

RATE-DISTORTION ADAPTIVE VECTOR QUANTIZATION FOR WAVELET IMAGE CODING

Qun Gu and Scott E. Budge

Electrical and Computer Engineering Dept.
Utah State University, Logan, UT 84322-4120
slkpc@cc.usu.edu, scott.budge@ece.usu.edu

ABSTRACT

We propose a wavelet image coding scheme using rate-distortion adaptive tree-structured residual vector quantization. Wavelet transform coefficient coding is based on the pyramid hierarchy (zerotree), but rather than determining the zerotree relation from the coarsest subband to the finest by hard thresholding, the prediction in our scheme is achieved by rate-distortion optimization with adaptive vector quantization on the wavelet coefficients from the finest subband to the coarsest. The proposed method involves only integer operations and can be implemented with very low computational complexity. The preliminary experiments have shown some encouraging results: a PSNR of 30.93 dB is obtained at 0.174 bpp on the test image LENA (512×512).

1. INTRODUCTION

With the boom in multi-media communication brought on by the Internet and the World-Wide-Web (WWW), image data compression has become an important element in reducing transmission time for digital imagery. One of the many compression methods which has drawn considerable attention in the literature and in standards bodies is compression based on the Discrete Wavelet Transform (DWT).

The wavelet transform takes advantage of joint time-frequency locality. In other words, wavelet transform coefficients in each subband contain not only frequency information but also spatial information. The spatial information in each subband leads to some *interband* relationships [1], which can be exploited to improve the *intra-band* coding. One example is the popular Embedded Zerotree of Wavelet coefficient (EZW) algorithm proposed by Shapiro [2]. The EZW algorithm uses an insignificance prediction based on a tree-structure called a *zerotree*. The insignificance prediction of the zerotree is based on the observation that DWT coefficients of a zerotree at different levels in the pyramid hierarchy are related to a same spatial area in the original image. If a coefficient at lower band is insignificant with respect to a given threshold, it will be reasonable to predict that all its descendant coefficients are insignificant as well.

Vector quantization (VQ) is a lossy compression technique that quantizes several samples simultaneously. VQ takes advantage of statistical dependencies among the samples in a vector. Although wavelet transform removes much of the correlation between image samples, high-order statistical dependencies may still exist within or across subband coefficients.

Residual VQ (RVQ) and Tree-structured VQ (TSVQ) are low-complexity VQ schemes with good rate-distortion performance [3]. Both of them provide successive approximation of the source or variable rate compression. Several researchers have applied RVQ, TSVQ or Tree-structured residual VQ (TRVQ), also known as multi-stage tree VQ, on the wavelet coefficients in their compression methods. Cossman *et al.* examined the variable-rate binary TSVQ on the wavelet coefficients with zerotrees [4]. They extended the zerotree thresholding by rate-distortion trade-offs. In their improved version [5], they applied TRVQ and claimed similar results. Jayashree Karlekar *et al.* applied RVQ on the finer levels of the pyramid hierarchy while still using EZW on the coarser levels [6]. Hamid Jafarkhani *et al.* employed a hierarchical pruned tree-structured vector quantizer (PTSQ) on wavelet coefficients [7]. They also tested the effect of non-square selection of vector shapes and concluded that the vector quantizer with rectangular vector shapes along the orientations of wavelet decomposition outperforms that with uniform square vector shapes.

In this paper, we propose an improved wavelet image coding method using TRVQ based on rate-distortion optimization of zerotree-like hierarchy. In section 2 we describe the algorithm and present results and conclusions in section 3.

2. IMAGE ENCODING ALGORITHM

2.1. Overview

In our proposed scheme, the input image is first transformed into wavelet domain by the two-dimensional separable octave decomposition which leads to a pyramid hierarchy. The top low-low subband is coded losslessly by a simple DPCM-Huffman coder. The other subbands are vector quantized based on a zerotree insignificance prediction. Only those significant vectors are actually VQ coded. The symbols of the significance map are embedded in VQ headers. A rate-distortion trade-off is applied instead of hard thresholding. The flow graph of this algorithm is shown in Fig. 1.

