MORPHOLOGICAL REPRESENTATION OF WAVELET DATA FOR IMAGE CODING

Sergio Servetto and Kannan Ramchandran and Michael Orchard

Coordinated Science Lab. and Beckman Institute
University of Illinois at Urbana-Champaign

ABSTRACT

We propose an improved statistical characterization of the field of wavelet coefficients of natural images. Based on this characterization, we introduce Morphological Representation of Wavelet Data (MRWD), a novel coding framework for both image and video coding applications. MRWD departs from existing wavelet-based coders in its use of a radically different set of primitive operations –non-linear, morphological operations–, for efficiently encoding the wavelet data field. Simulation results are very encouraging: a preliminary algorithm based on the morphological data structure is able to achieve about 0.5 dB of gain in SNR over Shapiro's state-of-the-art zerotree-based wavelet coder [1] at a coding rate of 1 bpp for the "Lenna" image.

1. INTRODUCTION

Since the introduction of wavelet bases for signal representation, much attention has focused on their application to image coding. This interest stems from the fact that wavelets provide a space-frequency decompostion of images that allows both good frequency compaction of energy (typically into low-frequency coefficients) and good space localization of energy around edges. To exploit these compaction properties, an image coding algorithm requires an accurate statistical characterization of the joint distribution of wavelet coefficients. This paper proposes an entirely new characterization (based on simple non-linear operators) of the spatial energy distribution across wavelet bands. The new characterization motivates the formulation of a new, high-performance coding algorithm based on morphological representation of wavelet coefficients.

Typical transform coding algorithms consist of three stages: a linear invertible transform (for decorrelation) followed by a (lossy) quantization stage, and entropy coding of the quantized data. An important feature of a good linear transform is its ability to compact maximum energy into few coefficients so that the second stage of the algorithm can quantize a large number of coefficients to zero. Substantial coding gain is achieved by exploiting the first-order distribution of these zeros (i.e. the marginal probability of zeros). The standard JPEG algorithm significantly improved on earlier approaches to transform coding (e.g. zonal quantization) through its use of zero-runlengths and end-of-block symbols to efficiently represent zeros. As further evidence of the importance of zero coefficients, recent research [5] has demonstrated substantial improvement on standard JPEG encoding by optimizing the quantization of zeros.

The wavelet's ability to compact energy into few, typically low-frequency, coefficients, while compacting edge energy into few, spatially localized, high-frequency coefficients makes it a very promising linear transform for application to transform coding. Compared with a DCT-based algorithm, one would expect a wavelet coder to achieve similar decoded image fidelity using fewer non-zero (denoted ***significant***) coefficients, with the reduced number of significant coefficients yielding improved coding efficiency. However, to realize the expected improvements in coding performance, an efficient method is needed for representing the large number of zero coefficients (equivalently, for representing the location of the significant coefficients). Early wavelet coders (e.g. [4]), based on classical bit allocation strategies, use a first-order probability model of frequency bands of wavelet coefficients as the basis for coding the high-band coefficients, thus effectively exploiting the marginal probability of zero-valued coefficients in these bands. However, recent work with wavelet transforms have found that the field of zero coefficients exhibit significant spatial dependencies. Specifically, the quantized zero-valued (likewise, for the significant) wavelet coefficients appear much more clustered than would be expected for a random, 2-D Poisson distribution of points having the same marginal probability (see Fig. 1). A natural coding framework for exploiting these spatial dependencies is to represent the field of coefficients as a composite of two distinct information sources: first, a source specifying the location of significant coefficients, and second, a source providing the values of significant coefficients. Both zerotree quantization, applied in [1, 2], and the algorithm presented in this paper are solutions for the coding problem posed by the first of these information sources: How does one efficiently represent the binary field specifying the location of significant coefficients? We refer to this binary field as the ***significance field***.

