ALGORITHMIC MODIFICATIONS TO SPIHT

Ulug Bayazit[1] and William A. Pearlman[2]

[1]Isik University, 80670 Maslak, Istanbul, TURKEY bayazit@isikun.edu.tr
[2]Rensselaer Polytechnic Institute, Troy, NY 12180, USA pearlman@ecse.rpi.edu

ABSTRACT

This paper proposes several low complexity algorithmic modifications to the SPIHT (Set Partitioning in Hierarchical Trees) image coding method of [3]. The modifications exploit universal traits common to the real world images. Approximately 1-2 % compression gain (bit rate reduction for a given mean squared error) has been obtained for the images in our test suite by incorporating all of the proposed modifications into SPIHT.

1.INTRODUCTION

SPIHT (Set Partitioning In Hierarchical Trees), presented in [2,3] as an efficient method for lossy and lossless coding of still images, was an advance over the innovative wavelet based image coding method of [1] which employed a tree representation of zeroes of wavelet coefficients for the coding of these coefficients. Zerotree image coding itself yielded several dB's of signal-to-noise-ratio improvement for most real world still images at low bit rates over the DCT (Discrete Cosine Transform) based older JPEG still image coding standard. SPIHT is nonetheless less efficient in coding performance than EBCOT, the wavelet based image coding algorithm presented in [4], which forms the basis of the more recent JPEG-2000 image coding standard. However, SPIHT has dramatically lower computational complexity than EBCOT, due in part to the fact that there is no explicit rate-distortion optimization in SPIHT. Other wavelet based image compression methods such as [5,6] that are competitive with SPIHT on a performance scale have considerably higher complexity than SPIHT as well.

It is a challenge, therefore, to improve the original SPIHT algorithm's coding performance without nsignificantly harming its algorithmic and computational simplicity. The algorithmic modifications to be discussed in this paper are a step towards achieving this goal.

The suggested modifications do not alter the process of uniform quantization of the wavelet coefficients and the resulting quantization values, but alter the way the sets of quantization levels are represented in the bitstream. Therefore the rate-distortion advantage comes in the form of a bit rate reduction for an image compressed for a given fidelity.

Three of the modifications target the tree representation of the sets of coefficients deemed significant with respect to the tested threshold (sorting pass). Only a single modification is proposed for the process of progressive coding of the significant coefficients (refinement pass).

The modification in the refinement pass exploits the nonuniformities in the frequencies of the binary quantization indices output by the successive approximation quantizer. The quantization indices are arithmetic coded ([7]) by employing symbols based on the sign of the (reconstructed) coefficient and the sign of the level output by the successive approximation quantizer.

Two of the other three modifications in the sorting pass also require an implementation with an adaptive arithmetic coder. i) The coefficient signs and significances are jointly arithmetic coded. ii) Different arithmetic models are used to code the significance information of Type B sets (set of spatially oriented descendants of a coefficient excluding their children) rooted at different levels of the spatial orientation tree.

The remaining modification can be implemented without an arithmetic coder. The significance of a Type B set (descendants excluding children) is coded and transmitted prior to the coding and transmission of the significances of the children.

Section 2 reviews the original SPIHT algorithm of [3]. Section 3 presents the modifications in detail. Simulation results demonstrating their effectiveness are presented in Section 4 for several still images. Section 5 presents the concluding remarks.

2. SPIHT (SET PARTITIONING IN HIERARCHICAL TREES) ALGORITHM

As mentioned in [3] the image is wavelet transformed by employing paraunitary filters prior to its encoding with SPIHT. The wavelet coefficients are encoded and transmitted in multiple passes. In each pass only the wavelet coefficients with magnitudes exceeding a certain threshold are encoded. The threshold is computed according to the expression

$$T_p = \frac{c_{\max}}{2^{p+1}} \qquad (1)$$

where $p = 0,1,..., P$ denotes the pass number and $c_{\max} = 2^{\left\lceil \log_2 \max_{(i,j)} |c_{i,j}| \right\rceil}$,

where $c_{i,j}$ is the coefficient at position (i, j) in the image. The significance of a set of coefficients $\Re$ in pass p is indicated as

$$S_p(\Re) = \begin{cases} 1, & \max_{(i,j)\in\Re}\{|c_{i,j}|\} \geq T_p \\ 0, & \text{otherwise} \end{cases}$$