2.2. Wavelet transform

To allow modeling of fast computation and compatibility with hardware implementations, we adopted the integer wavelet transform defined by Daubechies' 9-7 filter pair with lifting scheme [8]. The analysis low-pass filter has 9 taps, and the analysis high-pass filter has 7 taps. Both the analysis and synthesis high-pass filters have 4 vanishing moments. Therefore, it provides good smoothness with moderate filter lengths. Its symmetric property allows simple edge

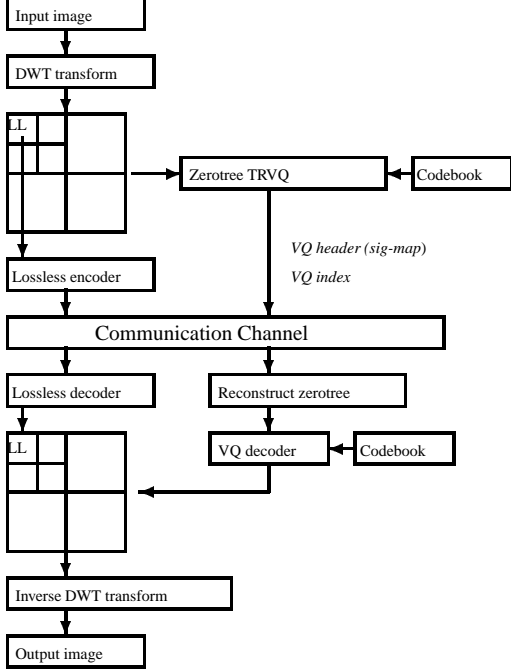


Figure 1: Flow graph of the proposed image coding algorithm.

treatment. Furthermore, it can be implemented to produce integer coefficients while still providing perfect reconstruction [9].

A two-dimensional octave DWT composition of an image will generate a pyramid hierarchy as Fig. 2, where the inter-band relations are also illustrated. Note that the HL subbands emphasize more horizontal details, the LH subbands stress more vertical information, and the HH subbands give more of the diagonal details of the image. In our proposed scheme, a 4-level two-dimensional DWT is applied.

2.3. Lossless coding of the Top Low-Low subband

Generally, most of the energy of images stays in the low-frequency band. After wavelet decomposition, the top low-low subband contains the low-frequency coefficients and contributes most to the visual fidelity of the reconstructed image. The top low-low subband coefficients are related to the local mean of the original image and their statistical distribution is also similar to the original image. For these reasons, we code the top low-low subband coefficients by a DPCM-Huffman coder. Differential Pulse Coded Modulation (DPCM) is commonly used to remove the statistical correlations between input samples. The resulting prediction errors can be efficiently coded by an entropy coder such as a Huffman encoder. The prediction of DPCM is achieved based on a three-point causal prediction window given as:

$$\tilde{u}(m, n) = \hat{u}(m, n - 1) + \hat{u}(m - 1, n) - \hat{u}(m - 1, n - 1) \quad (1)$$

where \hat{u} denotes the reconstructed value, \tilde{u} stands for the predicted value, m and n specify the location in the image. The Huffman table for the prediction-error is built based on the statistical distribution of a training sequence which is also used to develop the VQ codebook.

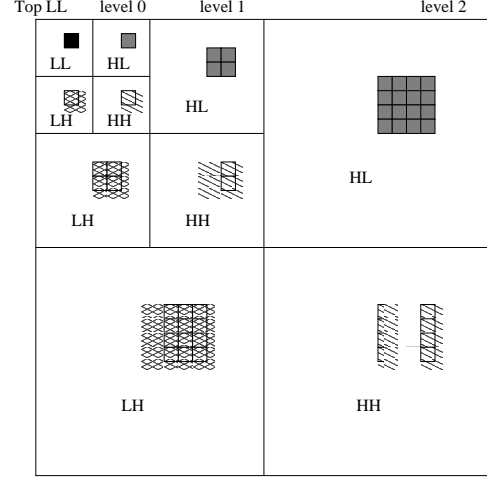


Figure 2: Pyramid hierarchy of 2-D octave DWT decomposition.

2.4. Tree-Structured Residual VQ (TRVQ)

RVQ is a structured VQ that reduces both the computational complexity and memory requirement of the coder. In addition, it can be implemented as a variable-rate VQ, which allows VQ to be performed with a rate-distortion trade-off. The basic idea of RVQ is to divide the VQ search into several VQ stages, where each stage uses a relatively small codebook to encode the residual error of the previous stage. However, RVQ has a significant performance degradation when the number of stages increases.

TSVQ is a very competitive VQ method which reduces the computational complexity with increasing memory requirement. In TSVQ, the VQ search is performed by a tree-structured search. TSVQ can also be applied as a variable-rate VQ since it can start from the tree root and stop at any intermediate tree level with decreasing distortion and increasing output rate.

TRVQ is an RVQ for which each stage is a TSVQ. TRVQ provides good VQ performance with moderate computational complexity and memory requirement. In our algorithm, a two-stage RVQ with each stage as a 3-level TSVQ is used to quantize the DWT coefficients, hence there are a total of 6 possible VQ encodings, or *coding units*. Each coding unit has 16 vectors in the codebook, producing an output of 4 bits per coding unit.

2.5. Rate-Distortion Trade-offs Based on Zerotrees

In the octave pyramid DWT decomposition, the HL, HH and LH subbands correspond to the horizontal, diagonal, and vertical directions respectively. We call these directional branches *direction trees* as shown in Fig. 3.

