Spherical Coding Algorithm for Wavelet Image Compression

Hasan F. Ates, *Member, IEEE,* and Michael T. Orchard, *Fellow, IEEE*

***Abstract*—In recent literature, there exist many high-performance wavelet coders that use different spatially adaptive coding techniques in order to exploit the spatial energy compaction property of the wavelet transform. Two crucial issues in adaptive methods are the level of flexibility and the coding efficiency achieved while modeling different image regions and allocating bitrate within the wavelet subbands. In this paper, we introduce the "spherical coder", which provides a new adaptive framework for handling these issues in a simple and effective manner. The coder uses local energy as a direct measure to differentiate between parts of the wavelet subband and to decide how to allocate the available bitrate. As local energy becomes available at finer resolutions, i.e. in smaller size windows, the coder automatically updates its decisions about how to spend the bitrate. We use a hierarchical set of variables to specify and code the local energy up to the highest resolution, i.e. the energy of individual wavelet coefficients. The overall scheme is nonredundant, meaning that the subband information is conveyed using this equivalent set of variables without the need for any side parameters. Despite its simplicity, the algorithm produces PSNR results that are competitive with the state-of-art coders in literature.**

***Index Terms*—image, wavelet, coding.**

I. Introduction

All image coders are based on some statistical model for natural images, and exploit the dependencies implied by that model. The coder is explicitly or implicitly optimized for the specific model and applied to sample images. Therefore, coding efficiency depends on how well the source model matches the true distribution of natural images. In other words, without a realistic source model to begin with, it is not possible to construct an efficient compression algorithm.

Building a good source model requires a convenient and efficient representation of the data. The success of transform domain techniques have proven that coders based on DCT or wavelet representations are superior to pixel domain methods. Wavelet domain is shown to provide a good match to the space-frequency characteristics of natural images. Hence, it is much easier to build a realistic image model in the wavelet domain than, say, in the pixel domain. That's why a simple coder in the wavelet domain could outperform a complex coder in the pixel domain. Within the last decade, wavelets have exemplified how a good image representation opens the doors to a variety of original and successful image models.

This research was supported in part by the NSF Grant DMS9872890 and by Isik University BAP-05B302 Grant.

H. F. Ates is with the Department of Electronics Engineering, Isik University, Sile, Istanbul, TURKEY (e-mail: hfates@isikun.edu.tr).

M. T. Orchard is with the Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005 USA (e-mail: orchard@ece.rice.edu).

In this paper, we further develop and analyze the "spherical representation" that has been introduced in [1] as a novel way of representing image information in wavelet domain. We first discuss the essential properties of a good wavelet-based image model and indicate the weaknesses of the existing models. Then, we show why spherical representation provides a robust framework for building efficient image coders. We suggest that wavelet subbands are best characterized by spatially varying non-homogeneous processes. Based on the spherical representation, we develop a coding algorithm that handles this non-homogeneity in an effective and non-parametric way.

Understanding the reasons behind the success of wavelet coders is important for predicting the future directions of image coding. Wavelet transform achieves energy compaction into few low-pass coefficients plus a sparse set of clustered high-pass coefficients. Such a compact representation is very suitable for a simple yet effective source model. The history of wavelet coders shows an evolution of the models employed for exploiting the energy clustering in the wavelet domain. At the beginning stages, image information was assumed to be naturally classified into statistically independent wavelet subbands, each of which was modeled as an independent identically distributed (i.i.d.) process. Successful coding schemes used in DCT-based algorithms, such as run-length coding, and vector quantization [2], [3] were tried in wavelet image coding, but demonstrated modest coding gains over standard transform-based algorithms. The breakthrough in wavelet image coding arrived with coders using hierarchical wavelet trees, such as EZW [4], SPIHT [5], SFQ [6], [7], [8], [9]. Grouping wavelet coefficients that belong to the same spatial region under a tree-structure, these coders were able to adapt to the properties of different regions in the image. Other coders, such as EQ [10], classification based algorithms [11], EBCOT [12], etc., achieved improved coding efficiency by introducing more spatial adaptivity in modeling the subbands.

The success of adaptive models is a direct consequence of the special characteristics of image information. Natural images consist of large smooth areas with localized high frequency structures (i.e. edges) separating them. Edges and texture come in arbitrary locations, orientations, shapes, and sizes in natural images. Since high-frequency information is rather localized, even coarse level information about the location of high activity areas allows the coding methods to be successfully adapted to the statistics of different regions. In other words, using such "location information", wavelet subbands are modeled as non-homogeneous processes and coded accordingly.

Recognizing the spatially changing properties of wavelet subbands is crucial for accurate modeling. Equally important is

the optimal allocation of bitrate to different parts of a subband having distinct statistical characteristics. Sophisticated adaptive techniques fine tune models for each coefficient based on the context of its local (scale and/or spatial) neighborhood. A good example is the EQ coder [10], which uses a generalized Gaussian distribution (GGD) with spatially adapted variance for modeling each subband; the variance at each point is estimated from the decoded values in its causal neighborhood unless all the neighborhood coefficients are quantized to zero. Based on the estimated variances, the coefficients are coded in a way that yields overall rate-distortion optimality.

Despite their success, the EQ coder and other adaptive methods could only offer a restricted view of image information in the wavelet domain. For instance, zerotrees of EZW coder [4] are able to provide a rather structured separation between significant and insignificant sets of coefficients. The EQ coder is more flexible; however, because of the way the variances are estimated, it assumes a slowly changing variance field for the wavelet subband. It is doubtful whether this level of adaptivity is adequate to accurately model the rich variety of local statistics of wavelet coefficients. A modeling mismatch for each coefficient will contribute to the loss of coding efficiency for the overall image. We claim that parametric descriptions of wavelet coefficient distributions are especially prone to mismatches. In other words, the wavelet image model shouldn't be tied down to a fixed parametric description. A more adaptive coding approach should be developed, which updates its modeling paradigm locally as more information becomes available about the underlying wavelet coefficients.

In this paper, we develop a wavelet-based representation that is general, flexible and realistic. The "spherical representation" is a hierarchical description of how total coefficient energy gets distributed within each wavelet subband. A hierarchical tree of subband energy is formed by summing up the squared coefficients. Phase variables are defined that describe how the energy in a given region is split into energies of two subregions. Phase variables are coded based on a simple and effective model. The non-homogeneity of wavelet subbands is handled through this non-parametric model of the hierarchy. We discuss why the spherical coding framework is more robust against modeling mismatches than typical parametric techniques. In particular, we explain how our coder improves the coding efficiency by allocating total bitrate according to the local sum of energies within the subband. The local energy is used to adapt the coder to the local statistics of wavelet coefficients. We claim that this approach makes it possible to build highly adaptive and flexible coding algorithms.

