# Rational systems have moderate risk aversion with respect to "gambles" on variable-resolution compression

J.A. Garcia,* Rosa Rodriguez-Sánchez,

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

**Abstract**

In an embedded wavelet scheme for progressive transmission, a tree structure naturally defines the spatial relationship on the hierarchical pyramid. Transform coefficients over each tree correspond to a unique local spatial region of the original image, and they may be coded bit-plane by bit-plane through successive-approximation quantization. After receiving the approximate value of some coefficients, the decoder can obtain a reconstructed image. Here we show a rational system for progressive transmission which, in absence of a priori knowledge about regions of interest, choose at any truncation time among alternative trees for further transmission, in such a way as to avoid certain forms of behavioral inconsistency. We prove that some rational transmission systems might exhibit aversion to risk involving "gambles" on tree-dependent quality of encoding while others favor taking such risks. Based upon an acceptable predictor for visual distinctness from digital imagery, we demonstrate that, without any outside knowledge, risk-prone systems as well as those with strong risk aversion appear not capable to attain the quality of reconstructions which can be achieved with moderate risk-averse behavior.

*Index terms:* Embedded coding, rate control optimization, decision problem, behavioral inconsistency, progressive transmission utility functions, risk aversion, information theoretic measure.

## I. INTRODUCTION

A progressive transmission scheme prioritizes the code bits according to their reduction in distortion (e.g., [1]). If the original image is $I$, the coding is actually done to $C = \Omega(I)$, where $\Omega$ represents a unitary hierarchical subband transformation [2]. The 2D array $C = \{c_{i,j}\}$ has the same dimensions of $I$, and each element $c_{i,j}$ is called transform coefficient

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*

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 often the mean square-error.

In an embedded wavelet scheme for progressive transmission, a tree structure, called spatial orientation tree (SOT) in [1], naturally defines the spatial relationship on the 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 over 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 1.

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 of the sorting pass, $BS_i$, is generated using the set-partitioning approach, [1], 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 paper we show that 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. 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

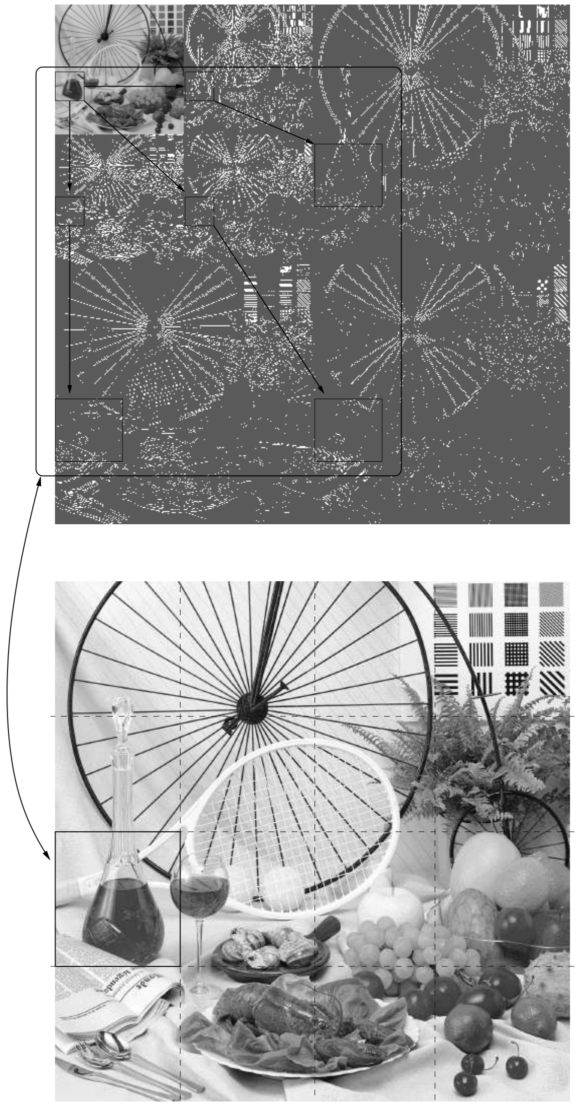

Fig. 1. Each spatial-orientation tree in the hierarchical pyramid corresponds to a spatial region of the original image.

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 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, at any truncation time, a system should choose without any outside knowledge among alternative spatial-orientation trees *in such a way as to avoid certain forms of behavioral inconsistency.*

This paper demonstrates that with several plausible assumptions we may restrict the form of the utility function for ordering preferences in progressive transmission. We conclude that *some rational transmission systems may then exhibit aversion to risk* with respect to "gambles" on SOT-dependent quality of encoding *while others favor taking such risks.* A final section presents a survey designed to analyze, by means of an objective criterion, the behavior of rational systems that exhibits either a risk seeking posture or risk aversion with respect to "gambles" on variable-resolution compression, i.e., some image regions may be retained at high (or full) fidelity, while sacrificing other regions.

## II. BASIC ELEMENTS OF THE SELECTION PROBLEM

We describe any situation in which choices are to be made, at truncation time $t$, among available spatial-orientations trees for their transmission by successive-approximation quantification 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 (i.e., perceived visual fidelity of decoded outcome). 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 at truncation time $t$ 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.

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 (perceived visual fidelity) 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 $\mathcal{R}$.*

## III. 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 [3]).

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 through 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 through 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.* See Appendix.

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 actual graylevel occurrences.

## IV. PROGRESSIVE TRANSMISSION UTILITY FUNCTIONS

This section presents the basic axiomatic characterization of a utility function. The objective is twofold: firstly, to characterize the utility function with a minimal number of properties which are natural and thus desirable; and secondly, to determine the form of all utility functions satisfying these properties which we have stated to be desirable for progressive transmission.

The following postulate states the assumption that the statistical characterization of the decoded output involving information transmitted over $R_i$ up to this time is independent of the characterization of the output corresponding to another SOT $R_j$.

**Postulate 6.** *Let* $R_i$ *and* $R_j$ *be any pair of spatial orientation trees (SOTs). Preferences for decoded outcomes at truncation time* $t$ *involving the two SOTs* $R_i$ *and* $R_j$ *depend only on the* *probability distributions that characterize reconstructions using the information transmitted for $R_i$ and $R_j$, respectively, up to this time $t$, and not on their joint probability distribution.*

In the absence of a priori knowledge about regions of interest, experience with progressive transmission indicates that, for any image region, preferences are for reconstructions of this region leading to the highest level of fidelity that is perceptible within this image portion and so emphasizing its particular features, with independence of the level of fidelity in any other region. It indicates that the value independence between reconstructions for different SOTs is a plausible assumption *without any outside knowledge about regions of interest.* Even if this property does not hold exactly, it leads to simplifications that make it a useful working hypothesis.