In our scheme, the zerotree prediction is implicitly applied in the rate-distortion optimization along each direction tree. The rate-distortion trade-off along a direction tree can be extended from the vector-based λ rate-distortion optimization [10]. The goal of rate-distortion optimization within a vector is to, for an N coding unit VQ, find the optimal number of the VQ coding unit n which yields the minimum cost J given λ :

$$J = \min_n (D(n) + \lambda R(n)), \quad (2)$$

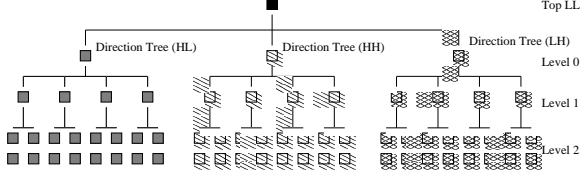


Figure 3: Direction trees in a pyramid hierarchy.

where $D(n)$ is the distortion of n -coding-unit VQ, and

$$R(n) = H(n) + \sum_{i=1}^n B(i) \quad (3)$$

is the bit length (rate) of coded VQ. $H(\cdot)$ is the length of the entropy code used to code the VQ header indicating the coding decision, and $B(i)$ is the length of the VQ indices in the i th coding unit. Similar to the EZW algorithm, we have four symbols for the significance map: SG-significant, RT-zerotree root, IZ-isolated zero, and CH-zerotree child. Note that CH only virtually exists. The significance symbols are embedded in the VQ header by defining the value of the VQ header as:

$$h = \begin{cases} 0 & \text{if } IZ, \\ 1 \sim N & \text{if } SG, \\ N + 1 & \text{if } RT, \end{cases} \quad (4)$$

where N is the total number of VQ coding units. Denote s as the significance of a certain vector in a direction tree. The cost T of the vector will be extended as:

$$T(s) = \begin{cases} J & s = SG \\ D(0) + \lambda H(N + 1) & s = RT \\ D(0) + \lambda H(0) & s = IZ \\ D(0) & s = CH \end{cases} \quad (5)$$

where $D(0)$ is the distortion of the vector without VQ coding (zero coding units). The value of the function $T(\cdot)$ is the actual cost of a vector subject to its significance. Note that $s = CH$ is not applicable to the finest level of the DWT. If we denote $W_{l,p}$ as the sum of the cost of p th zerotree node at l th DWT level and the costs of all its descendants, then it can be expressed recursively as:

$$W_{l,p} = T(s_{(l,p)}) + \sum_{q=4p}^{q < 4p+4} W_{l-1,q} \quad (6)$$

where q is the index of the four siblings belonging to a same parent. Note that l DWT levels consist of 4^l nodes, starting from Level 0, as 1, 4, 16, \dots . For an L level DWT decomposition, the bottom level cost is given by

$$W_{L-1,p} = T(s_{(L-1,p)}). \quad (7)$$

The rate-distortion optimization is applied along the three directional branches, HL, HH and LL, by minimizing $W_{0,0}$ with proper values of $s_{(l,p)}$, or in other words, by proper zerotree classifications within a directional branch. For a pyramid hierarchy of L -level DWT decomposition, this λ -adaptive VQ can be implemented by the following steps:

Step 1: Search for the optimal VQ coding unit number (n) for all the vectors of the directional tree by minimizing the rate-distortion cost of each individual vector. Record the minimal distortions, the optimal VQ coding unit, and the resulting VQ indices. Let $l = L - 1$.

Step 2: Let $p = 0$;

Step 3: Compute $\min_{s_{(l,p)}} (W_{l,p})$. The rule is to compare the three cases of $s_{(l,p)}$: SG, IZ, RT. If $s_{(l,p)} = SG$, then $W_{l,p} = T(SG) + \sum_{q=4p}^{q < 4p+4} W_{l-1,q}$ with the significance of all its descendants unchanged, *i.e.* the second part of 6 unchanged; if $s_{(l,p)} = IZ$, then $W_{l,p} = T(IZ) + \sum_{q=4p}^{q < 4p+4} W_{l-1,q}$ with the significance of all its descendants unchanged; if $s_{(l,p)} = RT$, then compute $W_{l,p} = T(RT) + \sum_{q=4p}^{q < 4p+4} W_{l-1,q}$ by changing the significance of all its descendants as CH. Find the minimal cost among these three cases. If choice of RT is minimal, update W for all its descendants. This is equivalent to the insignificance prediction of EZW in the sense of rate-distortion optimization.

Step 4: Let $p = p + 1$. If $p < 4^l$, go back to Step 3; else go to Step 5.

Step 5: If $l = 0$, quit; else let $l = l - 1$, go back to Step 2.

Once an encoding has been completed, the total number of bits generated is compared to the target, and λ adjusted iteratively to meet the target rate.

