normal law of probability or when the high-resolution hypothesis does not hold. A normal or Laplacian law of probability is a strong assumption on the prediction error of wavelet or DCT coefficients, or on predicted frames in video coding, especially when dealing with high compression rates. Moreover, many research works have pointed out that, although MSE is a relevant visual quality measure at high bit-rate, it is not always the case [2]. This fact has motivated a great deal of research on perceptual coding based on weighted quadratic distortion measures. These weighted quadratic measures are also interesting because they led to important results on rate-distortion theory and vector quantization [3]–[5].

Several approaches led to better visual results [6], [7] by taking into account human psycho-visual characteristics [8] or even the effects of the wavelet transform [9]. In this work, we focus instead on the distortion measure itself, in the framework of a simple wavelet-based image coder.

Many distortion measures have been proposed and analyzed as alternatives to MSE [3], [10]. Some are well-adapted to particular problems related to human vision, others are suitable for speech processing. However, under certain asumptions such as a Gaussian distribution of the source, these measures are equivalent to weighted MSE. So are many of the other distortion measures based on information, such as the Mahalanobis distance, the Kullback-Leibler divergence [11], the Itakura-Saito measure [12], and the quantizer mismatch [13], [14]. While entropy has been used to compare the efficiency of different quantizers, differential entropy has not been used as a distortion measure.

In this paper, we propose a new distortion measure based on differential entropy, which is not equivalent to any weighted distortion measure based on MSE. This estimator is both well-adapted to the PDF of the quantization error, and is robust to outliers. We here restrict interest to scalar quantization, in which case it is known that for asymptotically large rate uniform quantizers yield minimum entropy for fixed MSE distortion [1], [15]. We focus on the intrinsic properties of the proposed distortion measure, and we present a bit allocation algorithm based on it in order to illustrate its properties.

The paper is organized as follows. In section II we introduce a differential-entropy-based distortion measure. We show that this criterion can be interpreted as a measure of the information carried in the signal, and we study its properties. In section III, we propose a bit allocation algorithm based on the proposed distortion measure, and in section IV, we present some experimental results.

II. Differential entropy-based distortion measure

A. Notation and background

Let X be a random variable which describes the behavior of a signal x to be quantized. Let us denote by p_X its PDF. A quantizer $Q(x)$ is defined by a partition $\mathcal{S} = \{S_i = [t_i, t_{i+1})\}$, $t_{i+1} > t_i$, and a reproduction codebook $\{\hat{x}_i\}$ as $Q(x) = \hat{x}_i$ if $x \in S_i$, where i takes values in an index set which we take to be the nonnegative integers. We assume without loss of generality that the $\hat{x}_i$ are distinct. We denote by $p_i = \Pr(Q(X) = \hat{x}_i) = \Pr(X \in S_i)$ the probability mass function (PMF) of indexed quantizer output points (or of the indexes themselves).

The classical squared error distortion measure is defined by

$$d\left(x, Q(x)\right) = \left(x - Q(x)\right)^2 \tag{1}$$

and the average distortion (i.e., the MSE) is

$$\begin{aligned} D_{\textbf{MSE}} &= E[d(X,Q(X)] = \int_{\mathbf{R}} p_X(x)\left(x - Q(x)\right)^2\, dx \\ &= \sum_i \int_{t_i}^{t_{i+1}} p_X(x)(x-\hat{x}_i)^2\, dx. \end{aligned} \tag{2}$$

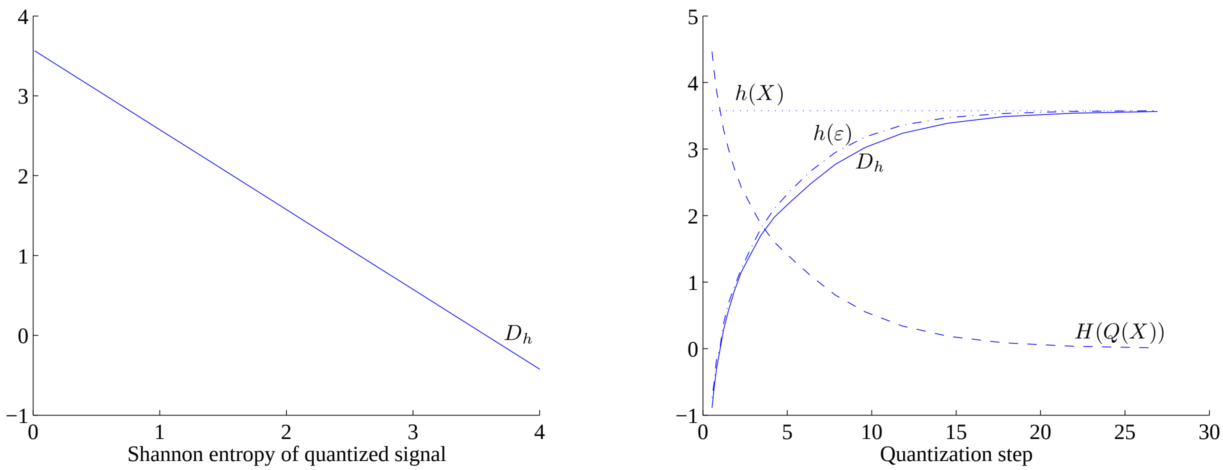
The MSE tends to favor high-energy coefficients, which is relevant at high bit-rate but does not always correspond to a better visual quality in the general case. In the following, we propose an alternative to this criterion.