In order to exploit the fact that the energy of an image is concentrated in the low frequency components, the coefficients are ordered in hierarchies, called *spatial orientation trees,* with roots in the lowest frequency subband, branching successively into higher frequency subbands at the same spatial orientation. The set of offspring, $O(i, j)$, of a tree node corresponding to a wavelet coefficient at coordinates (i, j), consists of the wavelet coefficients of the same spatial orientation in the next (finer resolution) level of the pyramid. Except for the nodes at the lowest (finest resolution) level of the pyramid all nodes have 4 offspring. The nodes at the lowest level have no offspring by definition of the spatial orientation tree. The set of all descendants of a node corresponding to a wavelet coefficient at coordinates (i, j), is termed Type A set and denoted by $D(i, j)$. The set of all descendants excluding the offspring of a node corresponding to a wavelet coefficient at coordinates (i, j) is termed a Type B set and denoted by $L(i, j) = D(i, j) - O(i, j)$.

Initially, 2x2 blocks of wavelet coefficients in the lowest frequency subband designate seven sets, four of which are the coefficients themselves. The remaining three sets are the three Type A sets of three of these coefficients (excluding the upper left coefficient of the 2x2 block). The upper left coefficient does not have any offspring or descendants, i.e. the spatial orientation subtree rooted at this node is degenerate. The horizontal, vertical and diagonal neighbors of the upper left coefficient are the roots of the subtrees spanning the spatially oriented coefficients of the horizontal, vertical and diagonally oriented subbands.

The algorithm decomposes each pass into a sorting pass for evaluating coefficient and set significances and a refinement pass for refining the reconstruction values of previously significant coefficients.

In its sorting pass SPIHT partitions the $D(i, j)$ (Type A set) into $L(i, j)$ (Type B set) plus four single coefficient sets $(k,l) \in O(i, j)$ whenever $S_p(D(i, j)) = 1$, and partitions $L(i, j)$ into four sets $D(k,l)$ with $(k,l) \in O(i, j)$ whenever $S_p(L(i, j)) = 1$.

Significance information is stored in three ordered lists, called list of insignificant sets (LIS), list of insignificant pixels (coefficients) (LIP), and list of significant pixels (coefficients) (LSP). During the sorting pass the significance with respect to T_p of the coefficients in the LIP is coded and the coefficients that become significant are moved to the end of the LSP and their signs are coded. Similarly the significance of the descendant sets of Type A and descendant sets of Type B in the LIS are coded and those that become significant are partitioned. The significances with respect to T_p of the four single coefficient sets resulting from partitioning of the Type A sets are coded. The significant single coefficient sets are added to the end of LSP and their signs are coded. Insignificant single coefficient sets are added to the end of LIP. Newly formed Type A and Type B sets are added to the end of LIS to be evaluated again before the same sorting pass ends.

The values of coefficients $c_{i,j}$ in the LSP except the ones included in the last sorting pass are refined in each refinement pass. Specifically a coefficient $c_{i,j}$ is coded by outputing the the p 'th most significant bbit of $| c_{i,j} |$.

In the sorting pass a coefficient deemed to be significant with respect to T_p is reconstructed at the encoder and the decoder as $\pm 1.5 T_p$ depending on the sign of the coefficient. During each refinement pass, the previous reconstruction error is quantized by a two level quantizer and a value of $\pm .5 T_p$ is added to the previous reconstruction value of the coefficient depending on the sign of the quantization level.

Optionally, the significance information of the coefficients in the LIP may be arithmetic coded by keeping blocks of 2×2 coefficients together in the LIP and representing the significance values of the previously insignificant coefficients of each block by a single symbol. Similarly, the significance of trees rooted in blocks of 2×2 coefficients may be arithmetic coded by representing the significance values of the previously insignificant trees of each block by a single symbol.

Since the threshold is halved with each pass, the maximum error in the reconstruction values of significant coefficients at the end of each pass is guaranteed to be halved with respect to that of the previous pass.