Section II defines what modeling mismatch is and shows its detrimental effects on the coder performance using a simple example. Section III motivates and explains the spherical representation, and discusses why this representation is more robust against modeling mismatch while coding the wavelet subbands. Then, Section IV describes the details of the spherical coding algorithm. In Section V, the algorithm is tested on standard test images. Compared to some of the state-of-art wavelet coders, the spherical coding algorithm provides better or as good coding performance.

II. Effects of Modeling Mismatch in Coding

Mismatch in source coding indicates the loss of coding efficiency resulting when a coder optimized for a certain source model is applied to a different model. This is an important problem in image coding, since there is no single source model that can successfully describe a variety of different image characteristics. Edges, texture, smooth regions require different type of characterizations. It is not easy to determine the exact statistical nature of each such region. Even if we assume that we could develop correct models for each and every pixel or wavelet coefficient of the image, we will probably need a large set of parameters to define these distributions and this incurs a heavy cost as side information for the coder. On the other hand, if the parametrization is restricted in some way, as it is done in all wavelet coders, modeling mismatch seems inevitable.

We provide a simple example to show quantitatively the effects of mismatch. In lossless coding, the performance loss due to mismatch is measured by the relative entropy between the two distributions, i.e. the distribution for which the coder is designed and the distribution to which the coder is applied. For lossy coding, results from high-rate vector quantization theory can be used to show that relative entropy between two continuous distributions is a good representative of the mismatch [13]. Suppose that we apply the optimal coder designed for an i.i.d. zero-mean Gaussian process to an independent non-homogeneous zero-mean Gaussian process with changing variances. The relative entropy is defined as

$$I(f||g) = \int d\mathbf{x} f(\mathbf{x}) \ln \frac{f(\mathbf{x})}{g(\mathbf{x})}. \tag{1}$$

Using

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{2^{k-1}} \prod_{i=1}^{2^k} \sigma_i} \exp\left(-\frac{1}{2}\sum_{i=1}^{2^k} \frac{x_i^2}{\sigma_i^2}\right) \tag{2}$$

and

$$g(\mathbf{x}) = \frac{1}{(2\pi)^{2^{k-1}} \sigma^{2^k}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{2^k} x_i^2\right), \tag{3}$$

we get

$$\begin{aligned} I(f||g) =& \int_{-\infty}^{\infty} d\mathbf{x} f(\mathbf{x}) \sum_{i=1}^{2^k} \left(\ln(\frac{\sigma}{\sigma_i}) + \frac{1}{2}\ln(e)(-\frac{1}{\sigma_i^2} + \frac{1}{\sigma^2})x_i^2\right) \\ =& \sum_{i=1}^{2^k} \left(\ln(\frac{\sigma}{\sigma_i}) + \frac{1}{2}\ln(e)(\frac{\sigma_i^2}{\sigma^2} - 1)\right). \end{aligned} \tag{4}$$

Each term in this sum is greater than or equal to zero, with equality being when $\sigma_i = \sigma$. Hence, the coder loses some efficiency over all coefficients, unless the variance estimate matches the true variance. In Figure 1, the relative entropy is plotted against $\ln(\sigma/\sigma_1)$ for $k = 1$.

This example illustrates the importance of accurate parameter estimation for coding a non-homogeneous process. The estimation errors accumulate to reduce the efficiency of the overall coding scheme. In wavelet subbands, this could be a major problem, since the statistics change rapidly from one

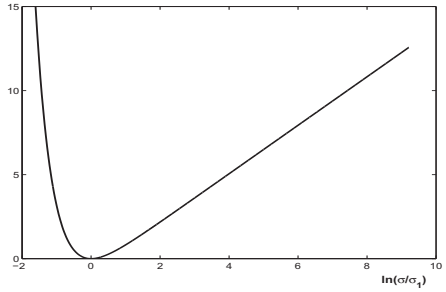


Fig. 1: Relative entropy vs. $\ln(\sigma/\sigma_1)$.

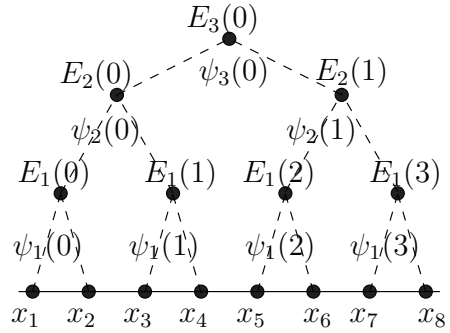


Fig. 2: Spherical representation in 1-D.

region to the next and there aren't enough samples to perform reliable estimation. The EQ coder tries to reduce mismatch by performing local variance estimation. However, it is not even clear whether the immediate neighbors of a coefficient are reliable enough to estimate its variance. This problem persists in general for all parametric coding algorithms, i.e. for algorithms that try to estimate certain parameters of the coefficient distributions, and code them accordingly. In the next section, we introduce our representation, and discuss how robust non-parametric modeling can be carried out in this framework, which we claim has the potential to significantly reduce the damaging effects of modeling mismatch.

III. The Spherical Representation

The clustering of energy in wavelet subbands motivates the use of spatially varying models in coding the wavelet information. All adaptive wavelet coders introduce some form of non-homogeneity, either explicitly by parameterizing the distribution of each wavelet coefficient (e.g. the EQ coder and classification-based coders), or implicitly by using different quantization and coding techniques in different parts of the subband (e.g. zerotree coding). In either case, care must be taken not to produce excessive side information in the form of model parameters or a classification map. This limitation compromises the freedom and the flexibility in choosing a matching model for the non-homogeneous nature of the wavelet subband. As discussed in Section II, model mismatches could result in severe performance loss.

Natural images offer many complications in modeling the existing non-homogeneity. Different image regions require different characterizations for efficient coding. There doesn't seem to be a small number of models one can easily define and use to capture the statistical variety observed in natural images. Due to the rich structure of edges and texture, statistical differences need to be recognized within windows of different shapes and sizes, ranging from large chunks of coefficients in smooth regions down to the level of single isolated locations. It is because of these challenges that we've decided to look for flexible representations that can deal with such varying information content.

Our representation and coding method share a similar philosophy with the EQ coder in its use of local energy, equivalently local variance, as a reliable measure of local information content. We suggest that local variance provides sufficient information about how the wavelet coefficients could be efficiently coded. Out of all wavelet coders in literature, we single out the EQ coder for its effort to offer an "infinite mixture model" for the wavelet coefficients. That is, each coefficient can possess a GG distribution with a different variance of any positive value. The problem in EQ is the obligation to use the causal neighborhood for variance estimation in order to avoid side information. In cases when local variances exhibit sudden changes, e.g. around high-frequency structures such as edges, the estimated variances are not accurate and this leads to model mismatch.