If coefficients were independently distributed within each band of wavelet coefficients (i.e. approximating a 2-D Poisson point process - possibly with different densities in different bands), the significance field could be optimally coded with classical approaches to optimal bit allocation among frequency bands. While studies have reported lack of ***correlation*** among high-frequency wavelet coefficients, there is ample evidence of substantial dependencies (e.g. image structure can be clearly recognized in the significance field). [3] presents empirical evidence of dependencies in the significance field, and uses this evidence to explain the improved

0-7803-2431-5/95 $4.00 © 1995 IEEE

Figure 1: Location of coefficients in the wavelet domain.

performance achieved by zero-tree quantization methods. Because groups of zero-valued wavelet coefficients occur in natural images with much higher probability than would be suggested by the marginal probability of zeros, coding performance can be improved by more efficiently representing groups of zeros. Zerotree quantization offers one approach for realizing this improvement. Algorithms incorporating zerotree quantization introduce a special symbol (zerotree symbol) to represent certain tree-structured sets of zero coefficients. Using a rate-distortion optimality criterion for deciding when to apply this special symbol, wavelet coders realize 1-2 dB PSNR improvement (at 1 b/p) on typical images.

Zerotree quantization can be viewed as a scheme for efficiently identifying zero regions in the significance field, leaving the coder to use standard methods to define the significance field everywhere outside these regions. The main drawback of zerotree quantization is that it approximates arbitrarily-shaped zero regions of the significance field as the union of a highly constrained set of tree-structured regions. Due to these constraints, certain zero regions, not well aligned with the tree-structure grid, may be very expensive (in bitrate) to represent, and many portions of zero regions are not included as zerotrees at all. The next section proposes a way to efficiently represent arbitrarily shaped regions of zeros in the significance field.

2. MORPHOLOGICAL PREDICTION OF SIGNIFICANCE

Zerotrees indirectly exploit clustering in the significance field by identifying regions of zero significance, thereby implying the clustering of significance in the remaining regions. We propose to directly exploit clustering, by characterizing the high probability of significance in regions close to coefficients which have already been found to have high energy. This approach assumes sequential transmission of both the significance field and the values of significant coefficients (i.e. once the significance of a coefficient is known, its value is immediately transmitted). Based on information provided by previously transmitted coefficients, our method will predict a region on which there is a high probability of significance. This section describes the operation of morphological prediction, which will serve as the basic building block of the algorithm defined in the next section.

At every stage i in the coding process (ignoring initialization), consider a partitioning of a certain band of wavelet coefficients into Q_i, the regions of the significance field that have been transmitted prior to stage i, and $\bar{Q}_i$, the remaining coefficients. Note: Q_i includes both significant coefficients and zero-valued coefficients. It is possible to code the coefficients on $\bar{Q}_i$ using standard methods which would characterize the first-order probability of values on $\bar{Q}_i$. However, we can outperform such an approach by first identifying some subset of $\bar{Q}_i$ whose probability of significance is much higher than on $\bar{Q}_i$ overall, and sending that subset with respect to an accurate first-order probability model. We use the morphological operation of dilation to identify such a subset with high likelihood of significance.

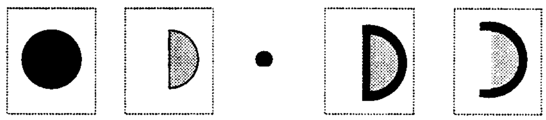

Figure 2: Morphological Dilation.

Let S_i denote a subset of Q_i containing all coefficients found to be significant. Let B represent some structuring element (an arbitrary set), and let $\oplus$ denote the dilation operator. We predict that the subset $(S_i \oplus B) - Q_i$, where $X - Y$ is defined as all elements of X not contained in Y, has high likelihood of significance. Figure 2 illustrates how this set is constructed. Assume the sets Q_i and $S_i \subset Q_i$, as shown. The structuring element B is an arbitrary set which has the intuitive role of defining a morphological distance. The elements of $(S_i \oplus B)$ can be viewed as all elements some distance from S_i, where B defines the distance. The dilation $(S_i \oplus B)$ is defined as the union of many translations of the set B, each centered on some element of S_i. Thus, for common structuring elements like B, dilation produces an enlarged set containing the original S_i, plus a set of nearby elements. [3] provides empirical evidence that, if S_i contains the large-valued wavelet coefficients, the set $(S_i \oplus B) - Q_i$ has much higher first-order probability of significance than the set $\bar{Q}_i$. (Note: In fact, the concentration of significance in $(S_i \oplus B) - Q_i$ could be viewed as a definition of clustering of significant coefficients.)