3. SUGGESTED MODIFICATIONS TO SPIHT

3.1. Joint Coding of Coefficient Signs and Significances

Adjacent coefficients (of the same subband) in the low-pass filtering direction tend to have positively correlated signs and adjacent coefficients (of the same subband) in the high-pass filtering direction tend to have negatively correlated signs [4]. These correlation properties may be exploited by joint coding the signs as well as significances of a 2x2 block of coefficients.

Each previously insignificant coefficient of the horizontally and vertically oriented subbands (i.e. LH,HL,LLLH,…) can be classified as one of insignificant or positively signed significant or negatively signed significant. The sign/significance classes for the previously insignificant coefficients of the 2x2 block may be mapped by means of a predetermined scan order to a single symbol that is arithmetic coded. For m previously insignificant coefficients in a 2×2 block there can be 3^m symbols. Of these 3^m symbols those corresponding to the above mentioned positive or negative correlations between signs tend to occur more frequently.

The joint coding of coefficient signs and significances is applied only in horizontally and vertically oriented subbands. The coefficient signs are independently coded in the diagonally oriented subbands. The scan order for the horizontally and vertically oriented subbands is depicted in Figure 1.

3.2. Priority of Type B Set Coding

Let us consider a Type A set that has more than 4 coefficients and that has been determined to be significant. In this case the original SPIHT algorithm first codes the significance information of the 4 offspring and then codes the significance information of the Type B set conditional on the number of significant offspring. The rate required for this process may be expressed as

$$R_1 = H(O_1, O_2, O_3, O_4) + H(B \mid O)$$

where $H(O_1, O_2, O_3, O_4)$ is the joint entropy of the offspring significances and $H(B \mid O)$ is the entropy of the Type B set significance conditional on the number of significant offspring, $O = \left| \{ O_i : i \in \{1,2,3,4\}, O_i = 1 \} \right|$.

By noting that $H(B \mid O) \geq H(B \mid O_1, O_2, O_3, O_4)$ the required rate is seen to have the lower bound

$$R_1 \geq H(O_1, O_2, O_3, O_4) + H(B \mid O_1, O_2, O_3, O_4) \qquad (2)$$
$$= H(O_1, O_2, O_3, O_4, B)$$
$$= H(B) + H(O_1, O_2, O_3, O_4 \mid B)$$

As suggested by the last line of Eqn. (2), one can improve the performance of the original algorithm by first unconditionally coding the significance information of the Type B set followed by coding the significance information of the 4 offspring conditional on the significance information of Type B set. Note that this order is preferred to the alternative of first coding the significance information of the 4 offspring and then coding the significance information of the Type B set conditional on the significance information of the 4 offspring (first line of Eqn. (2)) because the number of conditioning states (arithmetic models) in the first case is only two.

3.3. Different Arithmetic Models for Type B Sets Rooted at Different Pyramid Levels

The (conditional) entropy of the symbol representing the significance information of a Type B set depends on the level at which the Type B set is rooted. The histograms showing relative significance/insignificance frequencies of Type B sets rooted at each of the five levels of the spatial orientation tree at the end of 10^{th} pass for the Lenna 512x512 monochrome image are depicted in Figure 2. In a given pass Type B sets rooted at higher levels of the spatial orientation tree are more likely to be significant. This was exploited by employing a different arithmetic model for each of the five levels.

3.4. Exploitation Of The Nonuniformity in the Probability Mass Function Of the Successive Approximation Quantizer by Arithmetic Modelling

We note that the probability mass function of the wavelet coefficients peaks at zero magnitude and declines with magnitude. The probabilities of the output levels of the successive approximation quantizer used in the refinement pass are therefore unequal. In order to take advantage of this, the bi-level output of the successive approximation quantizer may be arithmetic coded. Specifically as shown in Figure 3 if the sign of the reconstructed coefficient is positive, Symbol 0 is coded to refine the reconstruction value by adding $-T_p/2$ and Symbol 1 is coded to refine the reconstruction value by adding $T_p/2$. If the sign of the reconstructed coefficient is negative, Symbol 0 is coded to refine the reconstruction value by adding $T_p/2$ and Symbol 1 is coded to refine the reconstruction value by adding $-T_p/2$. This ensures that the high probability level is coded with Symbol 0 for a negative refinement and the low probability level with Symbol 1 for a positive refinement in magnitude .