One way to overcome the problem in EQ is to represent local energy as part of the wavelet information to be coded, and not as extra parameters needed for modeling. In other words, local energy should be implicitly coded as part of the subband information content. If both encoder and decoder have access to local energy information, then coding could be adapted according to this local statistic without any need for side information. With that perspective, it is convenient to define local energy hierarchically, starting from the total energy of the full subband going down to smaller regions, even down to the energy of a single coefficient. As the size of the region is reduced, the local energy provides a better estimate of the variance of the coefficients in that region. Given the energy in a certain region, the encoder only needs to code how this energy is divided into its sub-regions. This coarse-to-fine strategy refines successively the available local information, and makes coding adaptation more successful.

Motivated by this reasoning, we propose to use the following hierarchical structure to represent a random process $\mathbf{X}$ (see Figure 2): In 1-D (one dimension) , for $\mathbf{X} = \{x_i\}_{1 \le i \le 2^k}$, and for $0 < m \le k$, $0 \le n < 2^{k-m}$,

$$E_m(n) = \sum_{i=2^m n+1}^{2^m(n+1)} x_i^2, \tag{5}$$

$$\psi_m(n) = \arctan\left(\sqrt{\frac{E_{m-1}(2n)}{E_{m-1}(2n+1)}}\right), \tag{6}$$

where $0 \le \psi_m(n) \le \pi/2$. Here, $\mathbf{X}$ could be seen as one of the wavelet subbands of a 1-D signal. In the next section, this formulation is easily extended to 2-D (two dimensions) subbands.

The variables $E_m(n)$ provide local energy information at different resolution levels m. The phase variables $\psi_m(n)$ indicate how the local energy gets split between the two

neighboring regions:

$$E_{m-1}(2n) = E_m(n)\sin(\psi_m(n))^2, \quad (7)$$
$$E_{m-1}(2n+1) = E_m(n)\cos(\psi_m(n))^2. \quad (8)$$

The phase variables in a sense represent the difference in information content between the two regions. Going from the top level ($m = k$) to the bottom level ($m = 1$) of the hierarchy, the phase values provide a refinement of the available information in each region of the subband.

When the total energy, $E_k(0) = \sum_{i=1}^{2^k} x_i^2$, and the $2^k - 1$ phases at all levels of the hierarchy are given, the coefficients are easily determined up to a sign bit; i.e. $x_i = \text{sign}(x_i)\sqrt{E_0(i)}$. The sign bits could also be defined as part of the representation if the phase values at the bottom of the hierarchy cover full range; i.e. $-\pi < \psi_1(i) \leq \pi$, and:

$$x_{2i} = \sqrt{E_1(i)}\sin(\psi_1(i)), \quad (9)$$
$$x_{2i+1} = \sqrt{E_1(i)}\cos(\psi_1(i)). \quad (10)$$

In this type of representation, we are able to use local energy not only to differentiate between statistically distinct parts of the process but also to provide direct information about the underlying coefficient values. Coding $E_k(0)$ and $\psi_m(n)$ can be seen as an alternative to coding x_i. We might say that, instead of cartesian coordinates, spherical coordinate system is used in representing the process; hence the name spherical representation.

The simple example of Section II helps us understand better the convenience of spherical representation for coding a non-homogeneous process. In case when the process $\mathbf{X}$ is i.i.d. zero mean Gaussian with variance σ^2, the local energies $E_m(n)$ are $\sigma^2\chi^2_{2^m}$, where $\chi^2_{2^m}$ is chi-square distributed with 2^m degrees of freedom. The ratio of two chi-square distributed random variables has F-distribution, $\text{F}_{2^m,2^m}$; the distribution of the phase variables could be computed accordingly. It can be shown that the joint pdf's satisfy:

$$p_{\chi^2_{2^m},\text{F}_{2^{m-1},2^{m-1}}}(x,y) = p_{\chi^2_{2^m}}(x)p_{\text{F}_{2^{m-1},2^{m-1}}}(y) \quad (11)$$

Therefore, $E_m(n)$ and $\psi_m(n)$ are mutually independent; it follows that $\psi_m(n)$ are independent random variables for all $0 < m < k$, $0 \leq n < 2^{k-m}$. In theory, we can design an optimal coder for these variables, and this coder is going to have a performance equal to that of the coder designed for i.i.d. Gaussian $\mathbf{X} = \{x_i\}_{1\leq i\leq 2^k}$.

If $\mathbf{X}$ is in fact non-homogeneous Gaussian with changing variances σ_i^2 for x_i, then the total loss of efficiency due to using the optimal spherical coder designed for the i.i.d. case has to be equal to the mismatch calculated in Section II. However, unlike the previous case, the phase variables $\psi_m(n)$ contribute in different proportions to the total mismatch; the phases at the top levels of the hierarchy cause more mismatch than then ones at the lower levels.

To show this, let's first look at the distribution of $E_m(0) = \sum_{i=1}^{2^m} x_i^2$, where x_i have variances σ_i^2 with mean variance $\bar{\sigma}^2 = (1/2^m)\sum_{i=1}^{2^m}\sigma_i^2$. Figure 3 plots the pdf of $E_m(0)$ for $m = 4$, $\bar{\sigma} = 1$ and $\text{var}(\sigma_i^2) = 0.1$. The dashed curve shows the pdf of χ^2_{16}. It turns out that, as long as σ_i^2's do not deviate too much from the mean value $\bar{\sigma}^2$, $E_m(0)$ is approximately distributed as $\bar{\sigma}^2\chi^2_{2^m}$.

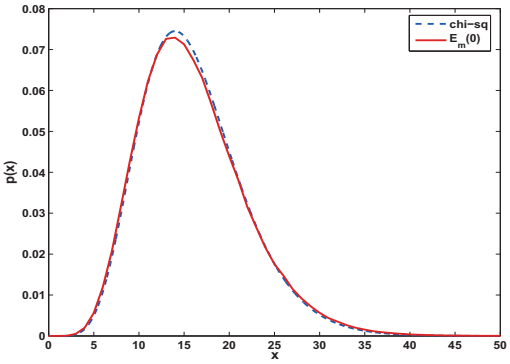


Fig. 3: Pdf for $E_m(0)$ and χ^2_{16} ($m = 4$ and $\bar{\sigma} = 1$).