We next invoke another specific preference pattern that may arose in the SOT selection process at any truncation time: Suppose that if the transmission system was willing to sacrifice a fraction of reconstruction fidelity corresponding to $R_i$ for an improvement in quality corresponding to $R_j$ then it would be also be willing to sacrifice the same fraction of reconstruction fidelity corresponding to $R_i$ with independence of the remaining information to be transmitted over $R_i$. Again this may be a plausible hypothesis but only if there is no a priori knowledge for the problem at hand.

**Postulate 7.** *The proportion of remaining data in $R_i$ still not delivered to the decoder which the transmission system is willing to give up at truncation time $t$ for an improvement in information transmitted over $R_j$ does not depend on the absolute amount of remaining information over $R_i$ involved.*

We now demonstrate that these two assumptions restrict the form of the utility function for consequences in delivering specific information over $R_i$:

**Proposition 2.** *If Postulates 6 and 7 both hold then a well-behaved utility function $u$ for consequences must have one of the following three functional forms ("well-behaved" means local, twice differentiable and that $\lim_{p\to 0} p\frac{u''(p)}{u'(p)}$ exists):*

$$\begin{aligned} u_1(\mathcal{P}_i, G_{l,i}) &= -(p_{l,i})^r \ , \quad \textit{for } r<0 \\ u_2(\mathcal{P}_i, G_{l,i}) &= \log p_{l,i} \ , \quad \textit{for } r=0 \\ u_3(\mathcal{P}_i, G_{l,i}) &= (p_{l,i})^r \ , \quad \textit{for } r>0 \end{aligned}$$

*where $p_{l,i} = p(G_{l,i} \mid S(R_i,t))$. If $r>1$ the resultant transmission system exhibits a risk*

*seeking posture with respect to "gambles" on SOT-dependent quality of encoding; whereas* $r < 1$ *implies risk averse behavior.*
*Proof.* See Appendix.

From Proposition 2 we know that some rational transmission systems may then exhibit aversion to risk involving "gambles" on SOT-dependent quality of encoding while others favor taking such risks.

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 derive alternative forms 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 spatial orientation trees.

**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* $R_i$*.*

*In a rational transmission system for which Postulates 6 and 7 hold, the possible functional forms for 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, are as follows:*

$$
\begin{aligned}
I(S(R_i, t)/\mathcal{Q}) &= \sum_l p(G_{l,i} \mid S(R_i, t)) \left[(p(G_{l,i} \mid t-1))^r - (p(G_{l,i} \mid S(R_i, t))^r\right], \quad \textit{for } r < 0 \\
I(S(R_i, t)/\mathcal{Q}) &= \sum_l p(G_{l,i} \mid S(R_i, t)) \log \frac{p(G_{l,i} \mid S(R_i, t))}{p(G_{l,i} \mid t-1)}, \quad \textit{for } r = 0 \\
I(S(R_i, t)/\mathcal{Q}) &= \sum_l p(G_{l,i} \mid S(R_i, t)) \left[(p(G_{l,i} \mid S(R_i, t))^r - (p(G_{l,i} \mid t-1))^r\right], \quad \textit{for } r > 0
\end{aligned}
$$

*where if* $r > 1$ *the system exhibits a risk seeking posture with respect to "gambles" on SOT-dependent quality of encoding while* $r < 1$ *implies risk aversion.*
*Proof.* See Appendix.

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 through 7, *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 $R_i$*

$$S(R_\dagger, t) = \arg_{S(R_i,t)} \left\{ \max_i \max_{S(R_i,t)} \frac{I(S(R_i,t)/\mathcal{Q})}{\#S(R_i,t)} \right\} \tag{1}$$

where $\#S(R_i,t)$ denotes the number of bits in $S(R_i,t)$.

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 at any truncation time 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[1].

We have developed an algorithm[7] 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. 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 are binary uncoded (without entropy coding). 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). 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.

## V. ASSESSMENT OF RISK ATTITUDES

This section is intended to elicit information on risk attitudes with respect to "gambles" on SOT-dependent quality of encoding. Experiments here reported were performed on a dataset composed of eighteen standard $512 \times 512$ grayscale test images as shown in Figure 2. Each image is decomposed by a 6-level 9-7 tap biorthogonal Daubechies filter [8]. It produces


Fig. 2. Dataset of standard $512 \times 512$ grayscale test images.

16 spatial orientation trees available for transmission (see Figure 1). The images are then compressed by the rational embedded wavelet image coder (REWIC) for $r$ values within a reasonable range for risk attitudes. This range for risk behavior is between $-10$ and 10, with $r = -10$ implying strong risk aversion and $r = 10$ implying high risk proneness. The risk neutrality is given by $r = 1$.

The images are also compressed by the set partitioning in hierarchical trees (SPIHT, proposed by Said and Pearlman in [1]). Results were 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 the art coder in progressive transmission.

A rational transmission system which has a certain risk attitude with respect to "gambles" on SOT-dependent quality of encoding, as given by its corresponding $r$ value, might be assessed by means of a predictor for visual distinctness of image reconstructions at different bit rates. In the coding community distortion measures like peak signal to noise ratio, root mean square error, and less often mean absolute error, 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[9],[10],[11],[12],[13],[14],[15]. Reference [9] suggests a new method for characterizing information of a target relative to its background. The resultant computational measure was there applied to quantify the visual distinctness of targets in complex natural backgrounds from digital imagery. The main result of the study in [9] is the finding that a generalization of the Kullback-Leibler joint information gain of various random variables, called "compound gain" in [9], appears to relate to visual distinctness as perceived by human observers. Bootstrap methods for assessing statistical accuracy were used to produce this inference. This result implies that the compound gain might be used to predict visual distinctness from digital imagery.

The form of the compound gain $\mathcal{E}_C^{Z_1,\cdots,Z_n}$ between a test image $I$ and decoded outcome $O$ is:

$$\mathcal{E}_C^{Z_1,\cdots,Z_n}(I,O) = \sum_{i=1}^{n} \sum_{l} p(I_{Z_i}) p(l/I_{Z_i}) \log \frac{p(I_{Z_i}) p(l/I_{Z_i})}{p(O_{Z_i}) p(l/O_{Z_i})} \tag{2}$$

with $Z_1, \cdots, Z_n$ being the significant locations of the test image $I$; $(p(l/I_{Z_i}))_l$ being the local histogram computed on a neighborhood of location $Z_i$ in the test image $I$; $(p(l/O_{Z_i}))_l$ being the local histogram computed on a neighborhood of $Z_i$ in the decoded outcome $O$. In the above equation, $I_{Z_i}$ and $O_{Z_i}$ denote the events that the feature at location $Z_i$ is highly significant in order to explain the information content of the test image $I$ and the reconstruction $O$, respectively; $p(I_{Z_i})$ and $p(O_{Z_i})$ being the a priori probabilities of occurrence of $I_{Z_i}$ and $O_{Z_i}$, respectively.