B. The proposed distortion measure

The quantization of a signal x should provide a compact representation of this signal with as little information loss as possible. In other words, it should minimize the amount of information contained in the

T. André, M. Antonini and M. Barlaud are with I3S Laboratory, Université de Nice-Sophia Antipolis, UMR 6070 of CNRS, - Bât. Algorithmes/Euclide B, 2000 route des Lucioles - BP121 - 06903 Sophia-Antipolis Cedex, France - Tel. +33 4.92.94.27.49 - Fax. +33 4.92.94.28.98 - e-mail: {andret,am,barlaud}@i3s.unice.fr

Robert M. Gray is with Stanford University, Information System Lab, Department of Electrical Engineering, 261 Packard Building, 350 Serra Mall, Stanford, CA 94305, USA - e-mail: rmgray@stanford.edu



(a) (b)

Fig. 1. (a) Proposed distorsion criterion as a function of the discrete Shannon entropy of the quantized subband (in bpp) — (b) Entropic distortion D_h as a function of the quantization step, compared with the discrete Shannon entropy $H(Q(X))$ of the quantized signal (in bpp), the differential entropy $h(X)$ of the original subband and the differential entropy $h(\varepsilon)$ of the corresponding quantization error ε (in bpp) (b). Image "Aerial" from the JPEG2000 database, high-frequency (HF) wavelet subband obtained using the $9/7$ filters [16].

quantization error $\varepsilon(x) = x - Q(x)$ [1]. Note that in the general case, and in particular for subband image and video coding, $\varepsilon(x)$ is real-valued. Toward this end we introduce a distortion measure defined as the self-information contained in the quantization error given the quantizer output, that is,

$$\begin{aligned} d(x, Q(x)) &= -\log p_{\varepsilon(X)|Q(X)}(\varepsilon(x)|Q(x)) \\ &= -\log p_{\varepsilon(X)|Q(X)}(x - Q(x)|Q(x)) \\ &= -\log p_{X|Q(X)}(x|Q(x)), \end{aligned} \tag{3}$$

the self-information in the input conditioned on the quantizer output. In this latter form the distortion measure can be identified as the distortion measure proposed in Theorem 3 of Najmi's Shannon rate-distortion theoretic approach to model selection and statistical inference [17]. The average distortion is easily computed as

$$D_h = E[d(X, Q(X)] \tag{4}$$

$$\begin{aligned} &= -\sum_i \int_{t_i}^{t^{i+1}} p_X(x) \log p_{X|X\in S_i}(x)\, dx \\ &= -\sum_i p_i \int_{t_i}^{t^{i+1}} p_{X|X\in S_i}(x) \log p_{X|X\in S_i}(x)\, dx \\ &= h(X|Q(X)), \end{aligned} \tag{5}$$

the conditional differential entropy of the input given its quantized value. This has the intuitive interpretation that small distortion corresponds to small conditional differential entropy given the quantized signal.

The quantity $h(X) + n$ can be interpreted as the average number of bits (assuming base 2 logarithms) required to describe X to n bits accuracy (see [18] p. 229). In the case where X describes pixel values, $h(X)$ is expressed in bits per pixel (bpp). Note that, unlike the discrete Shannon entropy H, the differential entropy h of a random variable can be negative or infinite [18], [19]. In the remainder of the paper, D_h will be referred to as "entropic distortion."

[1]Note that the distortion measure is not in general a difference distortion measure, that is, it does not depend only on the error $x - Q(x)$.