Figure 4 shows the frequencies of the lowest magnitude reconstructed coefficients for Lenna512 image at the end of the 10^{th} pass. By comparing the frequencies pairwise it can be seen that the (conditional) entropy of the output index of the successive approximation quantizer depends on the magnitude of the reconstructed coefficients. This dependency has been exploited by employing two arithmetic models for the refinement of the two lowest magnitude, nonzero reconstructed coefficients and another one for the refinement of all the other higher magnitude reconstructed coefficients in each pass.

4. EXPERIMENTAL RESULTS

The performance of the original SPIHT algorithm has been compared with that of the SPIHT algorithm with the four proposed modifications for the coding of several monochrome, 8bpp, 512x512 images. A 5-level pyramidal subband decomposition with 9/7 tap biorthogonal filters and mirror extension at the edges has been employed as in [3]. The bit rates are calculated from the actual size of the compressed files.

Table 1 shows the MSE vs. Rate performance obtained by the original SPIHT algorithm and the effect of incorporating the modifications introduced in Sec. 3.1 thru Sec. 3.4 on the bit rates for the three still images 'Lenna', 'Barbara' and 'Goldhill'. Original 'Lenna' image and its reconstruction after 8 passes with the SPIHT algorithm are shown in Figure 5.

The modifications, which make use of arithmetic coding, sometimes possess a slight performance penalty at very low bit rates due to the increase in the overhead bits for the increased number of arithmetic models. The modification of Section 3.4 yields a distinct performance advantage at all rates for all test images.

5. CONCLUSION

Low complexity modifications that exploit traits common to the real world images have been proposed for improving the performance of the image coding algorithm developed in [3]. Incorporating all modifications into the original algorithm yields approximately 1-2% compression gain at low to moderate rates.

6. REFERENCES

[1] J. M. Shapiro, "Embedded image coding using zeroes of wavelet coefficents," *IEEE Trans. Signal Processing*, vol. 41, pp. 3445-3462, Dec. 1993.

[2] A. Said and W. A. Pearlman, "An image multiresolution representation for lossless and lossy compression," *IEEE Trans. Image Processing*, vol. 5, pp. 1303-1310, Sept. 1996.

[3] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 243-250, June 1996.

[4] D. Taubman, "High Performance Scalable Image Compression with EBCOT," *IEEE Trans. Image Processing*, vol. 9, no. 7, pp.1158-1170, July 2000.

[5] B. A. Banister and T. R. Fischer, "Quadtree Classification and TCQ Image Coding," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 11, no. 1, pp. 3-8, January 2001

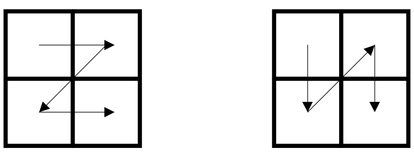
[6] D. Marpe et. al,"A two layer wavelet based algorithm for efficient lossless and lossy image compression," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 10, no. 7, pp. 1094-1102, October 2000.

[7] I. H. Witten, R. M. Neal and J. G. Cleary,"Arithmetic coding for data compression," *Communications ACM*, vol. 30, no.6, pp. 520-540, June 1987.

	MSE	RATE (bpp)				
		Orig	3.1	3.1-3.2	3.1-3.3	3.1-3.4
Lenna	500.40	0.0051	0.0053	0.0054	0.0054	0.0054
	293.9	0.0128	0.0132	0.0134	0.0133	0.0133
	155.0	0.0333	0.0340	0.0342	0.0340	0.0339
	77.39	0.0780	0.0789	0.0788	0.0783	0.0778
	36.79	0.1706	0.1716	0.1708	0.1702	0.1686