3. THE MRWD ALGORITHM

The morphological predictor defined in the previous section provides the basic building block used by the MRWD Algorithm to represent the significance field. This section details the algorithm structure. Our algorithm adopts the standard architecture for subband coders: a unitary linear transform for pixel decorrelation, quantization of the coefficients, and entropy coding of the resulting symbol stream.

Linear Transform using wavelets: We use the biorthogonal spline variant filters with less dissimilar length described in [4] to calculate the wavelet transform. These filters are chosen because they satisfy the perfect reconstruction property with linear phase, thereby permitting a graceful way of dealing with the image boundaries using symmetric extension. For computational reasons, as in [2], we treat our filters as being approximately orthogonal (with little loss in performance) so as to enable rate-distortion analysis (using a squared-error distortion metric) directly in the wavelet domain. We grow the wavelet tree to a depth of 4 in our simulations.

3.1. Quantization: Clusters of Signal Energy

In order to address the efficient management of the quantized wavelet data (we use simple uniform scalar quantization), we invoke a new data structure centered around *clusters* of wavelet coefficients. Our structure is motivated by the empirically verified hypothesis (see Fig. xxx) that signal energy in the high-frequency wavelet bands tends to be highly clustered or localized (corresponding to the presence of edges/textures in the original image). We propose to iteratively apply the morphological predictor described in the previous section to efficiently represent these clusters. The net effect of the iterative process is to partition the wavelet data into two sets: one set (corresponding to edges and textures) having relatively significant coefficients, and the other (residue) set having a very low probability of signficant values. The first set can be described as Q_I, for the final stage $i = I$ of the algorithm, and is entirely represented using morphological prediction. The second set $\bar{Q}_I$ can either be ignored completely (a reasonable strategy at low bit-rates), or isolated coefficients in $\bar{Q}_I$ can be represented with various ad-hoc approaches. Note that, in the first case where $\bar{Q}_I$ is ignored, arbitrarily shaped regions of zero coefficients are implicitly defined with *no* additional bitrate!

To efficiently represent the set of clusters via sequential morphological prediction, a seeding process is needed. (Recall: each stage of morphological prediction assumes a set Q_i of previously transmitted coefficients.) For each cluster of significant coefficients, the encoder initializes Q_0 to a "seed" value corresponding to a single significant coefficient location within the energy cluster (typically to the largest magnitude in the region), explicitly sending its location and value. Based on morphological prediction, the significance field (and corresponding significant values) is transmitted for an expanded region around the seed. Region growing finally terminates when the energy "island" is completely surrounded by a strip of insignificant values, where the strip width depends on the size of the structuring element size B. Thus, our coding data structure results in an efficient data-dependent way of capturing the shape of these energy clusters.

