A Rational Methodology For Lossy Compression

J.A. Garcia,* Rosa Rodriguez-Sánchez,

J. Fdez-Valdivia, and, Xose R. Fdez-Vidal

Abstract

In the coding community distortion measures like peak signal to noise ratio (PSNR), root mean square error (RMSE), and less often mean absolute error (MAE), are used to precisely measure and quantify the error present in a compressed image and great effort is expended toward minimizing such an error. However, these measures appear not capable to predict visual distinctness from digital imagery as perceived by human observers. Though the drawbacks with RMSE, PSNR, and MAE are becoming better known, they are still used to evaluate encoding schemes as well as in rate-distortion optimization.

To circumvent the lack of knowledge of what distortion measure is more suitable for optimization of the trade-off between image fidelity and coding rate, this paper shall introduce a novel mathematical methodology for rate control by organizing the progressive transmission in accordance with coherence constraints for avoiding forms of behavioral inconsistency. A set of postulates is provided for specifying the ways in which preferences need to be made precise and fit together if illogical forms of behavior are to be avoided. We shall show that the rational choice for transmission at truncation time t is to select bit streams which have the maximum increase in utility per coding bit, where "rational" must be understood in the sense that it cannot lead the transmission system into forms of behavioral inconsistency. This method is then used within a progressive transmission scheme to produce a new compression method called rational embedded wavelet image coding (REWIC).

Coding results should be presented which compare the new coding scheme against the state of art coder in progressive transmission and demonstrate its superior performance based on an objective criterion of evaluation. In this paper, it would be concluded if, in the language of statistical hypothesis testing, the rational coding is significantly better than the state of art in progressive compression.

J.A. Garcia, Rosa Rodriguez-Sánchez and J. Fdez-Valdivia are with the *Departamento de Ciencias de la Computación e I.A., E.T.S. de Ingeniería Informática, Universidad de Granada, 18071 Granada, Spain.* **E-mail:** jags@decsai.ugr.es. **Fax:** +34.958.243317. **Phone:** +34.958.244235.

Xose R. Fdez-Vidal is with the *Departamento de Fisica Aplicada, Facultad de Fisica, Universidad de Santiago de Compostela, 15706 Santiago de Compostela, Spain*

Index terms: Visual distinctness, distortion measures, rational embedded coding, rate control, decision problem, information theoretic measures, coder selection.

I. INTRODUCTION

Many examples can be enumerated regarding the key role of distortion measures in both theoretical and practical aspects of transmission systems. For example, in rate distortion theory the basic problem is: given a source distribution and a distortion measure, what is the minimum expected distortion achievable at a particular rate? or, what is the minimum rate description required to achieve a particular distortion? Even though an iterative algorithm for the calculation of the rate distortion function for a general i.i.d. source and arbitrary distortion measure was suggested by several authors [1], [2], [3], in image coding there is at present no real alternative to using the mean squared error as the distortion measure.

A second example is a progressive transmission scheme that prioritizes the code bits according to their reduction in distortion (e.g., [4]). If the original image is I, the coding is actually done to $C = \Omega(I)$, where Ω represents a unitary hierarchical subband transformation [5]. The 2D array $C = \{c_{i,j}\}$ has the same dimensions of I, and each element $c_{i,j}$ is called transform coefficient at coordinate (i, j) which for the purpose of coding it can be treated as an integer. In a progressive transmission scheme, the decoder initially sets the reconstruction vector $\hat{C}$ to zero and updates its components according to the coded message. After receiving the value (approximate or exact) of some coefficients, the decoder can obtain a reconstructed image $\hat{I} = \Omega^{-1}(\hat{C})$. In such a scheme, a major objective is to select the most important information–which yields the largest distortion reduction–to be transmitted first, where the distortion measure is the mean square-error [4].

A final example. The comparison of different compression schemes is illustrated in terms of rate, distortion, and execution time results which are reported on the operation of each coding algorithm on test images and the corresponding decoding algorithm on the resultant compressed files. The figures on rate are then calculated from compressed file sizes and on mean squared error or peak signal to noise ratio (PSNR) from the reconstructed images given by the decoding algorithm. For example, in a progressive transmission scheme the decoder reads the first bytes of the file up to the desired rate, calculates the inverse subband transformation, and then compares the recovered image with the original through the distortion which is measured by PSNR between the original and reconstructed images.

As a result, great pains are taken to precisely measure and quantify the error present in a compressed image, and great effort is expended toward minimizing an error measure that most often is of dubious value. These measures include PSNR, root mean square error (RMSE), and (less often) mean absolute error (MAE). The reason behind this is the fact that the perceived distortion in visual content is a difficult quantity to measure, as the characteristics of the human visual system are complex and not well understood [6], [7]. In order to illustrate this point we may perform a simple experiment to investigate the comparative results of various distortion measures for predicting image fidelity as follows.

Ten digitized mammographic images with clustered microcalcifications were selected from the Nijmegen database[8] and the MIAS database[9], and consultant radiologists review the original images to detect microcalcifications. The method for characterizing compression losses using images reconstructed under various degrees of lossy compression–the coding scheme was the international standard JPEG [10]—was the false positive to true positive (FP/TP) ratio for microcalcification detection. In order to derive the FP/TP detection ratio, the images reconstructed after compression at different levels were also reviewed to detect individual microcalcifications. Since it is rated diagnostic usefulness rather than general appearance or simply line or edge patterns, this study relates diagnostic accuracy to compression level. The performance of any distortion measure for predicting image fidelity in terms of diagnostic usefulness was then evaluated on the basis of the correlation coefficient $corr(y, z)$ between the FP/TP ratio and the distortion measure (which was applied to quantify the image distortion by means of the difference between the original mammographic image and images reconstructed after lossy compression). The comparative results of $RMSE$, $PSNR$, and MAE, and those of a simple measure of information gain–the Kullback-Leibler gain [11]– are illustrated in Table I. At the bottom of each of the columns is given the mean

$Image$	$corr\ (KL, \frac{FP}{TP})$	$corr\ (RMSE, \frac{FP}{TP})$	$corr\ (MAE, \frac{FP}{TP})$	$corr\ (PSNR, \frac{FP}{TP})$
# 1	0.843	0.731	0.740	0.630
# 2	0.951	0.922	0.929	0.854
# 3	0.971	0.896	0.908	0.821
# 4	0.886	0.826	0.850	0.736
# 5	0.789	0.835	0.839	0.737
# 6	0.818	0.750	0.774	0.652
# 7	0.824	0.775	0.792	0.700
# 8	0.848	0.723	0.737	0.623
# 9	0.944	0.829	0.843	0.727
# 10	0.823	0.792	0.808	0.701
$MEAN$	0.870	0.808	0.822	0.718

TABLE I

Comparative results of the Kullback-Leibler information gain and those of MAE, $RMSE$, and $PSNR$.

$\overline{corr}(y, z)$ of the correlation coefficients in that column. The Kullback-Leibler (KL) gain yields the highest value of the mean ($\overline{corr}(y, z) = 0.87$). Therefore, the KL gain shows the best overall performance.

But, how accurate are these data summaries? To analyze this question we may use bootstrap sampling for making statistical inference, [12]. Once we have the bootstrap sampling distribution of $\overline{corr}(y, z)$, the important point is to produce good confidence intervals for $\overline{corr}(y, z)$ automatically. The most advanced bootstrap intervals are the BC_a intervals, [12]. Table II shows the 95% BC_a confidence intervals for $\overline{corr}(y, z)$. From these results, statistical theory allows us to make the following inference: The true value of the mean $\overline{corr}(y, z)$ for the KL information gain lies in the interval $0.823 < \overline{corr}(y, z) < 0.944$ with 95% confidence. It says that if we ran a much bigger experiment, the value of $\overline{corr}(y, z)$ for the KL information gain probably wouldn't be too much different than the value of 0.87 that was estimated on the original set of ten mammographic images. From these confidence intervals,

BCa INTERVALS WITH 95 % CONFIDENCE	
Statistic	*Interval*
$\overline{corr}\left(KL, \frac{FP}{TP}\right)$	$(0.823, 0.944)$
$\overline{corr}\left(MAE, \frac{FP}{TP}\right)$	$(0.808, 0.843)$
$\overline{corr}\left(RMSE, \frac{FP}{TP}\right)$	$(0.792, 0.835)$
$\overline{corr}\left(PSNR, \frac{FP}{TP}\right)$	$(0.701, 0.737)$

TABLE II

95% BC_a CONFIDENCE INTERVALS FOR $\overline{corr}(y, z)$

BCa INTERVALS WITH 95 % CONFIDENCE	
Statistic	*Interval*
$\overline{corr}\left(KL, \frac{FP}{TP}\right) - \overline{corr}\left(MAE, \frac{FP}{TP}\right)$	$(0.017, 0.074)$
$\overline{corr}\left(KL, \frac{FP}{TP}\right) - \overline{corr}\left(RMSE, \frac{FP}{TP}\right)$	$(0.027, 0.087)$
$\overline{corr}\left(KL, \frac{FP}{TP}\right) - \overline{corr}\left(PSNR, \frac{FP}{TP}\right)$	$(0.118, 0.184)$

TABLE III

95% BC_a CONFIDENCE INTERVALS FOR THE DIFFERENCE OF MEANS

we also claim that the hypothesis $\overline{corr}(y, z) < 0.8$ is only rejected at 95% for the KL gain and MAE. Table III shows the 95% BC_a confidence intervals for the difference of means. Since they do not include the neutral value 0, in the language of statistical hypothesis testing the KL gain was found to be significantly better for predicting image distortion (in the sense of diagnostic usefulness) than MAE, $RMSE$, and $PSNR$.

