Wavelet tree classification and hybrid coding for image compression

C.-K. Su, H.-C. Hsin and S.-F. Lin

Abstract: A hybrid coding system that uses a combination of set partition in hierarchical trees (SPIHT) and vector quantisation (VQ) for image compression is presented. Here, the wavelet coefficients of the input image are rearranged to form the wavelet trees that are composed of the corresponding wavelet coefficients from all the subbands of the same orientation. A simple tree classifier has been proposed to group wavelet trees into two classes based on the amplitude distribution. Each class of wavelet trees is encoded using an appropriate procedure, specifically either SPIHT or VQ. Experimental results show that advantages obtained by combining the superior coding performance of VQ and efficient cross-subband prediction of SPIHT are appreciable for the compression task, especially for natural images with large portions of textures. For example, the proposed hybrid coding outperforms SPIHT by 0.38 dB in PSNR at 0.5 bpp for the Bridge image, and by 0.74 dB at 0.5 bpp for the Mandrill image.

1 Introduction

With the rapid growth of modern communications applications and computer technologies, image compression was (and still is) increasingly in demand. Three categories of image compression techniques have been developed: differential pulse code modulation, transform coding and subband coding [1–3]. State-of-the-art techniques are able to compress typical images by a factor ranging from 10 to 50 with acceptable quality [4]. The Joint Photographic Experts Group (JPEG) standard is the most widely used transform-coding algorithm. It shows good performances at moderate compression ratios [5]. Recently, the wavelet-based multi-resolution representation has received a lot of attention for compression applications. In wavelet domain, the higher detailed components of images are projected onto the shorter basis functions with higher spatial resolutions, and the lower detailed components are projected onto the larger basis functions with narrower bandwidths. This kind of trade-off between the space and spatial-frequency resolutions matches the characteristics of the human visual system [6]. Many wavelet-based image coders such as embedded zero-tree wavelets (EZW) [7], set partitioning in hierarchical trees (SPIHT) [8], morphological representation of wavelet data (MRWD) [9], group testing for wavelets (GTW) [10] and modulated wavelet subband image coding (MWSIC) [11] have been proposed with a great success.

It is noted that both EZW and SPIHT take advantage of the following hypothesis. If a wavelet coefficient at the coarse resolution is insignificant with respect to a given threshold, then all the corresponding wavelet coefficients at the finer resolutions are likely to be insignificant with respect to the same threshold. Consequently, these insignificant wavelet coefficients can be efficiently coded as a set of insignificant coefficients by using a single code symbol: zero-tree. The SPIHT algorithm has been modified for encoding large images in constrained memory environments [12], which is one of the important requirements in the JPEG 2000 standard [13, 14]. A vector extension of SPIHT, called VSPIHT [15], has been proposed to further improve the coding performance. It groups the wavelet coefficients of greyscale images into vectors and then performs successive refinement vector quantisation in the set partitioning framework. Moreover, for colour images, VSPIHT is able to exploit both the cross-subband dependency of each spectral component and the inter-component redundancy to improve on the scalar SPIHT. For video applications, a 3-D extension of SPIHT has been proposed [16]. It provides many advantages including scalability in both time and space for progressive transmission, precise rate control, and low complexity.

For natural images with textures composed mainly of the middle and high frequency components, however, zero trees of insignificant wavelet coefficients are rare, and therefore the compression performances of both EZW and SPIHT are usually unacceptable. A different, more appropriate coding strategy is therefore desirable for the aforesaid images, where the insignificant wavelet coefficients are found to be scattered as well as isolated in the tree representation. In this paper, a simple tree classifier is proposed to group the wavelet trees of images into two classes based on the amplitude distributions; and a hybrid, image coding system is thus developed by coding each class of trees with appropriate procedure to improve the compression performance.