Note that our new representation implies a *data-dependent* scanning order of the coefficients. Under the assumption of clusters surrounding high energy coefficients (with the amount of energy decaying with increasing distance), this scanning order results in a concentration of coefficients with large magnitudes at initial positions in the scan order. In order to achieve higher coding efficiency, we apply marginal return analysis in a rate-distortion (R-D) framework to determine the optimal termination point (i.e. *cluster length*) in the scan. That is, the encoder reserves the right to send an End-Of-Block (EOB) symbol to signal the end of the cluster-scan path to the decoder at any point in the stream. R-D optimization is performed by the encoder to determine *where* to optimally terminate the scan. In [5], a method was proposed to implement this optimization for JPEG, where each DCT block is scanned in a fixed order, and both the optimal location of the EOB symbol and the optimal subset of non-zero coefficients is found by minimizing a Lagrangian cost function. In our case, we regard clusters as generalized, data-driven-shaped blocks. As the scanning order and the shape of the generalized blocks both depend on the actual values sent, the analysis of the deletion of "midstream" coefficients becomes very complex. However, because high energy values tend to concentrate at the beginning of the cluster, we hope that by trimming its tail, considerable savings in rate can be obtained, for a moderate increase in distortion. To implement trimming, we define a Lagrangian cost function $J(C_n, \lambda) = D_n + \lambda R_n$, to be minimized for each cluster:

$$\begin{aligned}
C_i &= \text{First } i \text{ coefficients in cluster } C \\
D_i &= \sum_{k=0}^{i-1}(c_k - \hat{c}_k)^2 + \sum_{k=i}^{n-1} c_k^2 \\
R_i &= \log_2(NM) + \sum_{k=0}^{i-1} \log_2(1/P(\hat{c}_k))
\end{aligned}$$

where $\hat{c}_k$ are the quantized-dequantized coefficients, c_k are the unquantized coefficients, $P(\hat{c}_k)$ is the estimate of the first order probability mass function obtained from computing the histogram of the quantized bins, and the term $log_2(NM)$ in the rate accounts for having to explicitly send the seeds (NxM is the image size). Notice how this constant factor added to the cost of each cluster reflects the fact that this primitive is more efficient when applied to large, high energy clusters, and acts penalizing small ones with little contribution to distortion reduction

3.2. Entropy Coder: Adaptive Arithmetic Coding

The resulting stream of quantized bins is compressed via adaptive arithmetic coding. We use a size 2 alphabet for this coder, where each symbol is assigned a sequence of bits proportional in length to its magnitude, because we found empirically that this yields better compression than using an alphabet with the number of symbols equal to the number of quantization levels. We attribute the improved performance achieved with the small alphabet coder to its ability to adapt quickly to data statistics.

4. EXPERIMENTAL RESULTS

Two versions of the MRWD algorithm were tested on the standard Lena test image. Codec1 only codes clusters of coefficients, ignoring $\bar{Q}_I$, for $i = I$ the final stage of morphological prediction. At high bitrates, Codec1 has the disadvantage of not be able to represent isolated coefficients with significant energy. Codec2 codes clusters of coefficients like Codec1, but reserves some bitrate for representing isolated coefficients in $\bar{Q}_I$. For the Lena image, Codec1 achieves a performance of 39.6dB at 1bpp, matching results obtained with the standard zerotree algorithm, but falling 1dB below the rate-distortion optimized version of zerotrees. Codec2 improves over codec1 by 0.3dB at the same bitrate, suggesting a mixture model of cluster processes, with important energy existing in clusters with varying densities.

While the rate-distortion optimized zerotree algorithm (SFQ) of [2] outperforms the MRWD algorithm by the objective measure of PSNR, we suggest the possibility of advantages in subjective quality due to investing bits in areas of clustered energy (usually associated with signals) at the expense of bits to isolated coefficients (often associated with noise). Our final simulations explore this this noise reduction property of the proposed algorithm. We created a synthetic image (see Fig. 4), to which white noise was added. This addition of noise results, in the wavelet domain, in a number of randomly located low-energy coefficients. The noisy image was then encoded both with codec1 and SFQ, with a target bitrate = 0.1bpp. The SNR of the decoded images against the noisy original was 27dB for SFQ, and 26.85dB for codec1. However, the SNR against the noiseless original was 28.1dB for SFQ, and 28.7dB for codec1. In Fig. 5, the locations of coefficients for the noisy and noiseless image, and those selected by codec1 for transmission is shown.