Though the drawbacks of RMSE, PSNR, and MAE are already acknowledged within the coding community, PSNR is used to evaluate encoding schemes, yet, as well as RMSE is still employed to select the important information in embedded wavelet coding or in rate-distortion optimization. This is an interesting area for additional research and may lead to improvements in the visual quality of the coded images.

To circumvent the lack of knowledge of what distortion measure is more suitable for rate control (optimization of the trade-off between image fidelity and coding rate requires an ability to measure distortion), here we show a novel mathematical methodology for rate control by organizing the progressive transmission in accordance with coherence constraints for avoiding forms of behavioral inconsistency (Section II). To this aim, we shall describe the basic elements of this problem. Next, a set of postulates is provided for specifying the ways in which preferences need to be made precise and fit together if illogical forms of behavior are to be avoided. Then we show that the rational choice for transmission at truncation time t is to select bit streams which have the maximum increase in utility per coding bit, where "rational" must be understood in the sense that it cannot lead the transmission system into forms of behavioral inconsistency. This method will be used within a progressive transmission scheme to produce a new compression method called rational embedded wavelet image coding (REWIC).

Coding results should be presented which compare the new coding scheme against a state-of-the-art technique for lossy compression and demonstrate its superior performance based on an objective criterion of evaluation. The best instrument to measure image quality is the human eye. Unfortunately, visual tests are expensive to perform: Since human visual performance for targets or other relevant objects inherently shows a large variance and depends strongly on prior knowledge of the perceived scene, a typical experiment therefore requires a large number of observers to obtain statistically reliable data. Hence, other techniques to measure image quality are required.

As demonstrated in [6], the compound gain $\mathcal{E}_C^{Z_1,\cdots,Z_n}$ is a measure of information gain between two images such that satisfies three postulates which were found to be relevant for predicting visual distinctness, in addition to the basic axiomatic characterization of any proper measure of information gain. The main result in [6], [7] is the finding that the compound gain appears to relate to visual distinctness as perceived by human observers. This result implies that it might be used to predict visual distinctness from digital imagery. This is an important finding, since it may eliminate the need for psychophysical experiments, which are time consuming, and sometimes even impossible to perform.

In this paper, to assess the comparative compression performance of two progressive compression schemes (SPIHT by Said and Pearlman [4], and REWIC), various test images have been compressed, each with a huge number of coding rates. For each test image, the compound gain $\mathcal{E}_C^{Z_1,\cdots,Z_n}$ was then applied to quantify the image distortion by means of the difference between the original image and images reconstructed under various degrees of lossy compression.

For each test image, figures on rate-distortion are then calculated from compressed file sizes and on compound gain $\mathcal{E}_C^{Z_1,\cdots,Z_n}$ between the original image and reconstructed images given by each decoding algorithm (in this case, SPIHT and REWIC). The overall difference between two of these rate-distortion functions can be measured by subtracting the area beneath their graphs. This operation corresponds to a Kolmogorov-Smirnov (K-S) test. Based on the K-S test results, it will be then concluded if, in the language of statistical hypothesis testing, the new compression method is significantly better for embedded wavelet compression than the state-of-the-art SPIHT (Section IV).

II. RATIONAL EMBEDDED WAVELET IMAGE CODER (REWIC)

In an embedded wavelet scheme the coding is done to $C = \Omega(I)$, where Ω represents a unitary hierarchical subband transformation and I is the original image. A tree structure, called spatial orientation tree (SOT) in [4], naturally defines the spatial relationship on the resultant hierarchical pyramid. Figure 1 illustrates how a SOT is defined in a pyramid constructed with recursive four-subband splitting. Each node of the tree corresponds to a pixel, and its direct descendants (offspring) correspond to the pixels of the same spatial orientation in the next finer level of the pyramid. Transform coefficients in a spatial orientation tree correspond to a particular local spatial region of the original image, and thus, each SOT is associated with one spatial region as illustrated in Figure 2.

The embedded coding is distinctive from the conventional coding in the sense that any spatial orientation tree is coded bit-plane by bit-plane through successive-approximation quantization. Given an initial threshold T_0, the successive approximation consists in an iteration over two scans, called the sorting pass and the refinement pass. The bit stream

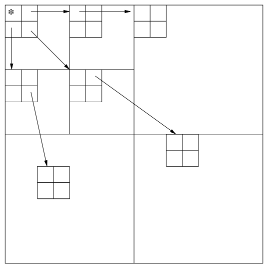


Fig. 1. A spatial-orientation tree defined in a pyramid constructed with recursive four-subband splitting.

of the sorting pass, BS_i, is generated using the set-partitioning approach, [4], to locate the significant wavelet coefficients with respect to the threshold. The bit stream of the refinement pass, called BS_{i+1}, results from the refinement of all the coefficients of the SOT which have not been quantized to zero so far. After a refinement pass, a new threshold may be computed as $T_i = T_{i-1}/2$ where i is the iteration number. The algorithm may be iterated by applying successive sorting and refinement bit streams, BS_i and BS_{i+1}. The result of the coding of the particular SOT is a completely embedded bit stream as given by a number of sorting and refinement bit streams $BS_1\, BS_2 \cdots BS_i\, BS_{i+1} \cdots BS_K$.

In this section we show that the performance of the embedded coder can be optimized by means of a rational strategy for rate control by organizing the progressive transmission with expected increase in utility per coding bit. We denote such coding strategy as rational embedded wavelet image coding (REWIC). REWIC allocates the available coding bits first to the optimal number of bit streams, for any spatial-orientation tree, with the maximum achievable expected increase in utility per coding bit. The resultant coding bitstream can be truncated at any point and still maintain an optimal utility-per-coding-bit performance at every truncation point.

The overall objective of developing a rational approach to choosing among SOTs–presented

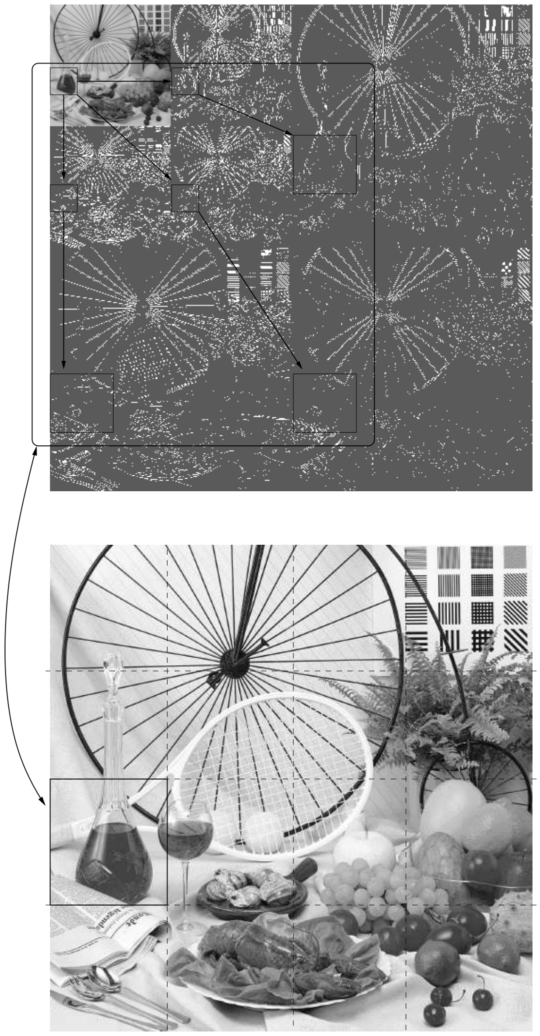


Fig. 2. Each spatial-orientation tree in the hierarchical pyramid corresponds to a spatial region of the original image.