C. Properties

We note that the proposed distortion d as well as its average D_h are neither symmetric nor non-negative, and thus cannot be considered as distances. Indeed, considering the Kullback divergence between the probability of the source and the conditional probability of the source knowing the quantizer, defined as

$$D_{KL}(p_X, p_{X|Q(X)}) = \int_{\mathbf{R}} p_X(x) \log \frac{p_X(x)}{p_{X|Q(X)}(x|Q(x))} dx, \tag{6}$$

we obtain the following relationship by developing the integral:

$$D_h = D_{KL}(p_X, p_{X|Q(X)}) + h(X). \tag{7}$$

This relationship illustrates again the fact that D_h is homogeneous to a divergence, i.e. a distortion measure, rather than a distance in the usual sense. D_h does not satisfy the usual properties of distances, and in particular, can be negative. However, this fact does not prevent one from using it as a measure of goodness [17].

An informative alternative form for the average distortion follows from results for mutual information of mixed continuous and discrete random variables (see, e.g., [20]). In particular, the mutual information between the continuous random variable X and the discrete random variable $Q(X)$ can be expressed in two forms as

$$I(X, Q(X)) = H(Q(X)) - H(Q(X)|X) \tag{8}$$

$$= h(X) - h(X|Q(X)), \tag{9}$$

where H denotes the discrete Shannon entropy. Since $Q(X)$ is a deterministic function of X, $H(Q(X)|X) = 0$ and hence the entropic distortion can be expressed as

$$D_h(Q) = h(X|Q(X)) = h(X) - H(Q(X)), \tag{10}$$

the difference between the continuous average self information of the original signal and the Shannon average self information in the quantized reproduction.

D. Comparison with MSE

The proposed distortion measure depends only on the quantizer partition $\mathcal{S}$ and *not* explicitly on the reproduction codebook. Thus the reproduction values can be picked in any reasonable way. For example, the $\{\hat{x}_i\}$ can be chosen as the usual centroids with respect to MSE. This would be the provably optimal choice if, for example, one used a Lagrangian distortion of the form $D_h + \lambda$MSE for even a small Lagrangian multiplier $\lambda > 0$. Typically in addition to a distortion measure, one also has a rate-constraint. For a fixed-rate code, the rate is measured as the log of the number of codewords, say $\ln N$. In this case the maximum entropy codebook has entropy $\ln N$ and is achieved by a partition with equal probability cells. Forcing the quantizer to be regular (or the cells to be intervals in the scalar case) and using MSE centroids for reproductions constrains the maximum entropy quantizer to be well behaved. In the case of variable-rate quantization with an entropy constraint $H(Q(X)) \leq R$, the optimal quantizer will have discrete Shannon entropy of R. It is interesting to note that in the asymptotic case of large codebooks, the optimum MSE variable-rate scalar quantizer maximizes entropy [21]. Thus for asymptotically high rate, both distortion measures lead to the uniform quantizer in the variable rate case.

Fig. 1a shows the behavior of the distortion criterion for uniform scalar quantizers on two subbands of a wavelet coder. Note that as the quantization step q grows larger, the entropy $H(Q(X))$ of the quantized signal tends to zero, and D_h converges toward the amount of information $h(X)$ contained in the original signal, as shown in Fig. 1b: $\lim_{q\to\infty} D_h = h(X)$. By comparison, the classical distortion D_{MSE} converges towards the power of the original signal.

Note also that D_h differs from the differential entropy $h(\varepsilon)$ of the quantization error except for $q \to 0$ and $q \to \infty$, as shown in Fig. 1b. The linear relationship (10) is not valid if $h(\varepsilon)$ is used to measure average distortion instead of D_h.

E. Computation in practice

According to equation (5), the proposed distortion criterion is the differential entropy of a real-valued signal. Differential entropy can be estimated using several methods. The most simple consists of computing a discrete histogram of the signal, with estimation bins $\{P^i\}$ centered on values $\{x_i\}$ being in a reasonable number; from this histogram, the discrete probability of each estimation bin can be computed, and the proposed distortion D_h is expressed as

$$D_h \approx -\sum_{j\in\cup P^i} \mathrm{Pr}_X^{\Delta}\{x_j\}\log_2 \mathrm{Pr}_X^{\Delta}\{x_j|\hat{x}_j\}, + \log_2 \Delta. \tag{11}$$

where Δ is the width of the estimation bins P^i and $\mathrm{Pr}_X^{\Delta}\{x_i\}$ is the probability for the outcome values of the signal X to be located in the bin P^i [18]. This method gives satisfactory results if the signal is described by a sufficient number of samples, and it has been adopted for experimental results of this paper. In our experiments, we adapted Δ so that 512 bins were used to estimate histograms. This value was set experimentally to obtain good results with large images. Other estimation methods exist and can be used as well. For example, we also tested an estimator using the k-th nearest neighbors [22] with no noticeable difference on large images.