Therefore, for slowly changing variances, we can assume that $E_m(n)$ is $\sigma_1^2\chi^2_{2^m}$, and $E_m(n+1)$ is $\sigma_2^2\chi^2_{2^m}$, where σ_1^2 and σ_2^2 are the mean variances for the corresponding coefficients. Then $E_m(n)/E_m(n+1)$ is $(\sigma_1^2/\sigma_2^2)\text{F}_{2^m,2^m}$, where the pdf of $\text{F}_{2^m,2^m}$ is given by,

$$g(x) = \Omega\frac{x^{(2^m-2)/2}}{(1+x)^{2^m}}, \quad x > 0, \quad (12)$$

and the pdf of $E_m(n)/E_m(n+1)$ is ($s = \frac{\sigma_1^2}{\sigma_2^2}$),

$$f(x) = \Omega\frac{s^{2^{m-1}}x^{(2^m-2)/2}}{(s+x)^{2^m}}, \quad x > 0, \quad (13)$$

where Ω is an appropriate normalization factor. The relative entropy between $(\sigma_1^2/\sigma_2^2)\text{F}_{2^m,2^m}$ and $\text{F}_{2^m,2^m}$ is given by:

$$I(f||g) = \int_0^\infty dx f(x)\ln\left(\frac{s^{2^{m-1}}(1+x)^{2^m}}{(s+x)^{2^m}}\right)$$
$$= 2^m\int_0^\infty dx f(x)\left(\ln(s^{1/2}) + \ln(\frac{1+x}{s+x})\right). \quad (14)$$

The relative entropy is proportional to 2^m, hence increases exponentially at higher levels of the hierarchy. Even though the relative entropy for $E_m(n)/E_m(n+1)$ is not an exact measure of coding mismatch for the phase variables, we can argue that coding mismatch has to be significantly higher at the top levels of the hierarchy. Since the total mismatch should be equal to the case when x_i's are coded as i.i.d. Gaussian, the coding mismatch has to be relatively small for the phase variables at the lowest level, i.e. $m = 1$ (when compared with the expression in Eqn. 4).

Since the upper levels contribute a major portion of the efficiency loss due to mismatch, improving the coding performance at the upper levels significantly improves the coding efficiency of the overall coding scheme. This makes the spherical representation robust against the non-homogeneity of the underlying process. In other words, without knowing the exact nature of this non-homogeneity and without any detailed parametrization, we only need reasonable models for upper-level phase variables in order to have good overall coding results. From a different viewpoint, if the upper levels are

optimally coded, then the lower levels will have access to optimally decoded local energies, and the coding at the lower levels will benefit from this information.

The attractiveness of the spherical representation is not limited by its ability to collect modeling mismatch at the few upper level phase values. It also creates a highly adaptive coding framework, where different coding techniques could be developed at different levels of the hierarchy without requiring any side information. Imagine the optimal codeword for $\mathbf{X}$ is given by $\hat{\mathbf{X}} = \{\hat{x}_i\}_{1 \leq i \leq 2^k}$. The decoded phase variables $\hat{\psi}_{m'}(n)$ at a certain level m' affect the decoded values of lower level energies $\hat{E}_m(n), m < m'$, and eventually the decoded coefficients $\hat{x}_i$. On the other hand, optimal coding of $\psi_{m'}(n)$ is directly related to the total decoded energy $\hat{E}_{m'}(n)$ and therefore on $\hat{x}_i$. This mutual dependency creates difficulties for rate-distortion optimization but also opens the door to endless possibilities for innovative coding strategies.

We resort to our original example to explain in simple terms how to perform model adaptation using the spherical representation. Imagine that the variances $\sigma_i^2, 1 \leq i \leq 2^k$ of non-homogeneous $\mathbf{X}$ are mutually independent. Then the maximum likelihood (ML) estimate for each σ_i^2 is equal to x_i^2. That is, the values of neighboring coefficients are useless in estimating the variance of x_i. Without any *a priori* information, the optimal spherical coder based on the i.i.d. assumption will use $\sigma^2 = \hat{E}_k(0)/2^k$ as the variance estimate of all coefficients. At a certain level m', assume that we have the decoded energy $\hat{E}_{m'}(n)$. Then, the descendants of this node (e.g. $\psi_{m'}(n)$ and other phase values that belong to the subtree below $\hat{E}_{m'}(n)$) can be coded using the optimal spherical coder for an i.i.d. Gaussian with variance $\hat{E}_{m'}(n)/2^{m'}$. In other words, each subtree will be coded based on its decoded energy and independent of the rest of the spherical tree. Ignoring the quantization effects, this new variance estimate provides a better match than $\hat{E}_k(0)/2^k$ for the coefficients at the leaves of this subtree. This means a reduction in the modeling mismatch for this subtree. As we go down the hierarchical tree, the variance estimate for every subtree gets refined at each level, and we could get a significant recovery from the performance loss due to the modeling mismatch of using $\hat{E}_k(0)/2^k$ as the variance estimate. This example illustrates how different levels of the spherical hierarchy provides a natural refinement of available information and how this new information could be used for successful model and coder adaptation.

In the next section, we extend the spherical representation to 2-D wavelet subbands and develop a simple coding algorithm that exploits the flexibility and robustness of the spherical framework for efficient coding of the wavelet information.

IV. Spherical Coder in Wavelet Subbands

Spherical representation could be easily extended to 2-D to be used in wavelet image coding. Partial squared sums need to be defined in both vertical and horizontal directions in an alternating fashion (see Figure 4). Let us represent the phase variables and the local energies as $\psi_{u,v}(s,t)$ and $\Gamma_{u,v}(s,t)$ respectively. Assume the subband is $(2^J \times 2^J)$, so that $0 \leq u, v \leq J$, and s, t are defined accordingly. The spherical coder described in this section codes the wavelet subband through $2^{2J} - 1$ phase variables plus the total energy and the sign bits.

In coding $\psi_{u,v}(s,t)$, the spherical coder acts on the following assumptions:

- The technique is applied independently at each subband. Even though we believe the energy information to be highly redundant across scales, it is a challenging problem to model the dependencies among phase variables in different subbands. We discuss this and other issues at the end of this paper.
- $\psi_{u,v}(s,t)$ in each subband are assumed to be independent random variables that are uniformly distributed in $[0, \pi/2]$. Independence assumption simplifies the coding procedure. Once again, modeling and coding the intricate dependencies of $\psi_{u,v}(s,t)$ is a challenging and open problem. The use of true histograms (see Figure 6) in entropy coding $\psi_{u,v}(s,t)$ achieve very little coding gain, which justifies the use of uniform distribution.
- Independence of phase variables at different levels of the hierarchy implies that $\psi_{u,v}(s,t)$ is also independent of $\Gamma_{u,v}(s,t)$, since $\Gamma_{u,v}(s,t)$ is determined by $\psi_{u+1,v}(s,t)$ or $\psi_{u,v+1}(s,t)$ (see the definitions below).
- Since distortion is measured with respect to the actual decoded values of wavelet coefficients, rate-distortion theory implies that optimal coding of phase depends on the decoded values of corresponding local energy. Specifically, assuming independence, optimal coding requires rate-distortion curve of each $\psi_{u,v}(s,t)$ to be normalized by $\hat{\Gamma}_{u,v}(s,t)$ (See Figure 5).
- Since the decoded value of the local energy is needed for coding the phase, decoding is performed hierarchically, starting from the top level of the spherical tree going down to the coefficient level (See Figure 4).