in the form of a series of axioms as follows–do not purport to describe the ways in which individuals actually do behave in making choices among SOTs for their progressive transmission, neither do they assert the ways in which should behave. The axioms simply are proposed to prescribe constraints which it seems to us imperative to acknowledge in the image transmission problem where a system should choose, at any truncation time, among alternative spatial-orientation trees, *in such a way as to avoid certain forms of behavioral inconsistency.*

In the following it will be shown that any progressive transmission system which aspires to analyze the SOT selection problem at any truncation time t in accordance with a rational approach should verify that numerical values attached to the consequences–foreseen if a particular SOT is taken–are represented in the form of a utility function (that is, spatial-orientation trees cannot be ordered without a specification of utilities–numerical values–for the consequences). Next, we characterize the form of the expected increase in utility provided by transmitting a number of bit streams $BS_j\, BS_{j+1} \cdots BS_{j+n}$ for a particular SOT, if preferences are described by proper utility functions.

At any truncation time, the rational strategy for rate control with the embedded coder is to deliver the bit streams $BS_j\, BS_{j+1} \cdots BS_{j+n}$ which provides the maximum achievable expected increase in utility per coding bit for any available spatial-orientation tree.

A. BASIC ELEMENTS OF THE SELECTION PROBLEM

We shall describe any situation in which choices are to be made among available spatial-orientations trees for their transmission by successive-approximation quantification (s.a.q.), in terms of a decision problem whose structure is determined by three basic elements:

1. A set $\{R_i, i \in I\}$ of available spatial-orientation trees, one of which is to be selected for transmission of a number of bitstreams $S(R_i, t) = BS_j\, BS_{j+1} \cdots BS_{j+n}$, at truncation time t;

2. For each SOT R_i, a set $\{G_{l,i}; l \in L\}$ describing the gray-level occurrences in the spatial region which is reconstructed using the bit streams $S(R_i, t)$ candidates to be transmitted at time t, in addition to bitstreams transmitted, before the time t, for R_i; where $G_{l,i}$ denotes the uncertain event pixel X within spatial region associated with R_i, taking gray-level value l;

3. Corresponding to each set $\{G_{l,i}; l \in L\}$, a set of consequences $\{c_{l,i}; l \in L\}$ that induces the transmission of a number of bit streams $S(R_i, t) = BS_j\, BS_{j+1} \cdots BS_{j+n}$ for the particular SOT R_i.

The idea is as follows. Suppose we choose spatial orientation tree R_i for transmission of a number of its bitstreams $BS_j\, BS_{j+1} \cdots BS_{j+n}$, at truncation time t; then one set $\{G_{l,i}; l \in L\}$

occurs and leads to the corresponding consequence set $\{c_{l,i}; l \in L\}$. Thus, the choice of a R_i that is required at any truncation time t produces an outcome (the corresponding set of graylevel occurrences) that is beyond our control and induces a particular set of consequences. The entire transmission of all bit planes for each SOT involve sequential considerations but, this may reduce, essentially, to repeated analyses based on the above structure.

It is clear that to choose a particular SOT R_i is to opt for the uncertain scenario labelled by the pair set $\{(G_{l,i}, c_{l,i}); l \in L\}$. Obviously, the perception of the state of uncertainty resulting from the choice of any particular SOT for further transmission, is very much dependent on the information currently transmitted. Further information being transmitted will change the graylevel occurrences that result of selecting for transmission R_i. Thus the above representation captures the structure of a decision problem as perceived at a particular truncation time.

Preferences about (partial) transmission of different SOTs depend on assessments of the graylevel occurrences attached to the corresponding events. And this is clearly subject to change as new information is transmitted.

In addition to representing the structure of the selection of a SOT for transmission by s.a.q. (on the information transmitted at the time) as a decision problem using the three elements discussed above, we must also represent the idea of preference as applied to the comparison of all of the pairs of available SOTs. We shall therefore need to consider a fourth basic element in the SOT selection problem which is reformulated in terms of a decision problem: The relation $\preceq$ which expresses the preferences between pairs of available SOTs at a particular truncation time, so that $R_i \preceq R_j$ signifies that R_i is not preferred, at the truncation time, to R_j for further transmission of a number of bitstreams.

It should be understood that the preferences among SOTs are dependent on the overall state of information transmitted at that time (the actual information transmitted for each SOT).

We now give a formal definition of the selection of a SOT for transmission at a truncation time as a decision problem:

Definition 1. SOT selection as a decision problem. *The decision problem of SOT selection at the truncation time t is defined by four elements $\{\mathcal{R}, \mathcal{G}, \mathcal{C}, \preceq\}$ where:*

(i) $\mathcal{R}$ is the set of available SOTs;

(ii) $\mathcal{G}$ is the class of any possible set of gray-level occurrences corresponding to the transmission for a particular SOT;

(iii) $\mathcal{C}$ is the class of any set of possible consequences associated with a gray level occurrence set by means of the selection of a SOT;

(iv) $\preceq$ is a preference order between spatial-orientation trees; taking the form of a binary relation between SOTs in $\mathcal{R}$.

B. BASIC AXIOMATIC FOR AVOIDING FORMS OF BEHAVIORAL INCONSISTENCY

The operational notion of preference between SOTs, formalized by the binary relation $\preceq$, provides a qualitative basis for comparing SOTs. The following coherence axioms (Axioms 1 to 3) are to be proposed to provide a minimal set of rules to ensure that qualitative comparisons based on $\preceq$ cannot have intuitively undesirable implications (following [13]).

The first postulate states the essence of what is required for an orderly and systematic approach to comparing among spatial-orientations trees: (a) if all consequences were equivalent, there would not be a decision problem; and (b) if the system aspires to make a rational choice between alternative spatial-orientation trees, then it must at least be willing to express preferences between different SOTs.

Postulate 1.

(i) Not all the consequences in $\mathcal{C}$ are equivalent; and

(ii) The transmission system is able to compare any pair of options concerning the selection of a SOT at truncation time t.

The second axiom is intended to impose rules of coherence on preference orderings that will exclude possibility of two types of inconsistencies: First, a SOT is strictly preferred to itself; second, willingness to suffer the certain loss of something of value, which the system found itself expressing the preferences $R_i \preceq R_j$, $R_j \preceq R_k$, and $R_k \preceq R_i$ among the three SOTs R_i, R_j, and R_k.

Postulate 2.

(i) $R_i \preceq R_i$*; and*

(ii) If $R_i \preceq R_j$ *and* $R_j \preceq R_k$*, then* $R_i \preceq R_k$*.*

The binary relation $\preceq$ may also provide a qualitative basis for comparing, by extension, consequences and gray level occurrence events. And the third axiom shall ensure the consistency of any kind of preferences (e.g., between consequences or graylevel occurrence events).

Postulate 3.

(i) Preferences between visual fidelity of decoded outcomes (i.e., consequences) should not be affected by the transmission of further information;

(ii) If a gray level occurrence set $\{G_{l,i}; l \in L\}$ *is more likely to relate to better decoded outcomes than another gray level occurrence set* $\{G_{l,j}; l \in L\}$ *then the selection of the SOT that produces the former set is preferred to that producing the latter;*

(iii) If we have that R_i *is preferred to* R_j *under the occurrence of event* G*, then comparison of options* R_i *and* R_j *(which are identical in preference if a different event occurs) depends entirely on consideration of what happens if* G *occurs.*

Postulates 1 to 3 then provide a minimal set of rules to ensure that qualitative comparisons based on the preference $\preceq$ cannot have intuitively undesirable implications. But we also need to introduce some form of quantification by setting up a standard unit of measurement such that enables the transmission system to assign a numerical value to any given SOT in the selection problem. In short, precision through quantification is achieved by introducing some form of numerical standard into the system already equipped with a coherent qualitative ordering relation (Postulates 1 to 3). We shall regard it as essential to be able to aspire to some kind of quantitative precision in the context of comparing spatial-orientation trees. It is therefore necessary that we have available some form of standard SOTs. This notion of quantification is given in the following two axioms:

Postulate 4. *In the transmission system, there exists some form of standard spatial-orientation trees, which will play a role analogous to the standard units of measurement.*

Postulate 5. *The standard family of spatial-orientation trees provides a continuous scale against which any consequence or event can be precisely compared.*

It is important to emphasize that it is not required the assumption that standard families of SOTs actually, physically exist. It is only required that it can be invoked such a set up as a mental image.