For any rational embedded wavelet coder $\mathcal{L}_r$ with a certain risk attitude as implied by the risk aversion parameter $r$, let $\{I^r_{q(1)}, \cdots, I^r_{q(K)}\}$ be the set of decoded images at bit rates $q(1), \cdots, q(K)$, respectively. The compound gain $\mathcal{E}_C^{Z_1,\cdots,Z_n}$ may then be applied to quantify the visual distinctness by means of the difference between the original image $I$ and decoded images at various bit rates:

$$f(r,i) = \mathcal{E}_C^{Z_1,\cdots,Z_n}(I, I^r_{q(i)})\,. \tag{3}$$

For any test image, the elicitation of the optimal risk attitude for a rational system consists of calculating, for all coders $\mathcal{L}_r$, function $f(r,i)$ which quantifies the visual distinctness by means of the compound gain between the test image $I$ and images reconstructed under various degrees of lossy compression. It allows us to analyze the behavior of rational coders, for all risk postures $r$, across bit rates. Optimal coders $\mathcal{L}_r$ tend to produce the lowest value of $f(r,i)$, for all $i = 1, 2, \cdots, K$, over lossy coders in the class being compared.

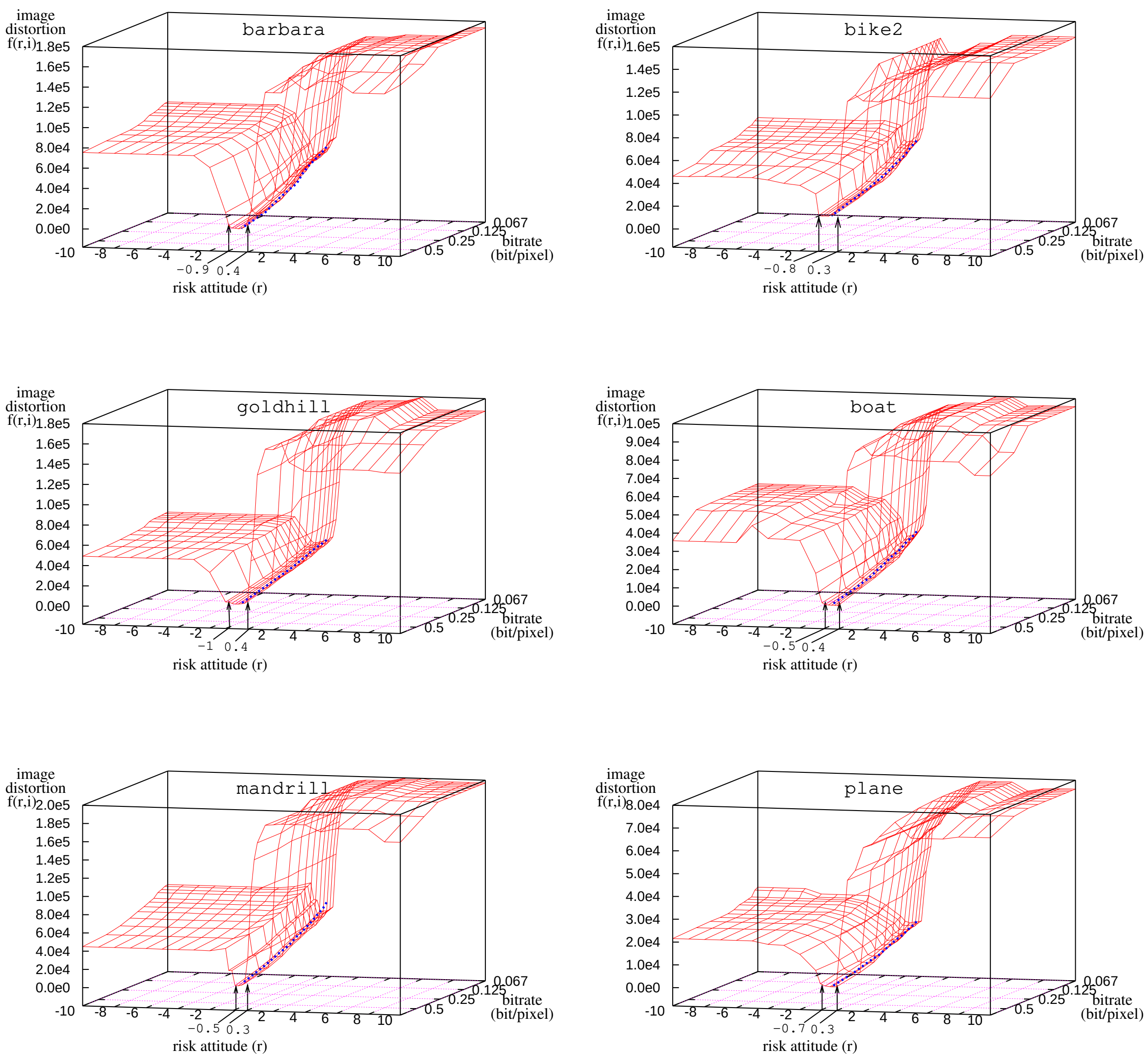


Fig. 3. 3-D plots on attitude-rate-distortion. Results for SPIHT are given by superimposing blue lines of dots.

Figures 3–5 show, for test images in the dataset, 3-D plots on attitude-rate-distortion as given by $f(r,i)$ for coders $\mathcal{L}_r$ with $-10 \leq r \leq 10$ and for bit rates from 0.067 bit/pixel to 0.8 bit/pixel. Results for SPIHT (i.e., $f(SPIHT,i)$ function) are illustrated by superimposing blue lines of dots onto 3D plots.