Given these modeling assumptions, the job of the encoder is to choose the optimal (in rate-distortion sense) codeword which is admissible within the spherical coding framework. A codeword is admissible if its spherical tree can be decoded with zero distortion at the designated bitrate. More specifically, for such a codeword, decoded phase and local energy coefficients should be exactly equal to their original values. At a given bitrate, the set of all admissible codewords defines the spherical codebook.

The spherical codebook has a rather complicated and nonlinear structure. It is not a trivial problem to find the optimal codeword which minimizes the distortion for a given set of wavelet coefficients. Building the spherical tree using true coefficient values and coding this original tree does not lead to an optimal answer. This could be easily understood by looking at how the coder behaves in smooth regions of the subband. Insignificant energies could add up to be significant, and the coder could end up spending bitrate coding the total energy, not knowing that this energy comes from an insignificant region of the subband. Since the coder wastes bitrate for coding insignificant sets, the resulting codeword cannot be the optimal one.

Ideally, spherical coding tree has to be constructed using optimally decoded wavelet values. However, there is no obvious

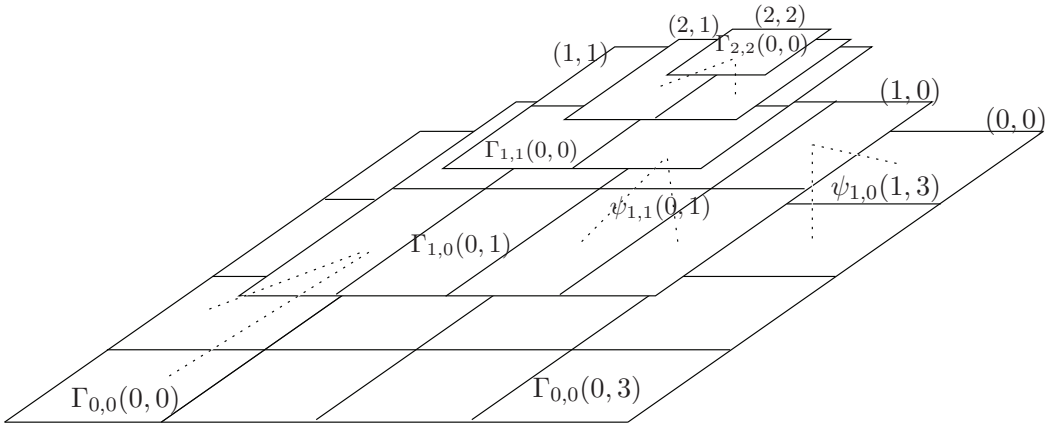


Fig. 4: Spherical coder.

way of determining these optimal values. Here, we propose a simple strategy for estimating the optimal spherical tree. First, wavelet coefficients are thresholded using a dead-zone interval. The dead-zone interval is used to find an initial set of coefficients that must be quantized to zero for rate-distortion efficiency. After thresholding, we perform a Lagrangian cost analysis to find any other set of coefficients that should also be quantized to zero. Going from the bottom to the top of the spherical tree, we compare the Lagrangian cost of zero-quantizing all coefficients of a given node to the best alternative associated with choosing not to do so. The latter is equal to the cost of coding the assigned phase value of the node plus the best cost of the two children nodes (See Figure 4). At the end, coefficients that belong to zero nodes are set to zero. The spherical tree is constructed and coded with these quantized coefficients. It turns out that this is an effective way of determining the set of zero-quantized coefficients, which is essential for successful coding performance.

In more detail, the spherical coding algorithm is given (for each wavelet subband at different scales and in different orientations) as follows (assume $0 \le m, n < 2^J$):

1) Use soft-thresholding to estimate zero-quantized wavelet coefficients:

$$\tilde{c}(m,n) = \begin{cases} c(m,n) & \text{if } |c(m,n)| > T \\ 0 & \text{else} \end{cases}$$

2) Define, for $0 \le u \le J$ ($v = u-1, u$),

$$\Gamma_{u,v}(s,t) = \sum_{m=2^u s}^{2^u(s+1)-1} \sum_{n=2^v t}^{2^v(t+1)-1} \tilde{c}(m,n)^2$$
$$0 \le s < 2^{(J-u)}, 0 \le t < 2^{(J-v)},$$

and, for $0 \le u < J$,

$$\psi_{u+1,u+1}(s,t) = \arctan\left(\frac{\sqrt{\Gamma_{u+1,u}(s,2t)}}{\sqrt{\Gamma_{u+1,u}(s,2t+1)}}\right)$$
$$0 \le s,t < 2^{(J-u-1)},$$
$$\psi_{u+1,u}(s,t) = \arctan\left(\frac{\sqrt{\Gamma_{u,u}(2s,t)}}{\sqrt{\Gamma_{u,u}(2s+1,t)}}\right)$$
$$0 \le s < 2^{(J-u-1)}, 0 \le t < 2^{(J-u)}.$$

The decoded values are represented as $\hat{\Gamma}_{u,v}(s,t)$ and $\hat{\psi}_{u,v}(s,t)$.