The following result serves to determine how numerical measures can be assigned to two of the elements of the SOT selection problem in the form of probabilities for graylevel occurrence events and utilities for consequences.

Proposition 1. *Any transmission system that aspires to analyze the SOT selection problem* $\{\mathcal{R}, \mathcal{G}, \mathcal{C}, \preceq\}$ *at any truncation time* t *in accordance with Postulates 1 to 5 should verify that:*
(a) Degrees of belief about gray-level occurrence sets $\{G_{l,i}; l \in L\}$ *are represented in the form of finite probability distributions* $\{p(G_{l,i} \mid S(R_i, t)); l \in L\}$*; where* $S(R_i, t) = BS_j\, BS_{j+1} \cdots BS_{j+n}$ *are the bit streams transmitted at time* t *for* R_i*, and with* $p(G_{l,i} \mid S(R_i, t))$ *denoting the probability of gray level* l *in the spatial region associated with SOT* R_i *and which was reconstructed using the bitstreams* $S(R_i, t)$ *in addition to the those transmitted before time* t*;*
(b) Numerical values attached to the consequences $\{c_{l,i}; l \in L\}$ *foreseen if a particular SOT* R_i *is taken are represented in the form of a utility function.*

Proof.

(a) Follows from Proposition 2.11 in [13] which establishes formally that coherent, quantitative measures of uncertainty about events must take the form of probabilities;

(b) follows from Propositions 2.21 and 2.23 in [13].

In conclusion, Proposition 1(a) demonstrates that, to avoid certain forms of behavioral inconsistency when a transmission system chooses, at any truncation time, among available spatial-orientation trees for transmission, the gray-level occurrence sets $\mathcal{G}_i \equiv \{G_{l,i}; l \in L\}$ should be represented by probability distributions $\mathcal{P}_i \equiv \{p(G_{l,i} \mid S(R_i, t)); l \in L\}$. In such a framework, the actions available to the system are the various probability distributions $\mathcal{P}_i$ over $\mathcal{G}_i$, the latter constituting the graylevel occurrence set corresponding to each SOT selection.

Also we know from Proposition 1(b) that options in the SOT selection cannot be ordered without a specification of utilities (numerical values) for the consequences. Thus, next we complete the specification of this decision problem by inducing the preference ordering through the introduction of a particular form of utility function $u(\cdot)$, which describes the numerical value $u(\mathcal{P}_i, G_{l,i})$; where the set of consequences $\mathcal{C}$ consists of all pairs $(\mathcal{P}_i, G_{l,i})$ representing the conjunctions of probability distributions and true graylevel occurrence.

C. MAXIMUM EXPECTED UTILITY INCREASE METHODOLOGY FOR SOT SELECTION

Reference [13] investigates the properties which such a function should possess in order to describe *a preference pattern which accords with what a scientific community ought to demand.* This special class of utility functions is often referred to as the class of score functions:

Definition 2. *A score function is appropriate in reporting pure scientific research if it satisfies three conditions:*

(i) the function $u(\mathcal{P}_i, G_{l,i})$ is smooth in the sense that small changes in the distribution to

produce only small changes in the obtained score; that is, it is continuously differentiable as a function of each $p(G_{l,i} \mid S(R_i,t))$ *in* $\mathcal{P}_i$*;*

(ii) a score function u *is proper if the supremum over* $\mathcal{P}$ *is attained if, and only if, it is computed on the actual distributions for the alternative graylevel occurrence events at the transmission time;*

(iii) the function u *is local if the value of a distribution* $\mathcal{P}$ *is to be assessed in terms of the probability it assigned to the actual outcome.*

The following result gives a characterization of proper local score functions:

Proposition 2. *If* u *is a smooth, proper, local score function for probability distributions then it must be of the form:*

$$u(\mathcal{P}_i, G_{l,i}) = A \log p_{l,i} + B_j \tag{1}$$

where $A > 0$ *and* B_j *are arbitrary constants; and* $p_{l,i} = p(G_{l,i} \mid S(R_i,t))$*.*

Proof. See Proposition 2.29 in [13].

As a result of Propositions 1 and 2 we can now provide, on the basis of the expected utility, an optimal solution to the problem of selecting, at time t, a spatial-orientation tree R_i for transmission of a number of bit streams. To this aim, firstly, we need to characterize the form of the expected increase in utility provided by transmitting $S(R_i,t)$ at time t for R_i, and secondly, we simply select for transmission at time t the bit streams $S(R_i,t)$ associated with the maximum expected increase in utility (per coding bit) over available SOT R_i.

Proposition 3. *Let* $p(G_{l,i} \mid t-1)$ *be the probability of gray level* l *in the spatial region associated with SOT* R_i *and which was reconstructed using the bit streams* $S(R_i,0),\cdots,S(R_i,t-1)$ *transmitted before time* t*. Let* $p(G_{l,i} \mid S(R_i,t))$ *be the probability of gray level* l *in the spatial region which was reconstructed using the bit streams* $S(R_i,t)$ *candidates to be transmitted at time* t*, in addition to bitstreams transmitted, before the time* t*, for* R_i*.*

If preferences are described by logarithmic score functions for the classes of probability *distributions $\mathcal{P}$, then the expected increase in utility provided by the transmission at time t of bit streams $S(R_i,t)$, when the initial probability distribution $\mathcal{Q} \equiv \{p(G_{l,i} \mid t-1); l \in L\}$ is strictly positive, is given by (up to a nonnegative multiplicative constant):*

$$I(S(R_i,t)/\mathcal{Q}) = \sum_l p(G_{l,i} \mid S(R_i,t)) \log \frac{p(G_{l,i} \mid S(R_i,t))}{p(G_{l,i} \mid t-1)} \tag{2}$$

where $I(S(R_i,t)/\mathcal{Q})$ is non-negative and is zero if, and only if, for all l, $p(G_{l,i} \mid t-1) = p(G_{l,i} \mid S(R_i,t))$.

Proof. From Proposition 2, we have the utilities of reporting $p(G_{l,i} \mid t-1)$ or $p(G_{l,i} \mid S(R_i,t))$ would be $A \log p(G_{l,i} \mid t-1) + B_j$ and $A \log p(G_{l,i} \mid S(R_i,t)) + B_j$, respectively. Thus, conditional on new data $S(R_i,t)$, the expected increase in utility provided by bit streams $S(R_i,t)$ is given by

$$\begin{aligned} I(S(R_i,t)/\mathcal{Q}) &= \sum_l p(G_{l,i} \mid S(R_i,t))[(A \log p(G_{l,i} \mid S(R_i,t)) + B_j) - (A \log p(G_{l,i} \mid t-1) + B_j)] \\ &= A \sum_l p(G_{l,i} \mid S(R_i,t)) \log \frac{p(G_{l,i} \mid S(R_i,t))}{p(G_{l,i} \mid t-1)}, \end{aligned}$$

which, by Theorem 1 in [6], is non-negative and verifies the nilpotence condition.

This proves Proposition 3.

This result motivates an operational procedure for selection of bit streams $S(R_\dagger,t)$ to be transmitted at truncation time t on a utility basis: In accordance with Postulates 1 to 5 as well as utility constraints given by Definition 2, *the rational choice for transmission at truncation time t is to select bit streams $S(R_\dagger,t)$ which provides the maximum achievable expected increase in utility $I(S(R_i,t)/\mathcal{Q})$ per coding bit, for any available spatial-orientation tree R_i*

$$S(R_\dagger,t) = \arg_{S(R_i,t)} \left\{ \max_i \max_{S(R_i,t)} \frac{I(S(R_i,t)/\mathcal{Q})}{\#S(R_i,t)} \right\} \tag{3}$$

where $\#S(R_i,t)$ denotes the number of bits in $S(R_i,t)$; and $I(S(R_i,t)/\mathcal{Q})$ is defined as given in equation (2).