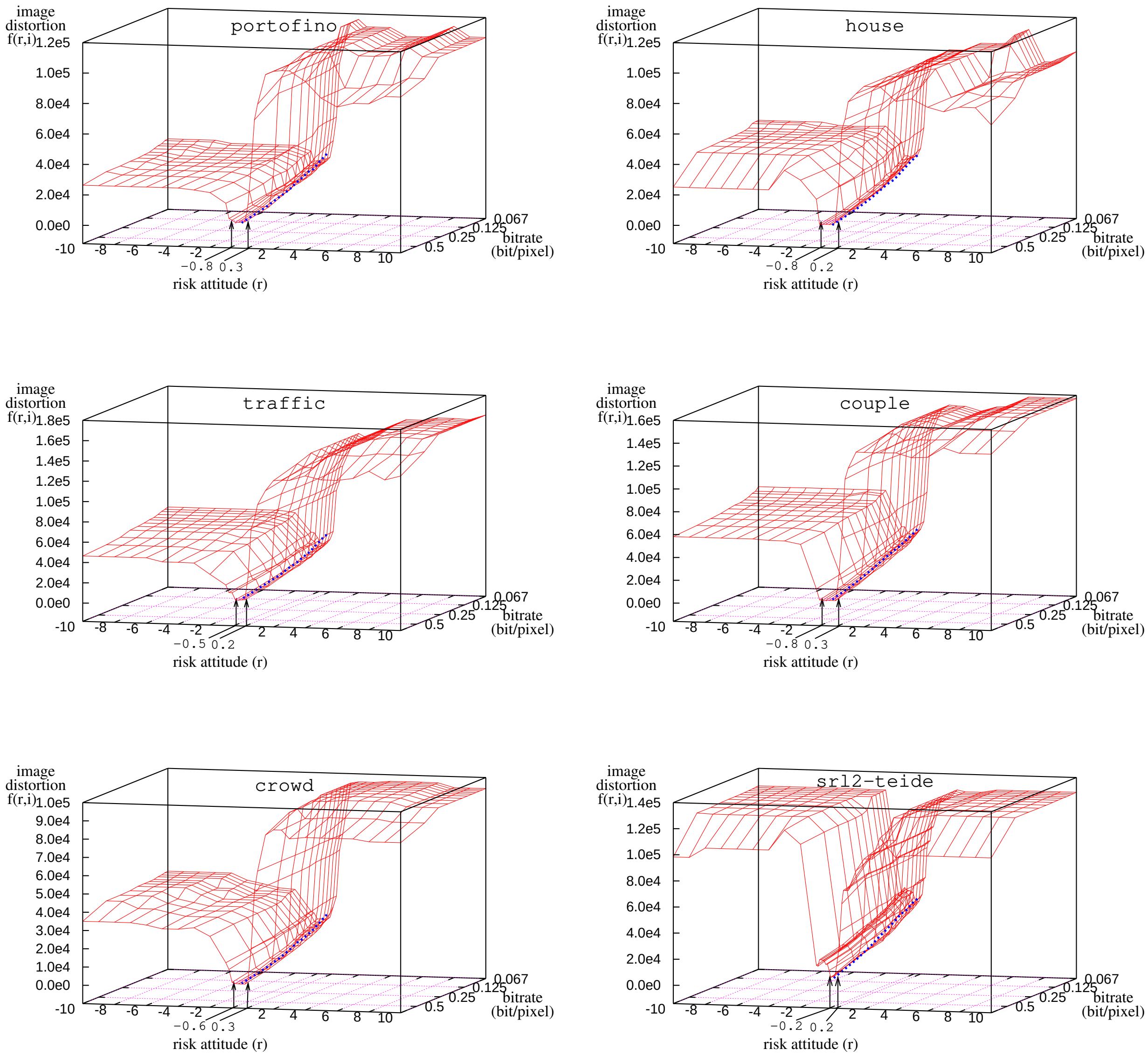


Fig. 4. 3-D plots on attitude-rate-distortion. Results for SPIHT are given by superimposing blue lines of dots.

The overall impression is that, at moderate risk aversion to “gambles” on variable-resolution compression, a rational embedded wavelet coder (REWIC) is best able to maintain acceptable image fidelity, as predicted by $f(r,i)$. For all test images, *in absence of a priori knowledge about regions of interest*, optimal systems for progressive transmission exhibited risk aversion with parameter $r$ around 0. These results also show that risk-prone systems as well as those with strong risk aversion appear not capable to attain the quality of reconstructions that was achieved with moderate risk-averse behavior.

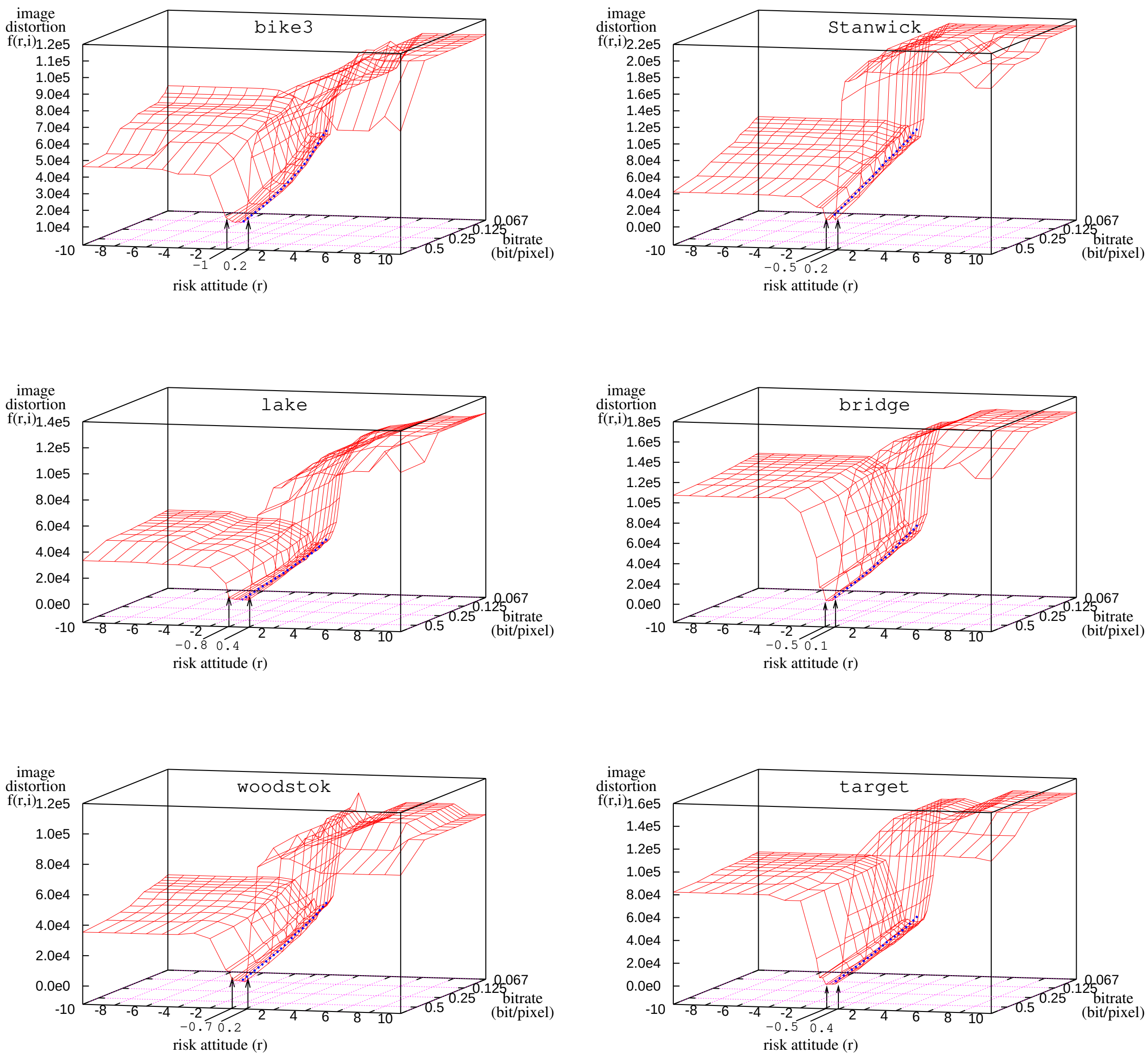


Fig. 5. 3-D plots on attitude-rate-distortion. Results for SPIHT are given by superimposing blue lines of dots.