3) **Optimizing spherical representation:** Compare the Lagrangian cost of sending coded values of the wavelet coefficients to the cost of quantizing them all to zero. If the latter cost is smaller, then quantize to zero. Define $L_{u,u}(s,t)$ as the Lagrangian cost. For $0 \le m, n < 2^J$,

$$L_{0,0}(m,n) = (c(m,n) - \tilde{c}(m,n))^2 + \lambda I(m,n)$$

where $I(m,n)$ represents the sign bit for coefficient $c(m,n)$, i.e.

$$I(m,n) = \begin{cases} 0 & \text{if } \tilde{c}(m,n) = 0 \\ 1 & \text{else} \end{cases}$$

Then, set $u = 1$. While $u < J$ do,

- For $0 \le s < 2^{(J-u)}, 0 \le t < 2^{(J-u+1)}$, define

$$L_{u,u-1}(s,t) = L_{u-1,u-1}(2s,t) + L_{u-1,u-1}(2s+1,t) + \lambda \log_2(K_{u,u-1}(s,t)+1)$$

where the Lagrangian cost for coding $\psi_{u,u-1}(s,t)$, i.e. $\lambda \log_2(K_{u,u-1}(s,t)+1)$ (see step 4), is added to the total cost of two subtrees in order to get the total cost of the coefficients related to this node. Consequently,

$$L_{u,u-1}(s,t) > \sum_{m=2^u s}^{2^u(s+1)-1} \sum_{n=2^{u-1}t}^{2^{u-1}(t+1)-1} c(m,n)^2$$
$$\Rightarrow \tilde{c}(m,n) = 0$$
$$\forall\, 2^u s \le m < 2^u(s+1), 2^{u-1}t \le n < 2^{u-1}(t+1).$$

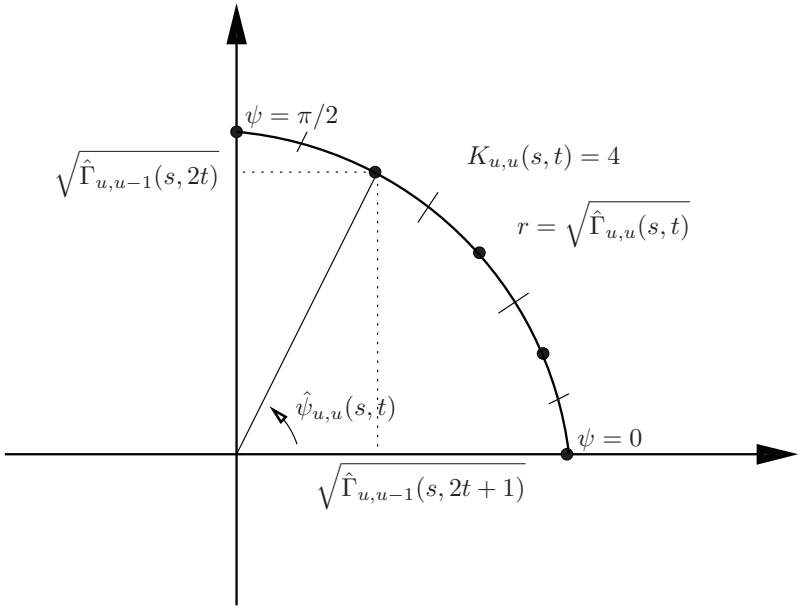


Fig. 5: Circular quantization.

- For $0 \le s,t < 2^{(J-u)}$, repeat the same procedure for $L_{u,u}(s,t)$.
- Increment u and repeat step 4.

4) **Decoding:** Code and send $\Gamma_{J,J}(0,0)$. Set $u = J$. While $u > 0$ do,
   - For $0 \le s,t < 2^{(J-u)}$, code $\psi_{u,u}(s,t)$ with a uniform quantizer for the interval $[0, \pi/2]$. Normalize the step size based on the magnitude, such that there are an integer number of quantization cells in the interval (see Figure 5):

$$q_{u,u}(s,t) = \frac{\pi}{2K_{u,u}(s,t)}$$

$$\text{where} \quad K_{u,u}(s,t) = \left\lfloor \frac{\pi/2\sqrt{\hat{\Gamma}_{u,u}(s,t)}}{q} + 0.5 \right\rfloor.$$

   Two quantization levels are placed at $\hat{\psi} = 0$ and $\hat{\psi} = \pi/2$. The other quantization points are chosen accordingly. Therefore, there are $K_{u,u}(s,t) + 1$ quantization levels (since 0 and $\pi/2$ have quantization intervals of half the size). Note that, if $K_{u,u}(s,t) = 0$, then there is no need for coding the phase values, and the iteration stops for such a node. In this case, all the coefficients of the node are quantized to zero.
   - For $0 \le s,t < 2^{(J-u)}$, decode the local energies,

$$\hat{\Gamma}_{u,u-1}(s,2t) = \hat{\Gamma}_{u,u}(s,t)\left(\sin(\hat{\psi}_{u,u}(s,t))\right)^2$$

$$\hat{\Gamma}_{u,u-1}(s,2t+1) = \hat{\Gamma}_{u,u}(s,t)\left(\cos(\hat{\psi}_{u,u}(s,t))\right)^2$$

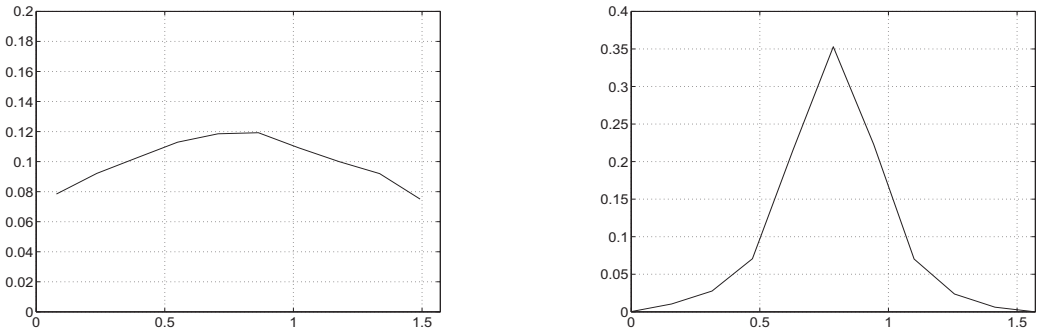
   - For $0 \le s < 2^{(J-u)}, 0 \le t < 2^{(J-u+1)}$, repeat the procedure for $\hat{\psi}_{u,u-1}(s,t)$ and $\hat{\Gamma}_{u-1,u-1}(2s,t), \hat{\Gamma}_{u-1,u-1}(2s+1,t)$.
   - Decrement u and repeat step 3.

   At the end of decoding, we have, for $0 \le m,n < 2^J$, the decoded wavelet coefficients:

$$\hat{c}(m,n) = \text{sign}(c(m,n))\sqrt{\hat{\Gamma}_{0,0}(m,n)}.$$

In the algorithm, q and T are chosen as the optimal quantization step size and the optimal dead-zone interval size, respectively, for best rate-distortion performance for a given Lagrangian multiplier λ. For a given bitrate, optimal λ is found using the convex bisection algorithm of [14]. Note that, optimal λ is equal to the slope of the rate-distortion curve of the spherical coder at its operating point. Starting from two extreme points of rate-distortion curve, the bisection algorithm shrinks the interval in which the optimal point lies until it converges. If the algorithm does not converge after a certain number of iterations, then the value of λ is incremented or decremented in small steps to find the optimal operating point.