It follows from Theorem 1 in reference [6] that the expected increase in utility provided by bit streams $S(R_i,t)$ is a measure of the information gain of the spatial region reconstructed using $S(R_i, t)$ in addition to $S(R_i, 0), \cdots, S(R_i, t-1)$ from the spatial region which was reconstructed using $S(R_i, 0), \cdots, S(R_i, t-1)$ transmitted before the time t. That is, the expected increase in utility as given in equation (2) is the Kullback-Leibler information gain which has a minimal number of properties that are natural, and, thus, desirable for predicting visual distinctness from 2D digital images (see [6] for further details).

III. CODING ALGORITHM

We have developed an algorithm which calculates the expected increase in utility with just a lookup table operation. It permits us to efficiently implement the utility-per-coding-bit optimization. An interesting property of embedded coding is that when the encoding or decoding is terminated during a pass for a particular spatial orientation tree, there are no artifacts produced that would indicate where the termination occurs. In other words, coefficients in some spatial orientation trees are represented with different precision of the others. The encoding algorithm can be stopped at any compressed file size or let run until the compressed file is a representation of a nearly lossless image (the compression may not be reversible as the wavelet transform filters chosen for lossy compression have non-integer tap weights and produce non-integer transform coefficients which are truncated to finite precision). Furthermore, since no training of any kind is required, the algorithm is fairly general.

If rate control optimization is based on the actual expected increase in utility per coding bit, the order of coding has to be transmitted to the decoder. To transmitting the coding order, the SOT descriptor and the number n of bitstreams $BS_j\, BS_{j+1} \cdots BS_{j+n}$ to be deliver to the decoder are encoded in the bit stream header. But the ordering data within the particular spatial orientation tree is not explicitly transmitted because of the utilization of the set partitioning sorting algorithm. Only if too many spatial orientation trees are available, the overhead to locate the bit streams with the largest actual expected increase in utility per coding bit begins to become a significant portion of the overall bit stream. In fact the coding scheme is so effective that it was not improved by entropy-coding its output, and thus, the bitstreams put out with the coding algorithm are binary uncoded (without entropy coding).

The progressive transmission method outlined above can be implemented with the coder and decoder algorithms as described in Appendix. The software and documentation may be accessed in the Internet site with URL of `http://decsai.ugr.es/diata/REWIC` or by anonymous ftp to `decsai.ugr.es` with the path `pub/REWIC` in the compressed archive file `rewic.tar.gz`. All materials are made available to other researchers for academic use only. The programs were not optimized to a commercial application level.

IV. COMPARATIVE STUDY

The purpose of this section is to assess the comparative performance of two progressive compression schemes using a proper coder selection procedure. A traditional approach for evaluating the comparative performance of lossy compression methods in an application is to simulate the application in a carefully designed experiment, gather necessary data in a way that interferes with the simulation as little as possible, and analyze the resulting data to assess the validity of the hypothesis in the application. This kind of evaluation is usually impractical both because of the funding constraints and because of the time required for thorough experiments and subsequent statistical analysis. It implies that a protocol for evaluation of lossy compression methods needs an acceptable technology for selecting coders deemed good enough. And the intent of the following subsection is exactly to develop such a technology.

A. CODER SELECTION PROCEDURE

The problem of selecting a coder for potential use in lossy compression of digital images may be solved using a suitable rate-distortion measure. In practice, however, this procedure may be difficult to apply for two reasons: 1) The lack of knowledge of what distortion measure is more suitable for digital images, [6], [7]; and 2) the behavior of image coding is not well understood for so-called non-difference distortion measures where the distortion is not measured by a function of the difference between the source and the decoded images, however non-difference distortion measures occur naturally in image coding, [14], [15].

Reference [6] suggests a new method for characterizing information of a target relative to its background. The resultant computational measure was applied to quantify the visual distinctness of targets in complex natural backgrounds from digital imagery. A generalization of the Kullback-Leibler joint information gain of various random variables, was shown to correlate strongly with visual distinctness as estimated by human observers. Bootstrap methods for assessing statistical accuracy were used to produce this inference. In the language of statistical hypothesis testing, the following postulates were found to be relevant for predicting visual distinctness between a reference image and another of input:

- **Postulate 1.** *A measure of how unexpected a single event of the reference image was, depends only upon its probability.*
- **Postulate 2.** *An estimate of how unexpected the reference image was from some probability distribution is simply defined as the mathematical expectation of how unexpected its single events were from this distribution.*
- **Postulate 3.** *The reference image is more unexpected from an "estimated" distribution than from the "true" distribution.*
- **Postulate 4: The effectiveness of the information.** *If the reference image contains relevant structures, unwanted detail and noise, the reduction of data without significant ensuing loss of information is of primary importance in the comparison of any input image with the reference image.*
- **Postulate 5: The information conservation constraint.** *If local information of the reference image at significant points is some constraint on the input image, then the selective measure of the error between them involves three steps: (i) detecting interest points in the reference image; (ii) for each interest point, zooming in (which is equivalent to modifying the scale parameter); and (iii) local comparison of the input image with the reference image.*
- **Postulate 6: The significance conservation constraint.** *If interest points of the reference image and their significance is some given constraint on the input image, then the*

selective information gain is also a function of the discrepancy between the significance of interest points in the reference image and their significance in the input scene.

From the results given in [6], a measure of information for predicting visual distinctness may satisfy the property of additivity, whereas the properties of nonnegativity and symmetry are not essential. The compound gain $\mathcal{E}_C^{Z_1,\cdots,Z_n}$ is a measure of information gain between two images such that, it satisfies Postulates 1 through 6. Also it has the property of additivity. The main result of the study in [6] is the finding that the compound gain appears to relate to visual target distinctness as perceived by human observers. This result implies that it might be used to predict visual distinctness from digital imagery.

Given two lossy coders $\mathcal{L}_\alpha$ and $\mathcal{L}_\beta$, let $\{I^\alpha_{q(1)}, \cdots, I^\alpha_{q(K)}\}$ be the set of decoded images at compression ratios $q(1), \cdots, q(K)$ using coder $\mathcal{L}_\alpha$; $\{I^\beta_{q(1)}, \cdots, I^\beta_{q(K)}\}$ be the set of decoded images at compression ratios $q(1), \cdots, q(K)$ using coder $\mathcal{L}_\beta$.

The compound gain $\mathcal{E}_C^{Z_1,\cdots,Z_n}$ may then be applied to quantify the image distortion by means of the difference between the original image I and the images reconstructed under various degrees of lossy compression:

$$f(\alpha, i) = \mathcal{E}_C^{Z_1,\cdots,Z_n}(I, I^\alpha_{q(i)}) \tag{4}$$

and $f(\beta, i)$ is defined similarly.

Once distortion functions $f(\dagger, i)$ have been calculated from reconstructed images given by each decoding algorithm, we may consider an objective criterion for coder selection based on the overall difference between these two functions $f(\alpha, i)$ and $f(\beta, i)$, which can be measured by a Kolmogorov-Smirnov (K-S) test to a certain required level of significance.

Definition 3: Coder Selection Procedure. *In the language of statistical hypothesis testing, the coding scheme $\mathcal{L}_\alpha$ is significantly better than $\mathcal{L}_\beta$ if the following two conditions are true:*

(1) $f(\alpha, i) \leq f(\beta, i)$, with $i = 1, 2, \cdots, K$; and

(2) We can disprove, to a certain required level of significance, the null hypothesis of a Kolmogorov-Smirnov test that two data sets $\{f(\alpha, i) \mid i = 1, 2, \cdots, K\}$ and $\{f(\beta, i) \mid i =$

$1, 2, \cdots, K\}$ *are drawn from the same population distribution function.*

Hence, optimal coders $\mathcal{L}_\alpha$ tends to produce the lowest value of $f(\dagger, i)$, with $i = 1, 2, \cdots, K$, for any lossy coder in the class being compared. And clearly, we should not seriously consider any criterion for selecting a lossy coder which might select $\mathcal{L}_\beta$ when $f(\alpha, i) \leq f(\beta, i)$ at a huge number of compression ratios $q(i)$. The proposed criterion for coder selection meets this condition.

As suggested in Definition 3, disproving the null hypothesis in effect proves data sets $\{f(\alpha, i) \mid i = 1, 2, \cdots, K\}$ and $\{f(\beta, i) \mid i = 1, 2, \cdots, K\}$ are from different distributions. The Kolmogorov-Smirnov test is applicable to data sets where each data point can be associated with a single number. In such cases the list of data points can be easily converted to an unbiased estimator of the cumulative distribution function from which it was drawn. Different data sets give different cumulative distributions. However, all cumulative distributions functions agree at the smallest allowable value of x (where they are zero), and the largest allowable value of x (where they are unity). So it is the behavior between the largest and smallest values that distinguishes distributions.