IEE Proceedings online no. 20050004

doi: 10.1049/ip-vis:20050004

Paper first received 5th January and in revised form 4th May 2005

C.-K. Su and S.-F. Lin are with the Department of Electrical and Control Engineering, National Chiao Tung University, Hsinchu, Taiwan

H.-C. Hsin is with the Department of Computer Science and Information Engineering, National Formosa University, Yun-Lin, Taiwan

E-mail: cks@chu.edu.tw

2 Overview of wavelet transform and SPIHT

Wavelet transform is well known as a multiresolution analysis that provides many advantages: joint space-spatial frequency localisation, clustered wavelet coefficients of significance with strong correlations between subbands, and exact reconstruction, which are truly beneficial to image compression. Discrete wavelet transform (DWT) decomposes a signal: $S_\ell(n)$ at resolution ℓ into two components:

$$S_{\ell+1}(n) = \sum_k S_\ell(k)\ h(2n-k)$$
$$D_{\ell+1}(n) = \sum_k S_\ell(k)\ g(2n-k) \qquad (1)$$

where $S_{\ell+1}(n)$ is its approximation at the next coarser resolution $\ell+1$, $D_{\ell+1}(n)$ is the detail information between the two successive resolutions: ℓ and $\ell+1$, $h(n) = <\phi, \phi_{-1,-n}>$, $g(n) = <\psi, \phi_{-1,-n}>$, $<\cdot,\cdot>$ is an inner product operator, ψ is a valid (mother) wavelet, ϕ is the scaling function, and $\phi_{-1,-n}(x) = 2^{-1/2}\phi(2^{-1}x - n)$. The original signal $S_\ell(n)$ can be exactly reconstructed from $S_{\ell+1}(n)$ and $D_{\ell+1}(n)$ by using the following inverse DWT (IDWT):

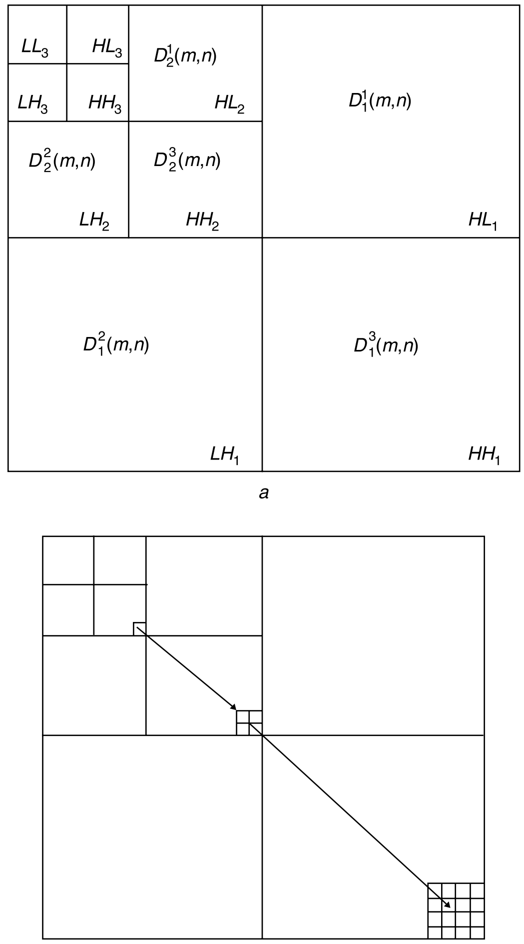
$$S_\ell(n) = \sum_k S_{\ell+1}(k)\tilde{h}(n-2k) + \sum_k D_{\ell+1}(k)\tilde{g}(n-2k) \quad (2)$$

where $\tilde{h}(n) = h(-n)$ and $\tilde{g}(n) = g(-n)$.