Arithmetic coding is used to code the phase variables. The spherical tree provides a natural context for adaptive arithmetic coding. The coding model of each phase value $\psi_{u,v}$ is adapted based on the corresponding number of quantization levels, $K_{u,v} + 1$, and the level of the tree, i.e. (u,v) pair. Figure 6 plots the histograms of phase variables at highest scale vertical subband at different levels of the spherical tree. Uniform distribution seems to provide a good fit to phase histograms at lower levels. The distribution gets more peaked around 45 degrees at higher levels of the spherical tree. However, since the number of phase variables drops as 2^{uv}, the use of true histograms in arithmetic coding does not provide much coding gain over uniform distribution. As a result, the bitrate of the arithmetic coder turns out to be only slightly better than the entropy estimate based on the uniform distribution. In other words, it is justified to use the self-information $\log_2(K_{u,v}+1)$ for estimating the bitrate of each phase variable.



(a) (b)

Fig. 6: Histogram of phase variables: (a) $u = 1, v = 0, 1$; (b) $u = 3, v = 2, 3$.

While encoding/decoding the spherical tree, once the algorithm reaches to a "zeronode", all the coefficients that belong to that node are set to zero and no further bitrate is spent for coding the remaining phase values. Therefore, the comparison of Lagrangian cost between two modes of quantization, namely "zeronode" quantization and "spherical" quantization is essential for achieving successful coding results.

The performance of the algorithm is very much dependent on how the spherical tree is constructed. Note that, the optimization step, i.e. step 3, tries to find the set of coefficients that are to be quantized to zero, and doesn't provide estimates for decoded values of remaining coefficients. In other words, zeronode quantization is introduced as the only alternative to standard spherical coding of phase variables. In principle, it is possible to include more sophisticated vector quantization modes into the search list. Yet, this surely turns Lagrangian optimization into a much harder problem.

It is rather challenging to figure out what the best strategy is, and how much better (in terms of total Lagrangian cost) the decoded values can get. The answer lies in the complicated structure of the codebook generated by the spherical quantization and coding strategy. The quantization of phase is very much dependent on the decoded magnitude, which is related to the previously decoded phase values. Therefore, the possible set of codewords have a rather complicated structure to visualize. It is an open research problem to develop a better understanding of the nature of this codebook and to improve the coding algorithm.

V. Simulations and Discussions

Spherical coder is implemented using biorthogonal linear phase filter pairs in a 6-level dyadic decomposition. Quantization step size used for all phase variables in all subbands is the same, up to the necessary normalization factor. Optimal q and T are chosen among the set $\{t : t = 0.1k\pi, k = 1, 2, ..., 150\}$. Low-pass subband is arithmetic coded, after applying an (8×8) DCT, using optimal scalar quantizer for a given λ.

The performance of the spherical coder is compared to that of some of the best performing coders in the literature [15], including SPIHT [5], SFQ [6], EQ [10], EBCOT [12] and EZBC [16]. *Lena*, *Goldhill* and *Barbara* images are used for comparison. All results are for the 9/7 filter pair, except for EQ which uses the 10/18 filter. The results are reported at 1.00, 0.50, 0.25 bpp (see Table I).

The spherical coder, called as SPHE in Table I, outperforms SPIHT, and is as good as SFQ, EBCOT and EZBC in most cases. The slightly better performance of EQ coder is partially due to the use of 10/18 filter. In Figure 7, PSNR for *Lena* is plotted against different bitrates for SPHE and EBCOT. Except for *Barbara*, the performance of SPHE is about the same as that of EBCOT, which is the algorithm used in JPEG2000 standard. Note that, EBCOT uses sophisticated contextual models which can adapt well to the local frequency content of textured regions in images such as *Barbara*. Considering the simplicity of the modeling choices we have made in the spherical coder, these results are rather encouraging for our future efforts in developing highly efficient and adaptive coding methods based on the spherical representation.

Table II provides PSNR results of SPHE using both 17/11 and 9/7 filter pairs. With 17/11 filter, PSNR is 0.05-0.1 dB better for *Lena* and *Goldhill*, and 0.3-0.4 dB better for *Barbara*. This is because 17/11 filter achieves better compaction of energy in wavelet subbands. This energy compaction is more pronounced for textured images such as *Barbara*.

The performance of the spherical coder could possibly be improved in many different ways. Using uniformly distributed independent phases simplifies the algorithm, and introduces some form of non-homogeneity. But this assumption is not quite right for modeling the actual non-homogeneity of wavelet subbands. There exist complicated high-order dependencies among phase variables. In addition, there exist inter-band dependencies among phase variables that belong to the same spatial locations in different subbands. Since the spherical representation is robust to coding mismatch, the spherical coder with the independence assumption is still very successful. Based on the discussions of Section III, if we can model these dependencies at different levels of the hierarchy, and manage to decode optimal or close to optimal local energies, we expect significant overall coding gains.

As for the computational cost of the algorithm, the most time consuming part is to find the optimal parameter set (q, T, λ) for a target bitrate. Due to this exhaustive optimization, the complexity of the algorithm is comparable to that of SFQ and EQ, and quite higher than the other algorithms

TABLE I: PSNR comparison of different coders.

Lena	PSNR (dB)					
Rate (bpp)	SPHE	SPIHT	SFQ	EQ	EBCOT	EZBC
1.00	40.67	40.46	40.52	40.88	40.55	40.62
0.50	37.40	37.21	37.36	37.69	37.43	37.47
0.25	34.28	34.11	34.33	34.57	34.32	34.35

Goldhill	PSNR (dB)					
Rate (bpp)	SPHE	SPIHT	SFQ	EQ	EBCOT	EZBC
1.00	36.85	36.55	36.70	36.96	36.75	36.90
0.50	33.37	33.13	33.37	33.44	33.38	33.47
0.25	30.72	30.63	30.71	30.76	30.75	30.74

Barbara	PSNR (dB)					
Rate (bpp)	SPHE	SPIHT	SFQ	EQ	EBCOT	EZBC
1.00	37.00	36.41	37.03	37.65	37.38	37.28
0.50	32.06	31.40	32.15	32.87	32.50	32.15
0.25	28.22	27.58	28.29	28.48	28.53	28.25

TABLE II: PSNR results for SPHE using 17/11 and 9/7 filters.