If both conditions in Definition 3 hold, it allows us to assess the fact that dataset $\{f(\alpha, i) \mid i = 1, 2, \cdots, K\}$ is significantly better than dataset $\{f(\beta, i) \mid i = 1, 2, \cdots, K\}$.

B. NUMERICAL RESULTS

Experiments here reported were performed on standard 512×512 grayscale test images: "Boat", "Baboon", "Goldhill", and "Bike". They are shown in Figure 3. Additional coding results (on "Barbara", remote-sensing images, biomedical images, military vehicle images, among others) can be found at the website `http://decsai.ugr.es/diata/REWIC`. The images are decomposed by a 6-level 9-7 tap biorthogonal Daubechies filter [16]. It produces 16 spatial orientation trees available for transmission (see Figure 2). The images are then compressed by the set partitioning in hierarchical trees (SPIHT, proposed by Said and Pearlman in [4]) and the rational embedded wavelet image coder (REWIC). Results are obtained without entropy-coding the bits put out with both REWIC and SPIHT algorithms. The well-respected SPIHT coder is used here as a reference of the state of art coder in progressive transmission. We therefore compare REWIC with SPIHT to show the advantage of the proposed rate control strategy.

Subjective comparisons of the coded images at different bit rates are illustrated in Figures 4-8. For clarity, only selected portions of the images are shown. Visual comparisons show that the subjective qualities of REWIC-coded test images are better than that of SPIHT. What is clearly evident is that the rational coder is better able to capture important objects of the original image and their natural texture, while reducing the ringing distortion seen in SPIHT results.

In the second set of experiments, we perform a more thorough comparison of $REWIC$ and $SPIHT$ based on the coder selection procedure as given in Definition 3.

The *first step* of the coder selection procedure consists of calculating, for each coder, function $f(\dagger, i)$ which quantifies the image distortion by means of the compound gain between the original image I and images reconstructed under various degrees of lossy compression (from 0.0625 b/pixel to 0.5 b/pixel). It allows us to analyze the behavior of image coders at different levels of performance across compression ratios.

Figure 9 shows, for each test image, plots on rate-distortion as given by $f(\dagger, i) = \mathcal{E}_C^{Z_1,\cdots,Z_n}(I, I^{\dagger}_{q(i)})$ for SPIHT and REWIC. It is worthwhile to point out that, for experiments performed on Goldhill, Baboon, and Boat, we have the result $f(\text{REWIC}, i) \leq f(\text{SPIHT}, i)$ from 0.0625 bbp to 0.5 bpp. Furthermore, on Bike image coded from 0.090 bpp to 0.5 bpp, we have a similar result. The overall impression is that at high compression ratios (i.e., bit rates below 0.5 bit/pixel) REWIC is best able to maintain acceptable image fidelity. At low compression ratios both coders seem to perform at about the same level for the test images, however, REWIC is better able to capture the natural texture of the original.

The *second step* of the coder selection procedure consists of characterizing an optimum coder on hypothesis testing basis, as suggested by Definition 3. The overall difference between functions $f(\text{REWIC}, i)$ and $f(\text{SPIHT}, i)$ can be measured by subtracting the area beneath

Kolmogorov-Smirnov (K-S) test results		
Image	K-S Statistic	Probability
bike	0.079646	0.851605
goldhill	0.176991	0.051066
boat	0.185841	0.035064
baboon	0.292035	0.000092

TABLE IV

K-S STATISTICS AND SIGNIFICANCE LEVEL FOR THE NULL HYPOTHESIS THAT THE DATA SETS ARE DRAWN FROM THE SAME DISTRIBUTION

their graphs. This operation corresponds to a Kolmogorov-Smirnov (K-S) test. Based on the K-S test results, it will be then concluded if, in the language of statistical hypothesis testing, REWIC is significantly better for embedded wavelet compression than the state-of-the-art SPIHT.

Table IV illustrates the results of the Kolmogorov-Smirnov test when it is applied for coder selection. From these results we conclude that, on Goldhill, Baboon, and Boat, the optimum coder is REWIC at the required level of significance (with 95% confidence). In the language of statistical hypothesis testing, this means that REWIC is significantly better than SPIHT.

V. CONCLUSION

From the results given in Section IV-A, we claim that the performance of an embedded coder can be optimized by means of a rational strategy for rate control by organizing the progressive transmission with expected increase in utility per coding bit. This scheme allocates the available coding bits first to the optimal number of bit streams, for any spatial-orientation tree, with the maximum achievable expected increase in utility per coding bit. The interesting point is that such a scheme follows from an axiomatic characterization that purports to describe the ways in which a transmission system should aspire to choose in such a way as to avoid certain forms of behavioral inconsistency. This is an interesting finding, since it may eliminate the need for a distortion measure in the trade-off between visual fidelity and rate control.

Here we use a new coder selection procedure to show the advantage of the proposed rate control strategy. The criterion for selecting lossy coders was based on (i) an information theoretic measure for predicting visual distinctness as estimated by human observes and (ii) Kolmogorov-Smirnov tests. The SPIHT coder was used as a reference of the state of art coder in progressive transmission.

Acknowledgments. This research was sponsored by the Dirección General de Enseñanza Superior, Spanish Board for Science and Technology, (CICYT) under grant TIC2000-1421.

References

[1] Blahut, R.E. "Computation of channel capacity and rate distortion functions." *IEEE Trans. on Information Theory,* Vol. 18, pp. 460-473, (1972).

[2] Arimoto, S. "An algorithm for calculating the capacity of an arbitrary discrete memoryless channel." *IEEE Trans. on Information Theory,* Vol. 18, pp. 14-20, (1972).

[3] Csiszar, I. "On the computation of rate distortion functions." *IEEE Trans. on Information Theory,* Vol. 20, pp. 122-124, (1974).

[4] Said, A., and Pearlman, W.A. "A new, fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. on Circuit and System for Video Technology*, Vol. 6, No. 3, pp. 243-250, (1996).

[5] Adelson, E.H., Simoncelli, E., and Hingorani, R. "Orthogonal pyramid transforms for image coding," *Proc. SPIE, Vol. 845 - Visual Communications and Image Processing II;* Cambridge, MA, pp. 50-58, (1987).

[6] Garcia, J.A., Fdez-Valdivia, J., Fdez-Vidal, Xose R., and Rodriguez-Sanchez, Rosa. "Information Theoretic Measure for Visual Target Distinctness." *IEEE Trans. on Pattern Analysis and Machine Intelligence,* Vol. 23, No. 3, March (2001).

[7] Garcia, J.A., Fdez-Valdivia, J., Fdez-Vidal, X.R., and Rodriguez-Sanchez, R. *Computational models for predicting visual target distinctness*, SPIE Optical Engineering Press, Bellingham, Washington USA, (2001).

[8] Karssemeijer, N. "Adaptive noise equalization and image analysis in mammography." in *13th Int. Conf. Inform. Processing Med. Imag.*, Flagstaff, pp. 472-486, (1992).

[9] Suckling, J. et al. "The mammographic Image Analysis Society Digital Mammogram Database." *Experta Medica. International Congress Series*, Vol. 1069, pp. 375-378, (1994).

[10] JTC1 Committee, "Digital compression and coding of continuous-tone still images," Int. Org. Standardization ISO/IEC, JTC1 Committee Draft, JPEG 8-R8, (1990).

[11] Kullback, S. *Information theory and statistics.* Gloucester, Mass. Peter Smith, (1978).

[12] Efron, B. and Tibshirani, R.J., *An introduction to the bootstrap*, Volume 57 in Monographs on Statistics and Applied Probability, Chapman and Hall/CRC, (1993).

[13] Bernardo, Jose M., and, Smith, Adrian F.M., *Bayesian Theory,* Wiley Series in Probability and Statistics, John Wiley and Sons, Ltd., Chichester, (1994).

[14] Linder, T., Zamir, R., Zeger, K. "High-resolution source coding for non-difference distortion measures: multidimensional companding," *IEEE Trans. on Information Theory,* Vol. 45, No. 2, pp. 548-561, (1999).

[15] Li, J., Chaddha, N., and Gray, R.M. "Asymptotic performance of vector quantizers with the perceptual distortion measure," Proc. IEEE Int. Symp. Information Theory, pp. 55, Germany, June, (1997).