Rational systems with risk aversion parameter $r$ around 0 and SPIHT seem to perform at about the same level for the test images, however, REWIC may be better able to capture the natural texture of the original. Subjective comparisons of decoded outcomes at different bit rates are illustrated in Figures 6–9. In a second set of experiments, we perform a more thorough comparison of $REWIC$ with a moderate risk-averse behavior and $SPIHT$, based on a novel coder selection procedure as follows.


Fig. 6. Visual comparison of $REWIC$ with a moderate risk-averse behavior and $SPIHT$.

### *A. Comparative Study*

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 bitrates $q(1), \cdots, q(K)$ using coder $\mathcal{L}_\alpha$; $\{I^\beta_{q(1)}, \cdots, I^\beta_{q(K)}\}$ be the set of decoded images at bitrates $q(1), \cdots, q(K)$ using coder $\mathcal{L}_\beta$.


Fig. 7. Visual comparison of $REWIC$ with a moderate risk-averse behavior and $SPIHT$.

Once distortion functions $f(\dagger, i)$ have been calculated (following equation (3)) 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 1: Coder Selection Procedure.** *In the language of statistical hypothesis*


Fig. 8. Visual comparison of $REWIC$ with a moderate risk-averse behavior and $SPIHT$.

*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 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.*


Fig. 9. Visual comparison of $REWIC$ with a moderate risk-averse behavior and $SPIHT$.

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

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 bit/pixel to 0.5 bit/pixel). It allows us to analyze the behavior of image coders at different levels of performance across bit rates.

Figures 10 and 11 show, for every test image in the dataset, 2D plots on rate-distortion as given by $f(\dagger, i)$ for SPIHT and REWIC$_r$ with $r$ set to $-0.1$ and $-0.5$. It is worthwhile to point out that, for thirteen test images of our dataset, there exists a risk aversion parameter $r$ (either $r = -0.1$ or $r = -0.5$) such that $f(\text{REWIC}_r, i) \leq f(\text{SPIHT}, i)$ from 0.0625 bpp to 0.5 bpp, and hence, the first condition of the coder selection procedure is true for thirteen out of eighteen test images of the dataset.

The *second step* of the coder selection procedure consists of measuring the overall difference between functions $f(\text{REWIC}_r, i)$ and $f(\text{SPIHT}, i)$ by subtracting the area beneath their graphs. This operation corresponds to a Kolmogorov-Smirnov (K-S) test. Table I illustrates the K-S test results when it is applied for coder selection: if K-S *Statistic* $>$ *Probability* then it follows that the corresponding data sets $\{f(\text{REWIC}_r, i) \mid i = 1, 2, \cdots, K\}$ and $\{f(\text{SPIHT}, i) \mid i = 1, 2, \cdots, K\}$ are from different distributions (i.e., the second condition of the coder selection procedure is true as well). Note that, the K-S tests are only computed for images such that pass the first condition in Definition 1 as given by $f(\text{REWIC}_r, i) \leq f(\text{SPIHT}, i)$ from 0.0625 bpp to 0.5 bpp. From Table I we have that the second condition of the coder selection criterion is true for thirteen out of thirteen test images that pass the first condition in Definition 1.

As summarized in Table 2, we have that, for thirteen out of eighteen test images in the data set, there exists a risk aversion parameter $r$ (either $-0.1$ or $-0.5$, which were the only two values tested in this second experiment) such that conditions (1) and (2) in Definition 1 are both true and hence the respective rational system REWIC$_r$ with moderate risk aversion parameter is significantly better than SPIHT, at least with 95% confidence level.