So far, we have presented the proposed distortion criterion from a theoretical point of view. In order to illustrate its relevance for image coding, we present in the following a bit allocation algorithm for wavelet-based image coder based on this criterion.

III. BIT ALLOCATION ALGORITHM

In a wavelet-based image coder, the bit allocation part consists of distributing the available bit budget between the subbands so that a certain distortion criterion is optimized. Usually, MSE is employed to this purpose. In this section, we make use of the proposed distortion criterion instead.

A. Proposed bit allocation method

According to (10), minimizing the linear criterion D_h is equivalent to maximizing the entropy $H(Q(X))$ of the quantized signal. In the framework of a wavelet-based image coder, such a minimization would thus lead to the trivial solution of maximum entropy. Classical optimal bit allocation algorithm rely on the fact that the MSE-distortion criterion is convex [23], [24], which is not the case here. They also assume a linear relationship between the global distortion and the distortion of each wavelet subband: this assumption is true for MSE distortion [25], but no relationship has been established between the entropy of a reconstructed image and the entropies of the corresponding wavelet subbands. For this reason, another approach is needed, adapted to the proposed distortion measure.

We propose the following allocation algorithm. According to the previous discussion, D_h represents the amount of information contained in the quantization error. Thus, it makes sense to distribute the total amount of error among the subbands proportionally to the amount of information they contain. In this case, the allocation procedure consists of quantizing each subband i so that its measured distortion D_i satisfies

$$D_i = \alpha a_i h_i, \tag{12}$$

where a_i accounts for the relative size of the subband, h_i is its differential entropy, and α is a parameter determined by the total target bit-rate. Note that a_i must be included so that all distortions are expressed in bits per pixel of the original image. For a N-levels dyadic wavelet decomposition, we have $a_i = 2^{-2n_i}$ where $n_i \in \{1, .., N\}$ is the decomposition level number of the subband i:

$$n_i = \min\left(N, \left[\frac{i}{3}\right] + 1\right). \tag{13}$$

The value of α must be determined according to the target bit-rate chosen by the user. The greater α, the higher the compression ratio. In the case where $\alpha \geq 1/a_i$, the subband i must be fully discarded. Thus, the target distortion D_i^* for the subband i will be:

$$D_i^* = \begin{cases} \alpha a_i h_i & \text{if } \alpha < \frac{1}{a_i}, \\ h_i & \text{otherwise.} \end{cases} \tag{14}$$

The special case $\alpha = 0$ corresponds to an “almost-lossless” coding of the signal.

Any iterative algorithm can be used to determine α. For example, the optimal quantization steps q_i can be iteratively defined for each subband i until the distortion D_i^* is reached with an acceptable precision.

However, this method suffers from two drawbacks, linked to the estimation of the differential entropy. First, if this estimation is not accurate, neither is the bit allocation. The problem can occur with relatively small images, especially if many wavelet decomposition levels are used: the lowest subbands do not contain enough pixels for the entropy estimation to be reliable. And second, the differential entropy estimation is a time consuming process, and the bit allocation algorithm described above requires it to be run several times, both on original subbands and on their quantization errors. In the following, we propose a modified version of the bit allocation algorithm which compensates for these two drawbacks.

B. Fast algorithm

Let us consider equations (10) and (14) again. According to them, the target entropic bit-rate $H(Q_i(X_i))^*$ for subband i must satisfy:

$$H(Q_i(X_i))^* = \begin{cases} (1 - \alpha a_i)h(X_i) & \text{if } \alpha < \frac{1}{a_i}, \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

Let us denote by L the number of subbands discarded by the proposed bit allocation algorithm, in other words the first L subbands. According to equation (15), all subbands j of relative size a_j (i.e. which belong to the same resolution level n_j) are simultaneously discarded if $1-\alpha a_j \leq 0$. Thus, for a dyadic spatial wavelet transform, we have:

$$L = 3. \left\lfloor \frac{\alpha}{4} \right\rfloor . \quad (16)$$

Then, we have $\forall i \leq L, H(Q_i(X_i))^* = 0$. By definition, the target bit-rate R_T satisfies:

$$R_T = \sum_i a_i H(Q_i(X_i))^* = \sum_{i>L} a_i H(Q_i(X_i))^*, \quad (17)$$

and it follows that

$$R_T = \sum_{i>L} a_i(1-\alpha a_i)h(X_i). \quad (18)$$

Finally, we obtain

$$\alpha = \frac{S_1 - R_T}{S_2}, \quad (19)$$

where $S_1 = \sum_{i>L} a_i h(X_i)$ and $S_2 = \sum_{i>L} a_i^2 h(X_i)$. The bit allocation algorithm can thus be modified as follows:

Parameters : target bit-rate R_T, precision ε

1) For each subband i:
   - Estimate $h(X_i)$
2) Compute L, S_1, S_2 and α using (19)
3) For each subband i: Compute $H(Q_i(X_i))^*$ using (15)
4) Quantization for each subband i :
   - Guess an initial quantization step q_i
   - Quantize the subband i using q_i
   - Evaluate the current value of $H(Q_i(X_i))$
5) Compute the total entropic bit-rate $H = \sum_i H(Q_i(X_i))$
6) If $|H - R_T| > \varepsilon$:
   - Update q_i
   - Resume to 4.

The actual distortion criterion is in fact never computed, nor are the differential entropies of the quantization errors. Only the differential entropy of each subband must be computed and only once, which reduces the probability of bad entropy estimation, thus of bad bit allocation. Moreover, this algorithm is comparable to the classical algorithms based on MSE in terms of complexity.

IV. Experimental results and discussion

In order to compare the proposed bit allocation algorithm with a classical optimal bit allocation algorithm [25] based on MSE distortion, we implemented both of them into a simple image coder which consists of a 5-levels 2D wavelet transform using the $9/7$ filters [16], a uniform quantizer, and an arithmetic coder. However, the proposed algorithm could be used in any image coder such as JPEG2000. Moreover, provided that EBCOT is used as entropy coder, the produced bitstream can be decoded using any JPEG2000-compliant decoder.

A. Effects of entropy-based bit allocation

Figure 2 compares the PSNR of the reconstructed image “Cafe” from the JPEG2000 database, using both bit allocation algorithms for different bit-rates. Figure 3 shows parts of the reconstructed images corresponding to the points marked in Fig. 2. Table I also shows the repartition of bit-rate between the different decomposition levels, for several bit allocation algorithms: MSE-optimal bit allocation with no psycho-visual weighting (MOBA), post-compression rate distortion algorithm used in [26] with fixed cisual weighting (PCRD+FVW), dynamic contrast-based quantization (DCQ) from [7], and the proposed entropy-based bit allocation with no psycho-visual weighting (EBA). Note that parameters of PCRD+FVW and DCQ have been optimized for a specific conditions of observation. Several things are then highlighted.

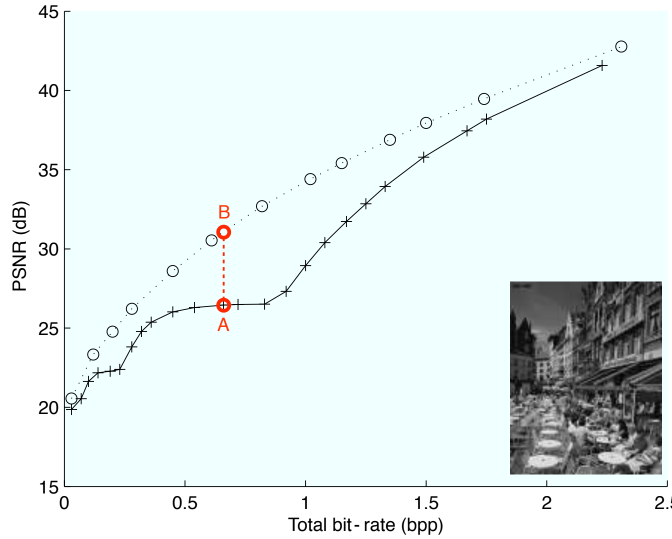


Fig. 2. PSNR (dB) of decoded image “Cafe” as a fuction of the target bit-rate (bpp), using a classical optimal bit allocation algorithm (dotted curve) and the proposed bit allocation algorithm (solid curve). Parts of images A and B are shown in fig. 3.

Algorithm	Decomposition level				
	1	2	3	4	5
MOBA	22.2%	47.3%	19.0%	6.1%	5.4%
PCRD + FVW *	0.5%	60.6%	26.8%	8.6%	3.5%
DCQ *	0.0%	50.4%	35.2%	11.0%	3.4%
EBA	**0.0%**	**39.4%**	**44.5%**	**10.7%**	**5.4%**

TABLE I
Bit-rate repartition between the different wavelet decomposition levels for different bit allocation algorithms. Image Woman compressed at 0.25 bpp. Level 1 corresponds to the finest scale. Results marked by * are extracted from [7].