[16] Antonini, M., Barlaud, M., Mathieu, P., and, Daubechies, I., "Image coding using wavelet transform," *IEEE Trans. on Image Processing,* Vol. 1, pp. 205-230, (1992).

Appendix

REWIC ALGORITHM

0. *VARIABLES:*
R_i i-**region associated to each SOT**
$R = \{R_i\}$
$d_i = \lfloor \mathbf{log}_2(max_{(n,m)}\{|c_{n,m}|\ |(n,m) \in R_i\}) \rfloor$
LSP_i **(list of significant pixels in** R_i**)**$=\{(n,m)\,|(n,m) \in R_i\ and\ 2^{d_i} \leq |c_{n,m}| < 2^{d_i+1}\}$
LIP_i **list of insignificant pixels in** R_i
LIS_i **(list of insignificant sets in** R_i**)** $=\{D(n,m)\ \ or\ \ L(n,m)\}$
$D(n,m)\ =\{(t,s)\,|(n,m) \in R_i\ and\ (t,s) \in R_i\ with\ (t,s)\ a\ descendant\ of\ the\ node\ (n,m)\}$
$L(n,m)\ =D(n,m) - O(n,m) =$
$\{(t,s))\,|(n,m) \in R_i\ and\ (t,s) \in R_i\ being\ a\ descendant\ but\ not\ an\ offspring\ of\ (n,m)\}$
$O(n,m)$ **set of coordinates of all offspring of (n,m)**
H_i **coordinate of the SOT root node corresponding to** $\mathbf{R}_i$
$MAXITER$ **Maximum number of bit streams to be analyzed at time t.**
Lookuptable **: for each region** R_i **and for each pass, Lookuptable stores the expected increase in utility (I) and the number of bits (BS) used between the last and current pass.**
For each region R_i **Next_Phase**$_i$ **={Sorting or Refinement}**

1. *INITIALIZATION*
- **(a) for each region** $\mathbf{R}_i$
  - $\mathbf{d}_i \leftarrow \lfloor \mathbf{log}_2(max_{(n,m)}\{|c_{n,m}|\ |(n,m) \in R_i\}) \rfloor$
  - **set** $\mathbf{LSP}_i$ **as an empty list**
  - **add the coordinate** $(n,m) \in H_i$ **to the** $\mathbf{LIP}_i$ **list**
  - **add all the coordinates** (n,m) **corresponding to nodes with descendants to the** $\mathbf{LIS}_i$ **list**
- **(b) endfor**
- **(c) Let d**$\leftarrow max_{i=1,\ldots,|R|}(d_i)$
- **(d) output d**
- **(e) for each region** $\mathbf{R}_i$
  - $\mathbf{dif}_i \leftarrow \mathbf{d}\text{-}\mathbf{d}_i$
  - **output** $\mathbf{dif}_i$
  - **n**$\leftarrow$**0**
  - $*$ **while (n <** $\mathbf{dif}_i$**) do**
    - $\mathbf{d}_i \leftarrow \mathbf{d}_i$**-1**
    - **n**$\leftarrow$**n+1**
  - $*$ **endwhile**
  - **Sorting Pass with arguments** $\mathbf{LIS}_i$**,** $\mathbf{LIP}_i$**,**$\mathbf{LSP}_i$ **and** 2^{d_i}
  - **Refinement Pass with arguments** $\mathbf{LSP}_i$ **and** 2^{d_i}
  - $d_i \leftarrow d_i - 1$
- **(f) endfor**
- **(g) for each region** $\mathbf{R}_i$
  - $numberpass_i \leftarrow (\lfloor log_2(d_i) \rfloor + 1) * 2$

- LIS^1aux_i ←**LIS**$_i$; LIP^1aux_i ←**LIP**$_i$; LSP^1aux_i ←**LSP**$_i$
- **current_pass**$_i$ **←0; daux**$_i^1$**←d**$_i$**;** k **←0; next_state = SORTING**

∗ **while** (k <*numberpass*$_i$) **do**
  - $LIS^2aux_i \leftarrow LIS^1aux_i$; $LIP^2aux_i \leftarrow LIP^1aux_i$; $LSP^2aux_i \leftarrow LSP^1aux_i$
  - **daux**$_i^2$**←daux**$_i^1$**; next_state_aux = next_state;** n **←0;**

  ∗ **while** (n < **MAXITER**) **do**
    ⋆ **if (next_state_aux)==SORTING then**
      - **Sorting Pass with arguments** LIS^2aux_i, LIP^2aux_i, LSP^2aux_i, **and** $2^{daux_i^2}$
      - **next_state_aux ← REFINEMENT**

    ⋆ **else if (next_state_aux == REFINEMENT) then**
      - **Refinement Pass with arguments** LSP^2aux_i, **and** $2^{daux_i^2}$
      - **next_state_aux ← SORTING**
      - $daux_i^2 \leftarrow daux_i^2$ **-1;**
    - **Lookuptable**$[i]$**.I**$[r]$← $I\left((BS_{initial}, ..., BS_{initial+n})/\mathcal{Q}\right)$
    - **Lookuptable**$[i]$**.BS**$[r]$ ← $\mid BS_{initial}$ **+...+** $BS_{initial+n} \mid$
    - $n \leftarrow n+1$; $r \leftarrow r+1$;

  ∗ **endwhile**

  ⋆ **if (next_state_aux)==SORTING then**
    - **Sorting Pass with arguments** LIS^1aux_i, LIP^1aux_i, LSP^1aux_i, **and** $2^{daux_i^1}$
    - **next_state ← REFINEMENT**

  ⋆ **else if (next_state == REFINEMENT) then**
    - **Refinement Pass with arguments** LSP^1aux_i, **and** $2^{daux_i^1}$
    - **next_state ← SORTING**
    - $daux_i^1 \leftarrow daux_i^1$ **-1**

∗ **endwhile**

- **endfor**

2. *REGION_SELECTION*

(a) $MaxAllRegion \leftarrow -\infty$

(b) **for each region R**$_i$
  - $U \leftarrow 0$; $B \leftarrow 0$; $max \leftarrow -\infty$; $k \leftarrow 0$

  ∗ **while** ($k < MAXITER$ **and** $(currentpass_i + k) < numberpass_i$ *** MAXITER) do**
    - $r \leftarrow currentpass_i + k$
    - $U \leftarrow Lookuptable[i].I[r]$; $B \leftarrow Lookuptable[i].BS[r]$

    ⋆ **if** ($max < \frac{U}{B}$) **then**
      - $max \leftarrow \frac{U}{B}$; $iteration \leftarrow k$
    - $k \leftarrow k+1$

  ∗ **endwhile**

  ∗ **if** ($MaxAllRegion < max$) **then**
    - $MaxAllRegion \leftarrow max$; $BestRegion \leftarrow i$; $BestIteration \leftarrow iteration$

(c) **endfor**

(d) $currentpass_{BestRegion} \leftarrow currenpass_{BestRegion} + (BestIteration + 1) * MAXITER$

(e) **return** *BestRegion* **and** *BestIteration*

3. *CODING ALGORITHM*

(a) **INITIALIZATION**

(b) **For (all region** R_i**) Next_Phase**$_i$ **← SORTING**

(c) **while (algorithm don't reach the desired bpp ratio) and (there exist any d**$_i \geq 0$**) do**
  - **REGION_SELECTION**
  - $n \leftarrow 0$
  - **output** *BestRegion*; **output** *BestIteration*

  ∗ **while** ($n \leq BestIteration$) **do**
    ⋆ **if (Next_Phase**$_{BestRegion}$ **of region** $R_{BestRegion}$ **)==SORTING then**
      - **Sorting Pass with arguments LIS**$_{BestRegion}$ **, LIP**$_{BestRegion}$**, LSP**$_{BestRegion}$ **and** $2^{d_{BestRegion}}$
      - **Next_Phase**$_{BestRegion}$ **←REFINEMENT**

    ⋆ **else if (Next_Phase**$_{BestRegion}$ **of region** $R_{BestRegion}$ **)==REFINEMENT the**
      - **Refinement Pass with arguments LSP**$_{BestRegion}$ **and** $2^{d_{BestRegion}}$
      - **Next_Phase**$_{BestRegion}$ **←SORTING**

- $\mathbf{d}_{BestRegion} \leftarrow \mathbf{d}_{BestRegion} - 1$
- $n \leftarrow n + 1$