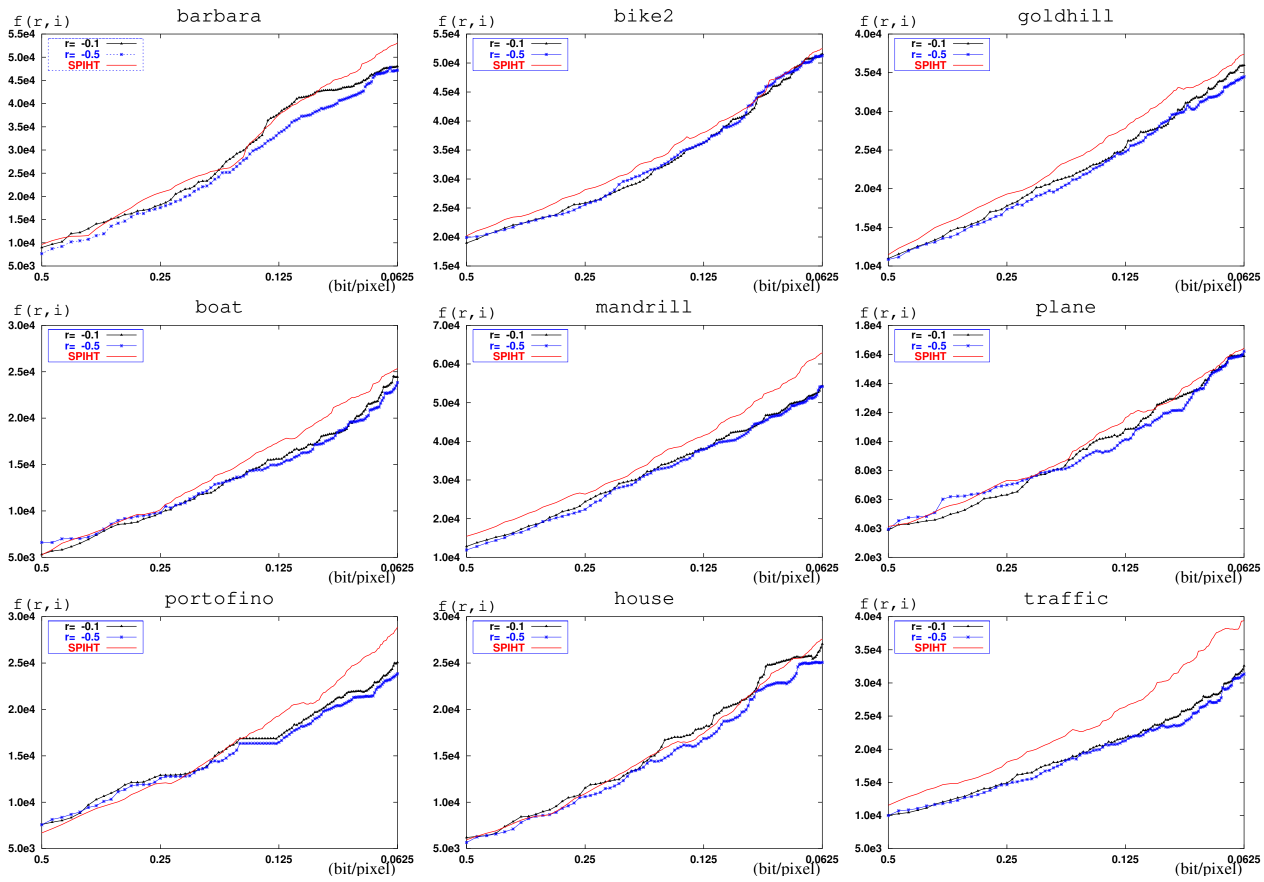


Fig. 10. 2D plots on rate-distortion for SPIHT and REWIC with $r$ set to $-0.1$ and $-0.5$.

## VI. CONCLUSION

Here we have proved that a rational system for progressive transmission might exhibit either a risk seeking posture or risk aversion with respect to “gambles” on variable-resolution compression. The rational embedded wavelet coder should choose, at any truncation time, among alternative spatial-orientation trees, in such a way as to avoid certain forms of behavioral inconsistency.

Using a dataset of standard test images, it was demonstrated that *in absence of a priori knowledge about regions of interest* a rational transmission system with moderate risk-aversion may be used with reasonable internal consistency and acceptable image fidelity. Based upon this same dataset, we found out that a rational system with moderate risk aversion is significantly better, on hypothesis testing basis, than SPIHT.

The validity and generalizability of the elicitation of risk attitudes for a rational coder as that presented here could be enhanced by exploring larger numbers of test images. While this empirical assessment should be regarded cautiously, on the basis of these preliminary findings it appears that the functional form used to represent utilities gives consistent results

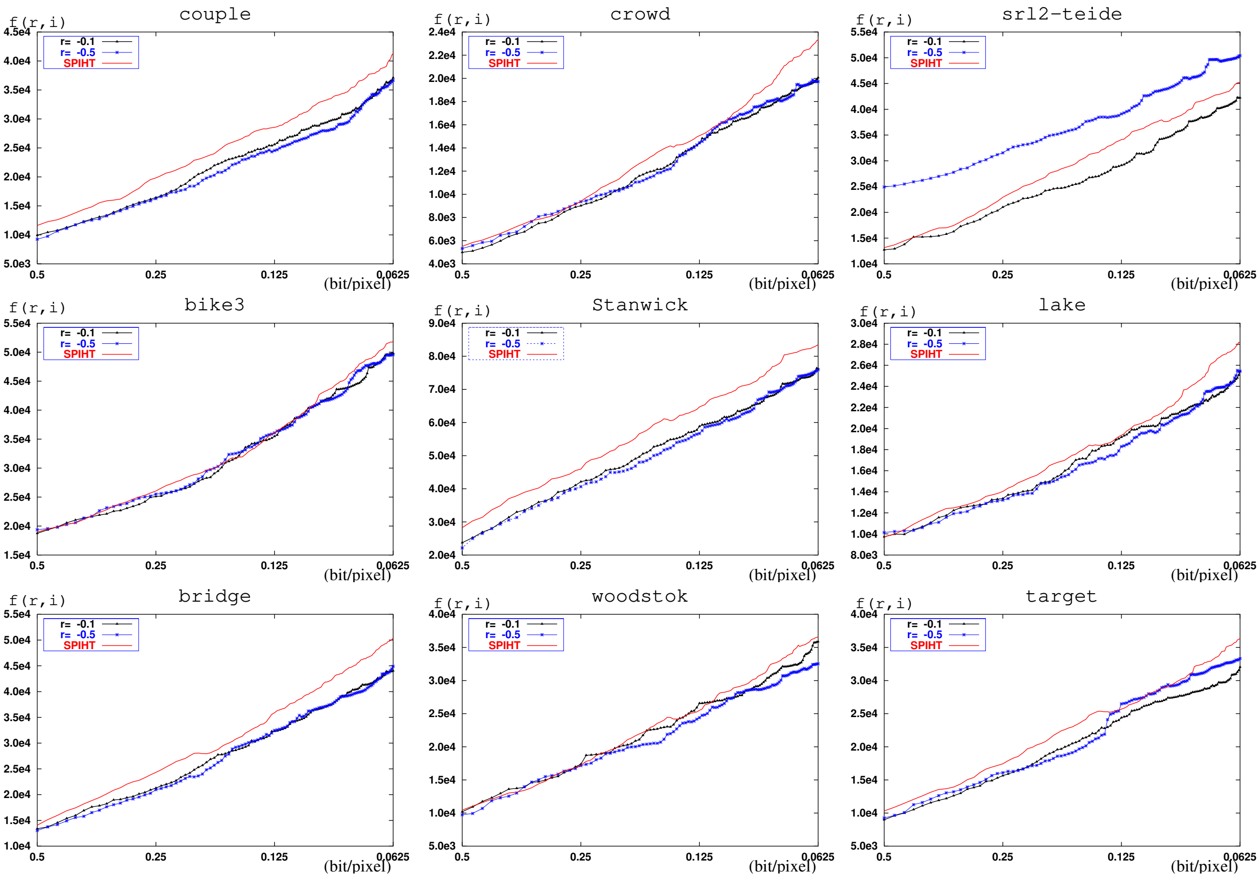


Fig. 11. 2D plots on rate-distortion for SPIHT and REWIC with $r$ set to $-0.1$ and $-0.5$.

and is suitable for further work.

**Acknowledgments.** This research was sponsored by the Spanish Board for Science and Technology (CICYT) under grant TIC2000-1421.

## References

[1] A. Said, and W.A. Pearlman, A new, fast and efficient image codec based on set partitioning in hierarchical trees, *IEEE Trans. on Circuit and System for Video Technology*, **6**(3), 243-250, (1996).

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

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

[4] R.L. Keeney and H. Raiffa, *Decisions with multiple objectives: preferences and value tradeoffs,* John Wiley and Sons, New York, (1976).

[5] J.S. Pliskin, D.S. Shepard, and M.C. Weinstein, Utility functions for life years and health status, *Operation Research,* **28**, 206-224, (1980).