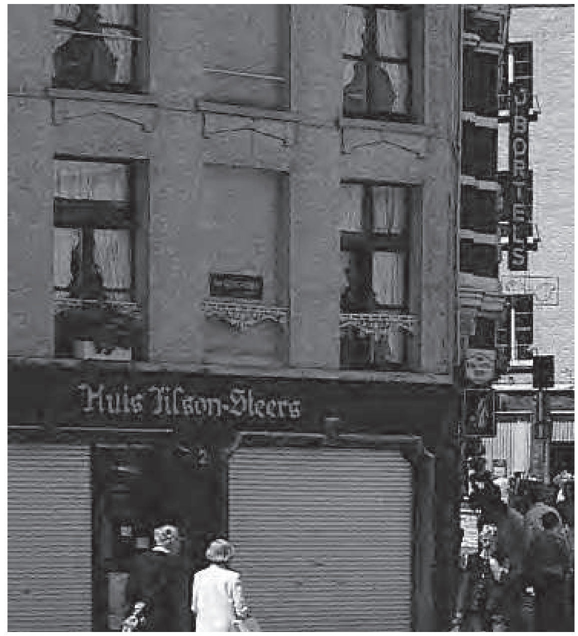
First, the PSNR is always better when the classical bit allocation algorithm is used, which was expected since MSE bit allocation maximizes the PSNR. However, the difference is small at some bit-rates, and the perceptual quality of images reconstructed using the proposed method is not accordingly worse (see the next subsection).

Second, the PSNR curve of the proposed method exhibits a few inflexion points. They correspond to the bit-rates at which an entire level of wavelet decomposition is discarded. In other words, the proposed bit allocation algorithm automatically performs a sort of spatial scalability (see also Table I). To some extent, it explains the observed visual results: for example, under a critical bit-rate, the reconstructed image is actually made of the low-pass subband of the first decomposition level. A spatial subsampling sometimes leads to better visual results than a compression of all subbands at very-low bit-rates: the proposed algorithm offers an automatic support for this.

B. Visual impressions

Figure 3 illustrates the effects of the proposed algorithm on the image “Cafe”. The most visible effects are a reduction of ringing and

(a) Proposed bit allocation



(b) MSE-optimal bit allocation

Fig. 3. Comparison of the two bit allocation methods on parts of image "Cafe". Total bit-rate is 0.65 bpp.

Image (bpp)	Algorithm	PSNR (dB)	DCTune2.0 (error)	SSIM (%)	NQM (dB)
Woman (0.25)	MOBA	**30.02**	37.20	0.990	39.21
	DCQ	28.15	**33.82**	**0.996**	**42.5**
	EBA	27.72	38.95	0.995	42.03
Cafe (0.65)	MOBA	**28.00**	**39.76**	0.998	40.05
	DCQ	24.79	44.23	**0.999**	39.77
	EBA	23.48	48.35	**0.999**	**41.36**

TABLE II
QUALITY OF DECODED IMAGES "WOMAN" AND "CAFE" EVALUATED USING DIFFERENT METRICS, FOR DIFFERENT BIT ALLOCATION ALGORITHMS.

geometric artifacts and a much better preservation of textures, which lead to a better visual impression atlow bit-rate. The reconstructed image looks like a low-resolution image rather than a highly-compressed image. More decoded images obtained using different algorithms are available online at http://www.i3s.unice.fr/~creative/entropy.

Interestingly, the results are very close to those obtained by algorithms taking the human psycho-visual system into account (see also table I), even though our approach is very different. The reconstructed images appear a bit blurred, and do not necessarily compare favorably with those obtained with the classical method when the original image is available for visual comparison.

We also compared (Table II) the quality of several decoded images using different visual quality assessment metrics, namely DCTune [27], noise quality measure (NQM, [28]) and structural similarity index (SSIM, [2]), on the images "Woman" and "Cafe". These results confirm that the proposed approach leads to results visually very close to those obtained using the DCQ algorithm. Note that DCTune was originally developed to assess the quality of DCT-based image compression algorithms.

C. Discussion

One can argue that similar behavior and visual impression can be obtained by replacing the differential entropies h_i by a constant (e.g. 1) in equation (12). This approach, which we will note EBA_1, consists of allocating the available bit budget in the subbands according to their relative size a_i only, disregarding their differential entropy. Another variant, motivated by the intuition that the differential entropy h_i of each subband i is related to its variance σ_i^2, could consist of replacing h_i by an approximation using σ_i^2. For example, in the case of a zero-mean Gaussian signal X of variance σ^2, it can be shown that $h(X) = \frac{1}{2}\log_2 2\pi e\sigma^2$ (see e.g. [18] p.225). This last approach will be denoted by EBA_{σ^2}. In the following, we compare these two methods with the original EBA algorithm.