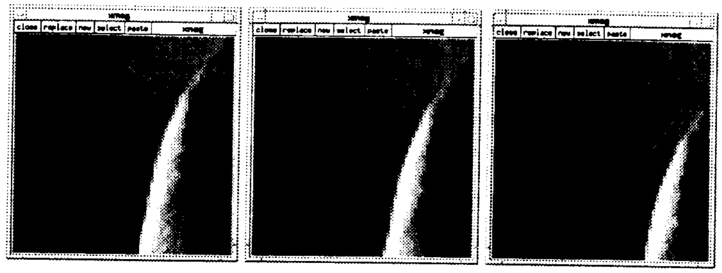

Figure 3: Zooming on Lena's hat: (a) original, (b) MRWD decoded, (c) SFQ decoded.

5. CONCLUSIONS AND FURTHER WORK

The new characterization of the random field of wavelet coefficients of image data has been proposed, leading to a morphological prediction representation of clusters of image energy. Based on this new model, two codecs were constructed, and their performance tested on real and synthetic data.

We anticipate 4 directions of future research: 1) expanding the current approach to support morphological prediction of significant values (in addition to the significance field); 2) developing codecs based on a mixture model of cluster densities; 3) integrating morphological predictors with motion compensation, for video coding; and 4) developing an embedded codec to support progressive transmission.

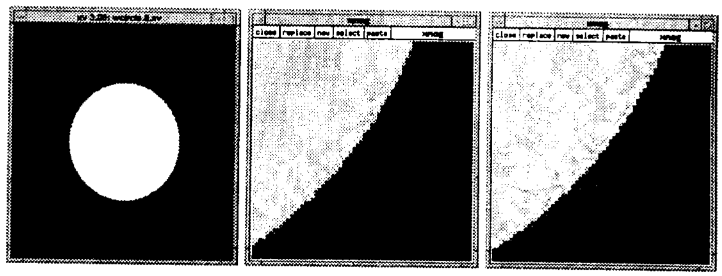

Figure 4: Reconstructions of the synthetic image: (a) original, (b) MRWD decoded, (c) SFQ decoded.

6. REFERENCES

[1] J. Shapiro. Embedded Image Coding Using Zerotrees of Wavelet Coefficients. *IEEE Transactions on Signal Processing*, 41(12):3445–3462, December 1993.

[2] Z. Xiong, K. Ramchandran and M. T. Orchard, "Joint optimization of scalar and tree-structured quantization of wavelet image decomposition", *Proc. Asilomar conference on signals, systems and computers*, Nov. 1993

[3] M. T. Orchard and K. Ramchandran, "An investigation of wavelet-based image coding using an entropy-constrained quantization framework", *Proc. Data Compression Conference '94*, p. 341-350, Snowbird, Utah, March, 1994.

[4] M. Antonini and M. Barlaud and P. Mathieu and I. Daubechies, "Image Coding Using Wavelet Transform", *IEEE Trans. on Image Proc.*, Vol. 1, no. 2, pp. 205-221, Apr. 1992.

[5] K. Ramchandran and M. Vetterli, "Rate–distortion optimal fast thresholding with complete JPEG/MPEG decoder compatibility", *IEEE Trans. on Image Proc.*, Vol. 3, no. 5, pp. 700-705, Sept. 1994.

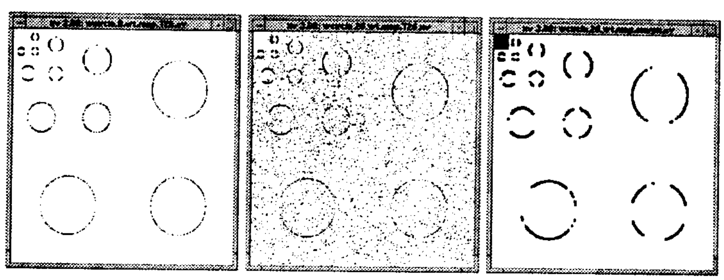

Figure 5: Location of coefficients in the wavelet domain for the synthetic image: (a) noiseless original, (b) noisy original, (c) MRWD selected.