[6] M.J. Machina, Choice under uncertainty: problems solved and unsolved, *Economic Perspectives*, **1**(1), 121-154, (1987).

| *Kolmogorov-Smirnov (K-S) Test Results* | | | | |
|---|---|---|---|---|
| | SPIHT/REWIC (r=-0.1) | | SPIHT/REWIC (r=-0.5) | |
| *IMAGE* | K-S Statistic | Probability | K-S Statistic | Probability |
| barbara | – | – | 0.221239 | 0.006488 |
| bike2 | – | – | – | – |
| goldhill | 0.168142 | 0.073017 | 0.256637 | 0.000895 |
| boat | 0.185841 | 0.035064 | – | – |
| mandrill | 0.292035 | 0.000092 | 0.300885 | 0.000050 |
| plane | – | – | – | – |
| portofino | – | – | – | – |
| house | – | – | 0.203540 | 0.015647 |
| traffic | 0.353982 | 0.000001 | 0.389381 | 0.000000 |
| couple | 0.230088 | 0.004065 | 0.274336 | 0.000298 |
| crowd | 0.247788 | 0.001510 | 0.265487 | 0.000521 |
| srl2-teide | 0.221239 | 0.006488 | – | – |
| bike3 | – | – | – | – |
| stanwick | 0.274336 | 0.000298 | 0.283186 | 0.000167 |
| lake | 0.212389 | 0.010169 | 0.230089 | 0.004065 |
| bridge | 0.283186 | 0.000167 | 0.265487 | 0.000521 |
| woodstok | – | – | – | – |
| target | 0.274336 | 0.000298 | – | – |

TABLE I


[7] J.A. Garcia, Rosa Rodriguez-Sanchez, J. Fdez-Valdivia, and, Xose R. Fdez-Vidal, A rational methodology for lossy compression, *Decsai Technical Report* **01-03-04**, 1-35, (2001).

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

[9] J.A. Garcia, J. Fdez-Valdivia, Xose R. Fdez-Vidal, and Rosa Rodriguez-Sanchez, Information Theoretic Measure for Visual Target Distinctness, *IEEE Trans. on Pattern Analysis and Machine Intelligence,* **23**(4), 362-383, (2001).

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

[11] A. Toet, Computing visual target distinctness, *TNO-report,* **TM-97-A039**, TNO Human Factors Research Institute, 74, (1997).

[12] X.R. Fdez-Vidal, J.A. Garcia, J. Fdez-Valdivia, Using models of feature perception in distortion measure guidance, *Pattern Recognition Letters*, **19**, 77-88, (1998).

[13] X.R. Fdez-Vidal, J.A. Garcia, J. Fdez-Valdivia, and R. Rodriguez-Sanchez, The role of integral features for perceiving image discriminability, *Pattern Recognition Letters*, **18**, 733-740, (1997).

[14] J. Martinez-Baena, J. Fdez-Valdivia, J.A. Garcia, and X.R. Fdez-Vidal, A new distortion measure based on a data-driven multisensor organization, *Pattern Recognition*, **31**(8), 1099-1116, (1998).

| *CODER SELECTION PROCEDURE WITH % CONFIDENCE* | | | | |
|---|---|---|---|---|
| IMAGE | Condition 1 (y/n ) | Condition 2 (-/y/n ) | $r$ | Confidence |
| barbara | $y$ | $y$ | $-0.5$ | 99% |
| bike2 | $n$ | – | – | – |
| goldhill | $y$ | $y$ | $-0.5$<br>$-0.1$ | 99%<br>90% |
| boat | $y$ | $y$ | $-0.1$ | 95% |
| mandrill | $y$ | $y$ | $-0.5$<br>$-0.1$ | 99%<br>99% |
| plane | $n$ | – | – | – |
| portofino | $n$ | – | – | – |
| house | $y$ | $y$ | $-0.5$ | 95% |
| traffic | $y$ | $y$ | $-0.5$<br>$-0.1$ | 99%<br>99% |
| couple | $y$ | $y$ | $-0.5$<br>$-0.1$ | 99%<br>99% |
| crowd | $y$ | $y$ | $-0.5$<br>$-0.1$ | 99%<br>99% |
| srl2-teide | $y$ | $y$ | $-0.1$ | 99% |
| bike3 | $n$ | – | – | – |
| stanwick | $y$ | $y$ | $-0.5$<br>$-0.1$ | 99%<br>99% |
| lake | $y$ | $y$ | $-0.5$<br>$-0.1$ | 99%<br>95% |
| bridge | $y$ | $y$ | $-0.5$<br>$-0.1$ | 99%<br>99% |
| woodstok | $n$ | – | – | – |
| target | $y$ | $y$ | $-0.1$ | 99% |

TABLE II

[15] X.R. Fdez-Vidal, A. Toet, J.A. Garcia, J. Fdez-Valdivia, Computing visual target distinctness through selective filtering, statistical features, and visual patterns, *Optical Engineering*, **39**(1), 267-281, (2000).

## Appendix

### A. Proof of Proposition 1

(a) Follows from Proposition 2.11 in [3] 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 [3].

*B. Proof of Proposition 2*

A well-behaved utility function is local and thus the value of distribution $\mathcal{P}$ is to be assessed in terms of the probability it assigned to the actual outcome. It leads to simplifications that make it a useful working hypothesis:

$$u(\mathcal{P}_i, G_{l,i}) = u(p_{l,i})$$

where $p_{l,i} = p(G_{l,i} \mid S(R_i, t))$, with the actual greylevel occurrence being $G_{l,i}$.

Let $u_{R_1,R_2,\cdots,R_n}(p_{l,1}, p_{l,2}, \cdots, p_{l,n})$ represents the utility function over SOTs $R_1, R_2, ..., R_n$. Keeney and Raiffa[4] have demonstrated that the hypothesis of value independence given in Postulate 6 holds if and only if, for all $R_i$, there are utility functions $u_{R_i}(p_{l,i})$ and constants $a_i$ such that:

$$u_{R_1,R_2,\cdots,R_n}(p_{l,1}, p_{l,2}, \cdots, p_{l,n}) = \sum_i a_i \times u_{R_i}(p_{l,i}) \,, \tag{4}$$

where $\sum_i a_i = 1$. That is $u_{R_1,R_2,\cdots,R_n}(p_{l,1}, p_{l,2}, \cdots, p_{l,n})$ has an additive form.