At first glance, images obtained using EBA_1 and EBA_{σ^2} are indeed quite similar to those obtained using EBA: one can observe the same blurring effect, and the bit-rate repartition between decomposition levels are also similar for certain images and compression rates, as reported in Table III. However, whereas EBA_1 distributes equally the bit-rate between subbands of a given decomposition level, EBA takes each one's entropy into account, thus leading to different results: as reported in Table III, EBA is better than EBA_1 in the sense of SSIM and NQM. As shown in Table IV, different subbands from a same decomposition level may indeed have very different entropies. Even though both algorithms produce blurry images due to the dependance on the size of the subbands, distributing the proposed distortion among the subbands according to their differential entropy improves the perceptual quality of reconstructed images.

Moreover, Table IV indicates that the Gaussian approximation of h_i is not valid for most subbands. This explains the results reported in Table III, according to which EBA is more efficient than EBA_{σ^2}.

V. CONCLUSION

In this paper, we have introduced a distortion measure for image coding, based on differential entropy. We have shown that this measure is a correct representation of the amount of information lost by quantization, and has some interesting properties. We have proposed a bit allocation algorithm based on this distortion measure which gives interesting results for image coding at low bit-rates. An extension to color image coding can be envisaged, for example based on kNN multivariate density estimator [29]. However, the coding algorithm used here for experiments still uses uniform scalar

Image (bpp)	Algorithm	Bit-rate repartition between levels 1	2	3	4	5	PSNR (dB)	DCTune2.0 (error)	SSIM (%)	NQM (dB)
"Woman" (0.25)	EBA	0.0%	39.4%	44.5%	10.7%	5.4%	27.72	38.95	0.995	42.03
	EBA_1	0.0%	44.9%	40.6%	9.7%	4.8%	28.10	36.12	0.994	39.48
	EBA_{σ^2}	0.0%	41.9%	42.9%	10.3%	4.9%	27.86	38.33	0.994	39.55
"Cafe" (0.65)	EBA	0.0%	63.4%	26.5%	7.2%	2.9%	23.48	48.35	0.999	41.36
	EBA_1	0.0%	63.1%	26.7%	7.3%	2.9%	23.48	48.32	0.998	38.68
	EBA_{σ^2}	0.0%	62.8%	27.0%	7.3%	2.9%	23.47	48.51	0.998	38.71

TABLE III

BIT-RATE REPARTITION BETWEEN SUBBANDS AND QUALITY OF DECODED IMAGES EVALUATED USING DIFFERENT METRICS, FOR THE ALGORITHMS EBA, EBA_1 AND EBA_{σ^2}

	Subband i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
"Woman"	h_i	2.6	3.9	3.5	3.6	4.2	3.9	3.5	3.6	3.4	2.6	3.0	2.8	2.5	3.4	3.0	7.4
	$\frac{1}{2}\log_2 2\pi e\sigma_i^2$	3.1	4.7	4.2	4.4	5.2	4.8	4.4	4.6	4.4	3.5	4.2	3.8	3.5	4.5	4.2	7.9
"Cafe"	h_i	3.5	4.6	4.8	4.3	5.0	5.1	4.3	5.2	5.1	4.5	5.4	5.2	4.7	5.4	5.5	7.6
	$\frac{1}{2}\log_2 2\pi e\sigma_i^2$	4.0	5.4	5.5	5.0	5.8	5.9	4.9	5.8	5.8	4.9	5.9	5.7	5.0	5.9	5.9	7.8

TABLE IV

DIFFERENTIAL ENTROPY h_i AND ITS VARIANCE-BASED APPROXIMATION (BITS PER SAMPLE) FOR EACH SUBBAND i. IMAGES "WOMAN" AND "CAFE", 5 DECOMPOSITION LEVELS.

quantization, which might not be optimal for the proposed criterion. Building an optimal quantizer in the sense of this distortion criterion is not trivial. This problem will be approached in future works.

Acknowledgements

This work was supported in part by the University of Nice-Sophia Antipolis (France), by the National Science Fundation under NSF Grant CCR-0309701 and by a gift from Hewlett-Packard. The authors would like to thank Dr. Damon Chandler and Prof. Sheila Hemami for providing source code and experimental results.

REFERENCES

[1] H. Girsh and J. Pierce, "Asymptotically efficient quantizing," *IEEE Transactions on Information Theory*, vol. IT-14, pp. 676–683, Sept. 1963.

[2] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. on Image Proc.*, vol. 13, no. 4, pp. 600–612, avril 2004.

[3] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design," *IEEE Transactions Commun.*, vol. 28, no. 1, pp. 84–95, January 1980.