SPHE	PSNR (dB)					
	Lena		*Goldhill*		*Barbara*	
Rate (bpp)	17/11	9/7	17/11	9/7	17/11	9/7
1.00	40.74	40.67	36.91	36.85	37.39	37.00
0.50	37.50	37.40	33.42	33.37	32.44	32.06
0.25	34.38	34.28	30.76	30.72	28.52	28.22

mentioned above. However, we believe that an exhaustive search for the optimal parameter set could be avoided by modeling the relationships between these parameters. As for the coding procedure for fixed values of (q, T, λ), the computational complexity is reasonable. For building the tree, the cost calculations require simple addition and comparison operations at each node. During decoding, the number of quantization bins is computed and uniform scalar quantization is performed for each node. A significant portion of the coding complexity is due to context-based arithmetic coding of the quantized phase variables. For hardware implementation, in both coding stages, the different nodes at a certain level of the hierarchy could be processed in parallel, which could significantly speed-up the execution.

VI. Conclusion

In this paper, we have introduced and analyzed the spherical representation as a convenient and flexible framework for developing adaptive models for wavelet information. A simple application of the framework in wavelet subbands has led to the spherical coding algorithm. The competitive results attained by the spherical coder point towards the potential of such energy-based representations in modeling wavelet subbands.

On a more philosophical point, spherical coding technique is based on an orthogonal representation which is quite different than the usual orthogonal bases of Cartesian coordinate system. Instead, the phase variables here could be seen as the basis vectors of the spherical coordinate system. To be more accurate, in its current form, the spherical coder is a combination of both coordinate systems, since wavelet transformation is applied first and the spherical coordinates are

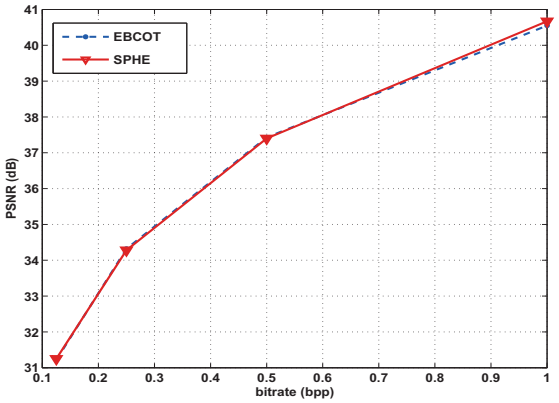


Fig. 7: PSNR versus bitrate for SPHE and EBCOT.

used independently in each subband. The phase coordinates could also be defined in different ways than the hierarchical way we did in our algorithm. One might think of various other ways to use these two types of representations together for developing interesting coding strategies. This could lead to a coding theory much more general than the theory of transform coding.

Spherical representation could find interesting applications in fields other than coding wavelet subbands. One such area is the study of turbulence. Multifractals are extensively used in this field [17], [18], mainly to describe the spatial dissipation of turbulent energy. There exist several techniques to analyze the multifractal nature of given data using different statistical tools, such as the multifractal spectra [19]. In contrast to such global descriptions, spherical representation could be used to develop more localized statistical models for these kind of processes.

Spherical coder is a basic implementation of the ideal adaptive coding methods that we are looking for, mainly because it doesn't rely on a deep understanding of natural images. We expect to develop more intelligent coding techniques and achieve much better results, if we can model the dependencies that exist among local energy and phase variables.

Spherical coder, as described in this paper, is a non-progressive lossy coding method. Our current work also focuses on different versions of spherical coder for lossless coding and for progressive coding by modifying the way in which the spherical tree is constructed and coded.

References

[1] H. Ates and M. Orchard, "Wavelet image coding using the spherical representation," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, Genova, Sept. 2005, pp. 89–92.

[2] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Processing*, vol. 1, no. 2, pp. 205–20, April 1992.

[3] ——, "Image coding using vector quantization in the wavelet transform domain," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 4, Albuquerque, NM, April 1990, pp. 2297–300.

[4] J. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3445–62, Dec. 1993.

[5] A. Said and W. Pearlman, "A new fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 6, no. 3, pp. 243–50, June 1996.

[6] Z. Xiong, K. Ramchandran, and M. Orchard, "Space-frequency quantization for wavelet image coding," *IEEE Trans. Image Processing*, vol. 6, no. 5, pp. 677–93, May 1997.
[7] ——, "Joint optimization of scalar and tree-structured quantization of wavelet image decomposition," in *Con. Rec. 33th Asilomar*, vol. 2, Pacific Groove, CA, Nov. 1993, pp. 891–5.
[8] K. Ramchandran and M. T. Orchard, "An investigation of wavelet-based image coding using an entropy-constrained quantization framework," *IEEE Trans. Signal Processing*, vol. 46, no. 2, pp. 342–53, Feb. 1998.
[9] Z. Xiong, N. Galatsanos, and M. Orchard, "Marginal analysis prioritization for image compression based on a hierarchical wavelet decomposition," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 5, Minneapolis, MN, April 1993, pp. 546–9.
[10] S. M. LoPresto, K. Ramchandran, and M. T. Orchard, "Image coding based on mixture modeling of wavelet coefficients and a fast estimation-quantization framework," in *Proc. Data Compression Conf.*, Snowbird, UT, March 1997, pp. 221–30.
[11] R. Joshi, H. Jafarkhani, and *et al.*, "Comparison of different methods of classification in subband coding of images," *IEEE Trans. Image Processing*, vol. 6, no. 11, pp. 1473–86, Nov. 1997.
[12] D. Taubman, "High performance scalable image compression with EBCOT," *IEEE Trans. Image Processing*, vol. 9, no. 7, pp. 1219–35, July 2000.
[13] R. Gray and T. Linder, "Relative entropy and quantizer mismatch," in *Con. Rec. 36th Asilomar*, vol. 1, Nov. 2002, pp. 129–33.
[14] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate-distortion sense," *IEEE Trans. Image Processing*, vol. 2, pp. 160–75, April 1993.
[15] http://www.icsl.ucla.edu/∼ipl/psnr_results.html, "Wavelet image coding: PSNR results."
[16] S. Hsiang, "Embedded image coding using zeroblocks of subband/wavelet coefficients and context modeling," in *Proc. Data Compression Conference (DCC '01)*, Washington, 2001, pp. 83–92.
[17] C. Meneveau and K. Sreenivasan, "Simple multifractal cascade model for fully developed turbulence," *Physical Review Letters*, vol. 59, no. 13, pp. 1424–7, Sept. 1987.
[18] ——, "The multifractal nature of turbulent energy dissipation," *Journal Fluid Dynamics*, vol. 224, pp. 429–84, March 1991.
[19] R. Riedi and I. Scheuring, "Conditional and relative multifractal spectra," *Fractals*, vol. 5, no. 1, pp. 153–68, March 1997.