Recalling that $S(R_j, t)$ denotes a bit stream, over remaining information in $R_j$, candidate to be delivered to the decoder, let $S^w(R_j, t)$ and $S^b(R_j, t)$ be the worst and best possible bit streams, respectively, that might be delivered over $R_j$ in the decision problem at hand. For example, $S^w(R_j, t)$ might correspond to some insignificant bit planes over remaining information in $R_j$ to be deliver to the decoder and $S^b(R_j, t)$ might correspond to the remaining bit planes over $R_j$ to achieve an excellent reconstruction. Without loss of generality we scale utility function $u_{R_j}(p_{l,j})$ so that $u_{R_j}(p^w_{l,j}) = 0$ with $p^w_{l,j} = p(G_{l,j} \mid S^w(R_j, t))$ and $u_{R_j}(p^b_{l,j}) = 1$ with $p^b_{l,j} = p(G_{l,j} \mid S^b(R_j, t))$. Then Postulate 7 states that

$$u_{R_1,R_2,\cdots,R_n}(p_{l,1}, p_{l,2}, \cdots, p_{l,i}, \cdots, p^w_{l,j}, \cdots, p_{l,n}) = u_{R_1,R_2,\cdots,R_n}(p_{l,1}, p_{l,2}, \cdots, q \times p_{l,i}, \cdots, p^b_{l,j}, \cdots, p_{l,n}) \,, \tag{5}$$

for some $q$, with $0 < q < 1$, and for all $p_{l,i}$. That is, if we assume Postulate 7, to achieve a decoded output corresponding to $R_i$ characterized by $p_{l,i}$ and another output to $R_j$ characterized by $p^w_{l,j}$ is equivalent to achieve an output to $R_i$ characterized by $q \times p_{l,i}$ and another to $R_j$ that is characterized by $p^b_{l,j}$, where $1 - q$ is the proportion of $p_{l,i}$ corresponding to $R_i$ that would be given up for an improvement in reconstruction fidelity corresponding to $R_j$ from $p^w_{l,j}$ to $p^b_{l,j}$.

Substituting equation (4) in equation (5) and recalling that $u_{R_j}(p^w_{l,j}) = 0$ and $u_{R_j}(p^b_{l,j}) = 1$ we find that

$$a_i \; u_{R_i}(p_{l,i}) = a_i \; u_{R_i}(q \; p_{l,i}) + a_j \,, \tag{6}$$

for all $p_{l,i}$.

Following Pliskin et al.[5], we first show that the three functional forms are consistent with Postulate 7. If $u(p) = \log(p)$ then $a_i = \frac{1}{1-\log q}$ and $a_j = \frac{-\log q}{1-\log q}$ are consistent with equation (6). Similarly, if $u(p) = p^r$ or $u(p) = -p^r$ then the values of $a_i = q^{-r}$ and $a_j = 0$ are consistent with equation (6). Summarizing we have already proved that $\log(p)$, $p^r$ for $r > 0$, and $-p^r$ for $r < 0$, are consistent with Postulate 7 when two particular probabilities $p^w_{l,j}$ to $p^b_{l,j}$ are involved in the tradeoffs. By a corollary given in [5] we can easily prove that this is sufficient for Postulate 7 to hold for any pair of probabilities.

We next prove that these three functions are the only ones consistent with Postulate 7 (to this aim we follow a proof suggested in another context by Keeney and Raiffa[4]). We twice differentiate equation (6) and divide the second derivative of each side by the first derivative which gives (let us now suppress the subscripts):

$$\frac{u''(p)}{u'(p)} = q\frac{u''(qp)}{u'(qp)} \,. \tag{7}$$

Recursively substituting $qp$ for $p$ in equation (7) it follows that:

$$\frac{u''(p)}{u'(p)} = q^n\frac{u''(q^n p)}{u'(q^n p)} \,, \tag{8}$$

for all $p$ and all $n \geq 0$. Given that $u$ is a "well-behaved" function, it implies the existence of $\lim_{p\to 0} p\frac{u''(p)}{u'(p)}$. Then for any $p_1$ we have that

$$\begin{aligned} \lim_{p\to 0} p\frac{u''(p)}{u'(p)} &= \lim_{n\to\infty} q^n p_1 \frac{u''(q^n p_1)}{u'(q^n p_1)} \\ &\overset{(a)}{=} p_1\frac{u''(p_1)}{u'(p_1)} \,, \end{aligned} \tag{9}$$

where (a) follows from equation (8). And similarly for any $p_2$ we have that

$$\begin{aligned} \lim_{p\to 0} p\frac{u''(p)}{u'(p)} &= \lim_{n\to\infty} q^n p_2 \frac{u''(q^n p_2)}{u'(q^n p_2)} \\ &= p_2\frac{u''(p_2)}{u'(p_2)} \,. \end{aligned} \tag{10}$$

From equations (9) and (10), it follows that for any $p_1$ and $p_2$:

$$p_1\frac{u''(p_1)}{u'(p_1)} = p_2\frac{u''(p_2)}{u'(p_2)} \,, \tag{11}$$

or, equivalently,

$$-p\frac{u''(p)}{u'(p)} = c \,, \tag{12}$$

By integration we obtain that the utility function $u(p)$ must have one of the three functional forms $\log(p)$ for $r = 0$, $p^r$ for $r > 0$, and $-p^r$ for $r < 0$, where $r = 1 - c$, and $c$ is the constant in equation (12).

Following Machina[6], we know that the shape of $u(p)$ determines risk attitudes. We have that, for $r < 1$, the utility is a concave function: (i) $p^r$ for $0 < r < 1$; (ii) $\log(p)$ for $r = 0$; or (iii) $-p^r$ for $r < 0$. Since a system with a concave utility function will in fact always prefer receiving a sure gain (of the expected value of the "gamble" on SOT-dependent quality of encoding) to the "gamble" itself, concave utility functions are termed risk averse. For $r > 1$ the utility function must have the form $p^r$ which is a convex function. It implies that the resulting system prefers bearing the risk rather than receiving the sure gain of the expected value of the "gamble" on SOT-dependent quality of encoding. Hence, such utility function is termed risk loving.

This proves Proposition 2.

*C. Proof of Proposition 3*

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

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

which, by Theorem 1 in [9], is non-negative and verifies the nilpotence condition. That is, the expected increase in utility in this case 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 [9] for further details).

Similarly it can be proved that if preferences are described by utility function $u_1$ 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 given by:

$$I(S(R_i,t)/\mathcal{Q}) = \sum_l p(G_{l,i} \mid S(R_i,t)) \left[ (p(G_{l,i} \mid t-1))^r - (p(G_{l,i} \mid S(R_i,t))^r \right] , \quad \text{for } r < 0 \tag{13}$$

while if preferences are described by utility function $u_3$ then the expected increase in utility is given by:

$$I(S(R_i,t)/\mathcal{Q}) = \sum_l p(G_{l,i} \mid S(R_i,t))\left[(p(G_{l,i} \mid S(R_i,t))^r - (p(G_{l,i} \mid t-1))^r\right] \ , \quad \text{for } r > 0. \tag{14}$$

This proves Proposition 3.