[4] T. Linder, G. Lugosi, and K. Zeger, "Recent trends in lossy source coding," *Journal on Communications (Hungary)*, vol. XLV, pp. 16–22, Mar 1994.

[5] T. Linder, R. Zamir, and K. Zeger, "On source coding with side-information-dependent distortion measures," *IEEE Transactions on Information Theory*, vol. 46, pp. 2697–2704, Nov 2000.

[6] A. Watson, "The cortex transform: Rapid computation of simulated neural images," *Comput. Vis. Graph. Image Processing*, vol. 39, pp. 311–327, 1987 1987.

[7] D. Chandler and S. Hemami, "Dynamic contrast-based quantization for lossy wavelet image comresssion," *IEEE Transactions on Image Processing*, vol. 14, no. 4, pp. 397–410, Apr. 2005.

[8] N. Jayant, J. Johnston, and R. Safranek, "Signal compression based on models of human perception," *Proc. of the IEEE*, vol. 81, no. 10, Oct 1993.

[9] A. Watson, G. Tangand, J. Solomon, and J. Villasenor, "Visibility of wavelet quantization noise," *IEEE Transactions on Image Processing*, vol. 6, no. 8, pp. 1164–1175, Aug 1997.

[10] M. Gutman, "On uniform quantization with various distortion measures," *IEEE Transactions on Information Theory*, vol. IT-33, pp. 169–171, Jan. 1987.

[11] S. Kullback and R. Leibler, "On information and sufficiency," *Annals of Mathematical Statistics*, vol. 22, pp. 79–86, 1951.

[12] F. Itakura and S. Saito, "Analysis synthesis telephony based upon maximum likelihood method," in *Repts. 6th Int. Congr. Acoustics*, Y. Kohasi, Ed., 1968.

[13] R. Gray and T. Linder, "Mismatch in high rate entropy constrained vector quantization," *IEEE Transactions on Information Theory*, vol. 49, pp. 1204–1217, May 2003.

[14] A. Aiyer, K.-S. Pyun, Y.-Z. Huang, D. O'Brien, and R. Gray, "Lloyd clustering of gauss mixture models for image compression and classification," *Signal Processing : Image Communication*, vol. 20, pp. 459–485, june 2005.

[15] A. Gersho and R. Gray, *Vector Quantization and Signal Compression*. Norwell, MA: Kluwer Academic Publisher, 1992.

[16] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transforms," *IEEE Transactions on Image Processing*, vol. 1, no. 2, pp. 205–220, Apr. 1992.

[17] A. Najmi, "Data compression, model selection and statistical inference," Ph.D. dissertation, Stanford University, 1999.

[18] T. Cover and J. Thomas, *Elements of Information Theory*. N. Y.: John Wiley & Sons, Inc., 1991.

[19] R. Gray, *Source Coding Theory*. Norwell, MA: Kluwer Academic Publisher, 1991.

[20] ——, *Entropy and Information Theory*. New York, USA: Springer-Verlag, 1990.

[21] D. Messerschmitt, "Quantizing for maximum entropy," *IEEE Transactions on Information Theory*, September 1971.

[22] M. Goria, N. Leonenko, V. Mergel, and P. N. Inverardi, "A new class of random vector entropy estimators and its applications in testing statistical hypotheses," *Journal of Non-parametric Statistics*, vol. 17, no. 3, pp. 277–297, 2005.

[23] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 9, pp. 1445–1453, 1988.

[24] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit allocation for dependent quantization with applications to multiresolution and mpeg video coders," *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 533–545, Sept 1994.

[25] B. Usevitch, "Optimal bit allocation for biorthogonal wavelet coding," in *Proc. of Data Compression Conference*, Mar. 1996, pp. 387–395.

[26] D. Taubman, "High performance scalable image compression with EBCOT," *IEEE Transactions on Image Processing*, vol. 9, no. 7, pp. 1158–1170, July 2000.

[27] A. Watson, "DCTune: A technique for visual optimization of dct quantization matrices for individual images," *Society for Information Display Digest of Technical Papers*, vol. XXIV, pp. 946–949, 1993.

[28] N. Damera-Venkata, T. Kite, W. Geisler, B. Evans, and A. Bovik, "Image quality assessment based on a degradation model," *IEEE Trans. on Image Proc.*, vol. 9, no. 4, pp. 636–650, avril 2000.

[29] S. Boltz, E. Debreuve, and M. Barlaud, "High-dimensional statistical distance for region-of-interest tracking: Application to combining a soft geometric constraint with radiometry," in *IEEE International Conference on Computer Vision and Pattern Recognition*, Minneapolis, USA, 2007.