3. EXPERIMENT RESULTS AND CONCLUSIONS

The codebook with single-sized vectors for the TRVQ was developed based on 16 training images of the image library at Utah State University. The Huffman codes for the VQ headers and the DPCM-Huffman coder were trained on the same data. These images varied in size from 512x512 to 2048x2048. We note that many wavelet-based VQ techniques use non-squared vector shapes to take advantage of directional characteristics of octave wavelet decomposition [4, 7]. However, in these methods, the size of zerotrees grows less than powers of 4, and will reduce the efficiency of insignificance prediction. In our scheme, this problem is solved by using a single-sized VQ and applying rate-distortion trade-offs along the three directions separately.

The test image 512 \times 512 LENA with 8-bit gray level was used to evaluate the performance of the coding scheme. A PSNR of 30.93 dB was obtained at about 0.174 bpp as shown in Fig. 5, which is apparently better than some published similar approaches [4, 6, 7].

Our experimental results have shown that applying vector quantization on the DWT coefficients greatly outperforms directly applying vector quantization on the image pixels. Fig. 6 illustrates the comparison of rate-distortion curves between our wavelet-based VQ and two other VQ techniques (single-dimension adaptive-search residual VQ and adaptive-search hierarchical VQ) similar to the key frame coding method described in [10].

4. REFERENCES

- [1] P. C. Cosman, R. M. Gray, and M. Vetterli, "Vector quantization of image subbands: A survey," *IEEE Trans. Image Processing*, vol. 5, pp. 202–225, Feb. 1996.



Figure 4: Original LENA image.



Figure 5: LENA encoded by the proposed algorithm with 4 level 9-7 transform, 2x2 2-coding unit 3-level TSVQ, PSNR= 30.93 dB, Compressed rate= 0.174 bpp

- [2] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing*, vol. 41, pp. 3445–3462, Dec. 1993.
- [3] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Dordrecht, Netherlands: Kluwer Academic Publishers, 1992.
- [4] P. C. Cossman, S. M. Perlmuter, and K. O. Perlmuter, "Tree-structured vector quantization with significance map for wavelet image coding," in *Proc. 1995 IEEE Data Compression Conf. (DCC)*, (Snowbird, Utah, USA), Mar. 1995.
- [5] S. M. Perlmuter, K. O. Perlmuter, and P. C. Cossman, "Vector quantization with zerotree significance map for wavelet image coding," in *Proceedings Twenty-ninth Asilomar Conference on Signals, Systems and Computers*, (Pacific Grove, CA), Nov. 1995.
- [6] J. Karlekar, P. G. Poonacha, and U. B. Desai, "Image compression using zerotree and multistage vector quantization," in *Proceedings of the 1997 International Conference on Image Processing*, (Washington, DC, USA), pp. 610–612, Oct. 1997.
- [7] H. Jafarkhani and N. Farvardin, "A scalable wavelet image coding scheme using multi-stage pruned tree structured vector quantization," in *International Conference on Image Processing (3-Volume Set)*, 1995, pp. 81–84, 1995.
- [8] I. Daubechies and W. Sweldens, "Factoring wavelet transform into lifting steps," tech. rep., Bell Laboratories, Lucent Technologies, 1996.
- [9] A. R. Calderbank, I. Daubechies, W. Sweldens, and B.-L. Yeo, "Wavelet transforms that map integers to integers," tech. rep., Dept. of Mathematics, Princeton University, Aug. 1996.
- [10] K. Liang, C.-M. Huang, A. K. Huber, and P. D. Israelsen, "Variable block size multistage VQ for video coding," in *Conference of the IEEE Industrial Electronics Society (IECON)*, Nov. 1999.

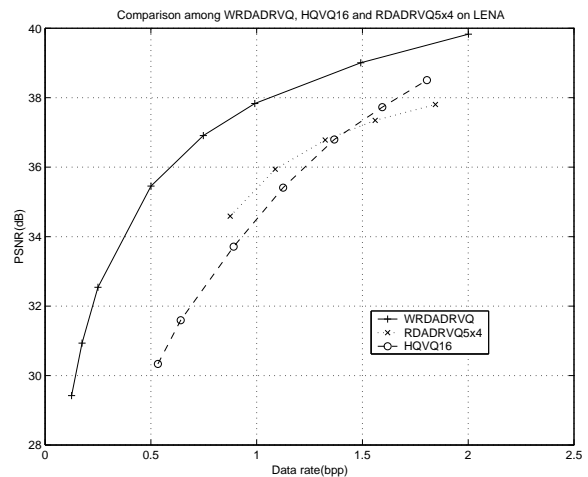


Figure 6: PSNR improvement for proposed algorithm, where WRDADRQ is the wavelet-based method, RDADRQ5x4 is a single blocksize (5x4) method, and HQVQ16 is a hierarchical method.