∗ **endwhile**

(d) endwhile

UNREWIC ALGORITHM

---

1. *INITIALIZATION*

(a) for each region $\mathbf{R}_i$
- **set $\mathbf{LSP}_i$ as an empty list**
- **add the coordinate $(n, m) \in H_i$ to the $\mathbf{LIP}_i$ list**
- **add all the coordinates (n, m) corresponding to nodes with descendants to the $\mathbf{LIS}_i$ list**

(b) endfor

(c) input d

(d) for each region $\mathbf{R}_i$
- **input $\mathbf{dif}_i$;** $d_i \leftarrow d$; **n←0**

∗ **while (n < dif $_i$) do**
- **$\mathbf{d}_i$←$\mathbf{d}_i$-1; n←n+1**

∗ **endwhile**
- **Sorting Pass with arguments $\mathbf{LIS}_i$, $\mathbf{LIP}_i$,$\mathbf{LSP}_i$ and 2^{d_i}**
- **Refinement Pass with arguments $\mathbf{LSP}_i$ and 2^{d_i}**
- $d_i \leftarrow d_i - 1$

(e) endfor

2. *DECODING ALGORITHM*

(a) INITIALIZATION

(b) For (all region R_i) Next_Phase$_i$ ← SORTING

(c) while (algorithm don't reach the desired bpp ratio) and (there exist any $\mathbf{d}_i \geq 0$) do
- **input** *BestRegion*; **input** *BestIteration*
- $n \leftarrow 0$

∗ **while ($n \leq$ *BestIteration*) do**

⋆ **if (Next_Phase$_{BestRegion}$ of region $R_{BestRegion}$)==SORTING then**
- **Sorting Pass with arguments $\mathbf{LIS}_{BestRegion}$, $\mathbf{LIP}_{BestRegion}$, $\mathbf{LSP}_{BestRegion}$ and $2^{d_{BestRegion}}$**
- **Next_Phase$_{BestRegion}$ ←REFINEMENT**

⋆ **else if (Next_Phase$_{BestRegion}$ of region $R_{BestRegion}$)==REFINEMENT the**
- **Refinement Pass with arguments $\mathbf{LSP}_{BestRegion}$ and $2^{d_{BestRegion}}$**
- **Next_Phase$_{BestRegion}$ ←SORTING**
- $\mathbf{d}_{BestRegion} \leftarrow \mathbf{d}_{BestRegion} - 1$
- $n \leftarrow n + 1$

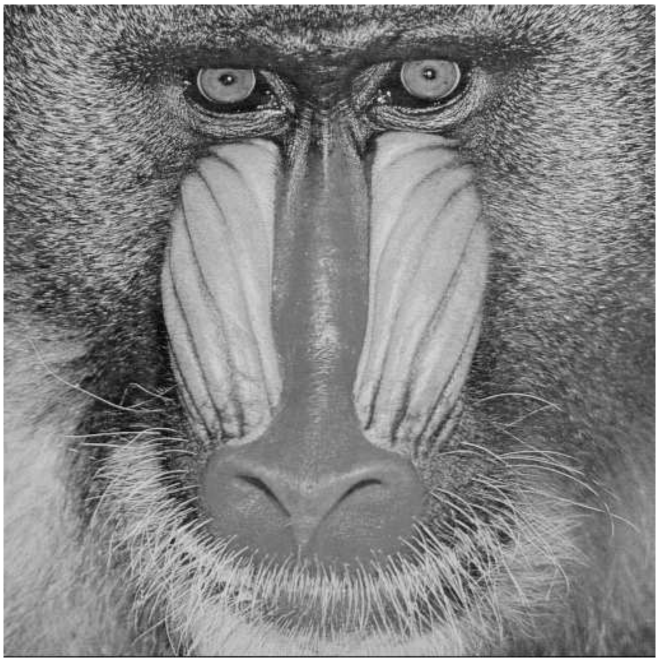
∗ **endwhile**

(d) endwhile

(Sorting Pass and Refinement Pass as defined in [4])

(A)

(B)



(C)

(D)

Fig. 3. Test images: (a) Bike; (b) Goldhill; (c) Baboon; and (d) Boat.

(A)

(B)

Fig. 4. Bike image at 0.2 bpp with (a) SPIHT and (b) the proposed method.

(A)

(B)

Fig. 5. Baboon image at 0.125 bpp with (a) SPIHT and (b) REWIC.

(A)

(B)

Fig. 6. Baboon image at 0.16 bpp with (a) SPIHT and (b) REWIC.

(A)

(B)

Fig. 7. Goldhill image at 0.13 bpp with (a) SPIHT and (b) REWIC.

(A)

(B)

Fig. 8. Boat image at 0.14 bpp with (a) SPIHT and (b) REWIC.

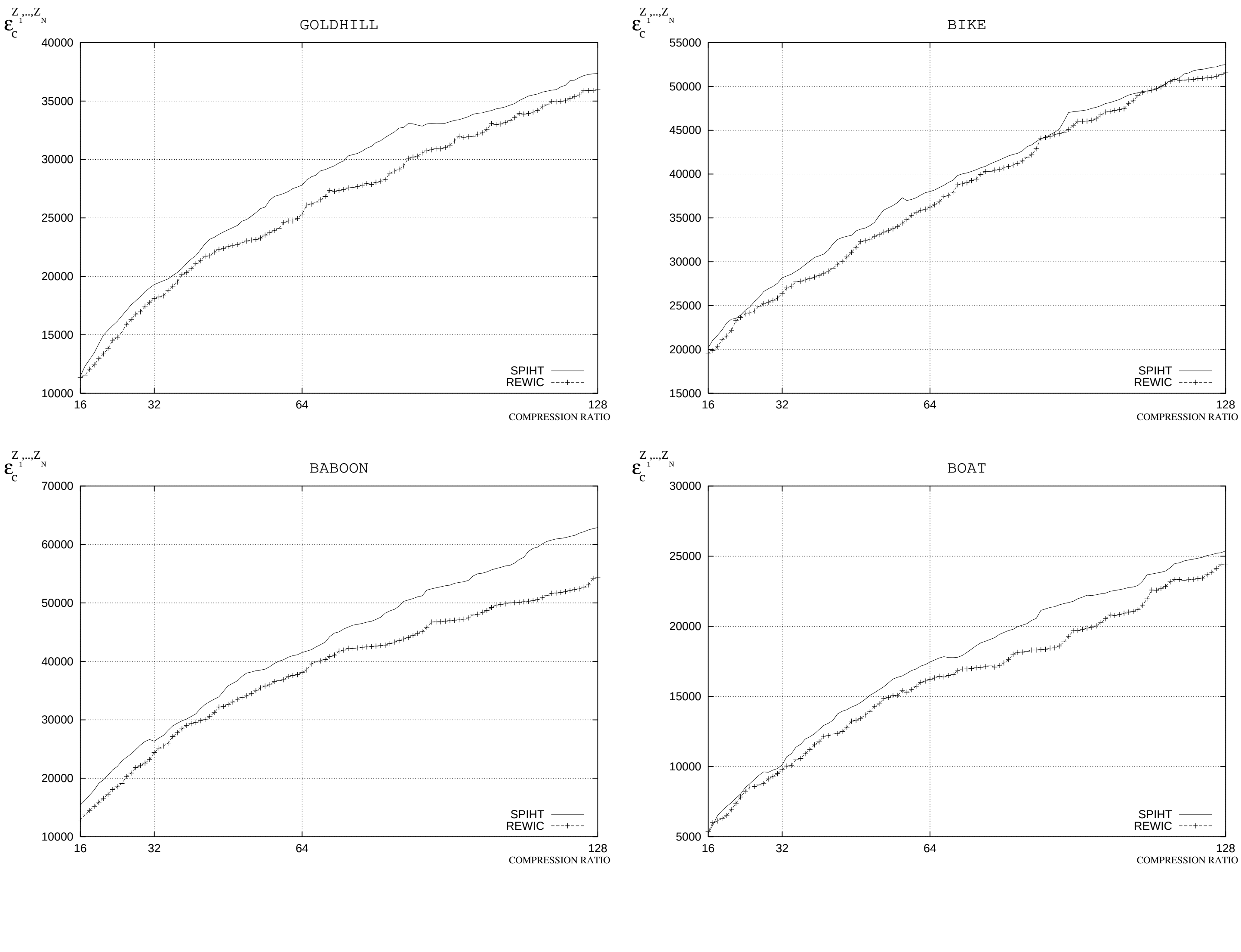


Fig. 9. Figures on rate-distortion for Goldhill, Bike, Baboon, and Boat.