	17.65	0.3498	0.3502	0.3491	0.3483	0.3447
	8.463	0.7132	0.7138	0.7126	0.7115	0.7041
Barbara	760.8	0.0054	0.0057	0.0058	0.0057	0.0057
	513.5	0.0143	0.0148	0.0149	0.0148	0.0148
	344.5	0.0430	0.0438	0.0439	0.0438	0.0435
	166.2	0.1582	0.1591	0.1588	0.1586	0.1572
	63.93	0.3751	0.3761	0.3755	0.3747	0.3672
	23.56	0.7234	0.7235	0.7229	0.7216	0.7084
	8.378	1.2530	1.2526	1.2518	1.2505	1.2294
Goldhill	478.0	0.0037	0.0039	0.0040	0.0040	0.0040
	317.4	0.0099	0.0104	0.0105	0.0105	0.0104
	201.7	0.0280	0.0286	0.0288	0.0288	0.0285
	118.2	0.0809	0.0817	0.0814	0.0813	0.0806
	62.77	0.2186	0.2185	0.2176	0.2173	0.2148
	28.82	0.5434	0.5414	0.5399	0.5392	0.5331
	10.93	1.1991	1.1952	1.1933	1.1926	1.1785

Table 1 MSE vs. Rate for original SPIHT and the proposed modifications incorporated one at a time



a) Horizontally oriented subbands   b) Vertically oriented subbands

Figure 1 Scan orders used for mapping coefficient sign/significance classes to a single symbol.

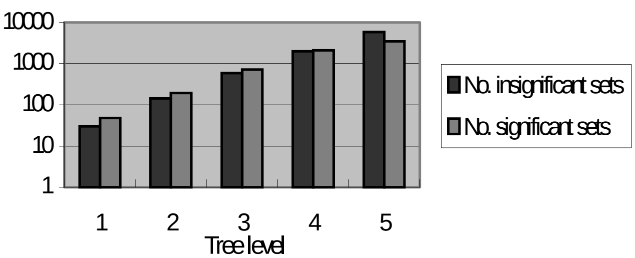


Figure 2 Number of significant and insignificant Type B sets rooted at each of the five levels of the spatial orientation tree at the end of 10 passes on Lenna512. Sets rooted at the highest level (Level 1) are more likely to be significant than those rooted at the lowest (Level 5).

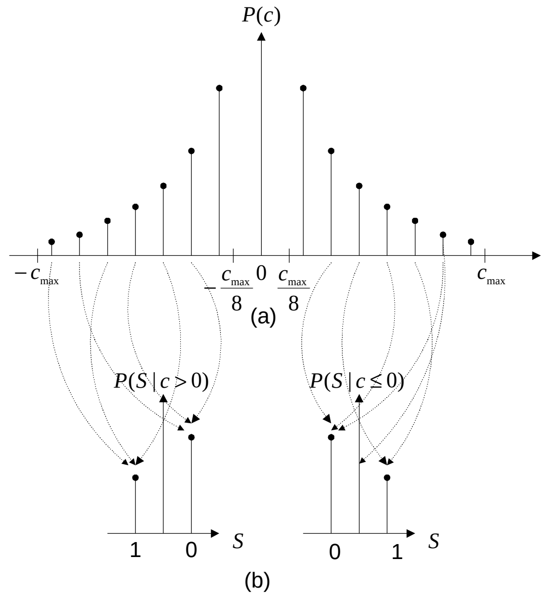


Figure 3 a)Illustration of the probability mass function for the reconstructed coefficients at the end of 3 passes ($p=3$). Probabilities decline with magnitude. b) Symbols are assigned to the output of the bi-level quantizer used in the refinement pass based on the sign of the reconstruction values (S:Symbol). High probability level is coded with S=0 for a negative and low probability one with S=1 for a positive refinement in magnitude .

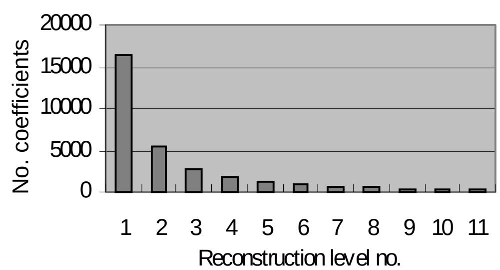


Figure 4 Histogram of the number of coefficients getting mapped to the 11 lowest magnitude nonzero reconstruction levels at the end of 10th pass for Lenna512. (Reconstruction level i has magnitude $(i+0.5)T_p$)

Figure 5 Left: Original Lenna512 Right: Reconstructed Lenna512 after 9 passes (MSE = 17.65, Original SPIHT Rate = 0.3498, SPIHT with all four modifications Rate = 0.3447)