For image applications, the two-dimensional DWT can be obtained by using the tensor product of two one-dimensional DWT, i.e. the row processing followed by column processing, or vice versa. Figure 1*a* shows a 3-level, 2-D DWT in a pyramid structure. HL_ℓ, LH_ℓ and HH_ℓ are the wavelet subbands composed of the wavelet coefficients $D_\ell^1(m,n)$, $D_\ell^2(m,n)$ and $D_\ell^3(m,n)$, representing the detail information at resolution ℓ in the horizontal, vertical and diagonal directions, respectively. LL_3 is composed of the scaling coefficients $S_3(m,n)$ representing the approximation at the coarsest resolution 3, and the original image is usually considered the scaling coefficients $S_0(m,n)$ at the finest resolution 0. $S_\ell(m,n)$ can be decomposed into $S_{\ell+1}(m,n)$, $D_{\ell+1}^1(m,n)$, $D_{\ell+1}^2(m,n)$ and $D_{\ell+1}^3(m,n)$ by using the 2-D DWT. And, the 2-D IDWT obtained by using the tensor product of two 1-D IDWT exactly reconstructs $S_\ell(m,n)$ from $S_{\ell+1}(m,n)$, $D_{\ell+1}^1(m,n)$, $D_{\ell+1}^2(m,n)$ and $D_{\ell+1}^3(m,n)$.

In wavelet domain, an image is decomposed into subbands with orientation selectivity. Wavelet coefficients taken from all the subbands of the same orientation are rearranged to form the wavelet trees. The tree hierarchy is based on the resolution level. The wavelet coefficients at coarse resolution are called parent nodes, each of which has four children nodes at the next finer resolution. Tree roots are at the coarsest resolution, and tree leaves are at the finest resolution. Figure 1*b* shows a wavelet tree in the diagonal direction. Many natural images are composed of large portions of homogeneous regions, textures, together with a small portion of edges, which are typically the low, middle and high frequency components, respectively. The significant wavelet coefficients of the homogeneous regions are usually at the coarser resolutions, i.e. in the lower frequency subbands, while those near the noticeable edges are usually clustered in the higher frequency subbands with strong similarities across subbands. If a non-leave node is insignificant, then all the descendants at the finer resolutions are likely to be insignificant. This cross-subband dependency of wavelet coefficients can be exploited to improve the image compression performance.

The SPIHT algorithm has received a lot of attention since its introduction for image compression in 1996. It contains two passes: the sorting pass and refinement pass, which can



b

Fig. 1 *Example of 3-level, 2-D DWT and a wavelet tree in the diagonal direction*

a 3-level, 2-D DWT with subbands delimited by thick lines
b A wavelet tree in the diagonal direction

be combined to form a single scan pass. Three symbols: zero tree (ZT), insignificant pixel (IP) and significant pixel (SP) are used to code the wavelet tree structure of images, which are stored in their respective lists: list of insignificant sets (LIS), list of insignificant pixels (LIP) and list of significant pixels (LSP). Below is the encoding algorithm presented in four steps [8].

(1) Compute $b = \lfloor \mathbf{log}_2(\mathbf{max}_{(m,n)}|c_{m,n}|)\rfloor$, where $c_{m,n}$ is the wavelet tree node at coordinate (m,n). Set the initial threshold $T = 2^b$.
(2) Sorting pass: identify the coefficients such that $T \leqslant |c_{m,n}| < 2T$; output their respective coordinates and signs.
(3) Refinement pass: output the b^{th} (most significant) bit of all the tree nodes with $|c_{m,n}| \geqslant 2T$ following the same order used to output the coordinates in previous sorting passes.
(4) Decrease b by one, halve the threshold T and go to Step 2.

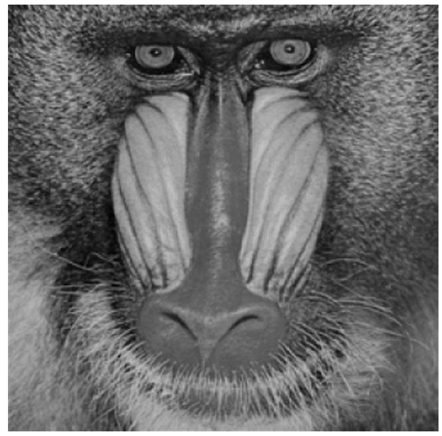
The scan pass (i.e. Step 2 followed by Step 3) of SPIHT is performed in a recursive manner until the expected bit rate is reached. In the sorting pass, the coefficients in LIS and LIP are evaluated as follows. For coefficients whose magnitudes are greater than or equal to the current threshold, they become significant and will be moved to LSP. For insignificant coefficients whose magnitudes are less than the current threshold, they will be stored in LIS if all their descendants are also insignificant with respect to the same threshold; or otherwise, stored in LIP. A sequence of successively smaller thresholds can be obtained by using the following recursive equation:

$$T_k = 0.5T_{k-1} \tag{3}$$

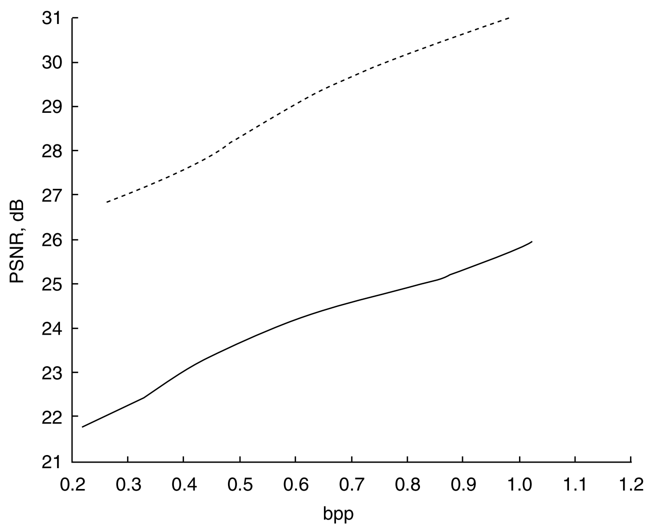
where the initial threshold T_1 must be greater than or equal to half the maximum magnitude of the transform coefficients. After the *kth* sorting pass, tree nodes whose magnitudes are in the range: $[T_k, T_{k-1})$ for $k>1$ (or $[T_1, \infty)$ for $k=1$) will be stored in LSP with one bit per node to indicate their respective signs. In refinement pass, the significant nodes stored in LSP are refined with one bit per node to update their respective information. The great success of SPIHT is attributed to the important hypothesis of wavelet transform: if a parent node is insignificant, then all its descendants are likely to be insignificant with respect to the same threshold and therefore these insignificant nodes can be efficiently coded with a single symbol ZT.

3 Proposed hybrid coding

For images with textures composed mainly of the middle and high frequency components, there are many significant nodes whose ancestors are insignificant. It follows that zero trees of insignificant nodes are very rare. Figure 2*a*, for example shows a 256×256 greyscale Mandrill image with large portions of high frequency textures. Empirically, we have classified the wavelet trees into two classes based on the magnitude distribution. The compression performance of SPIHT is evaluated for each class of wavelet trees. As shown in Fig. 2*b*, where the horizontal and vertical axes are the compression rates measured in bits per pixel (bpp) and peak-signal-to-noise-ratio (PSNR) values measured in dB, respectively, the SPIHT algorithm is much more effective for one class of wavelet trees than the other.



a



b

Fig. 2 *Rate-distortion curves of the low frequency (dotted line) and high frequency (solid line) wavelet trees of Mandrill image by using the SPIHT algorithm*

a 256×256 greyscale image
b Rate - distortion curves

3.1 Wavelet tree classification

High quality image compression at low bit rates can be achieved by coding each individual wavelet tree using a distinct, suitable procedure. It is noted that the SPIHT algorithm will not be suitable for coding wavelet trees with a large amount of significant nodes scattered in the higher frequency subbands, and therefore a different coding strategy is desirable. Moreover, a tree classifier that can efficiently divide the wavelet trees of images into two classes based on the magnitude distribution of the dominant wavelet coefficients is required.

For computation simplicity, a tree classifier based on the average magnitude of wavelet coefficients of each subband has been utilised to divide the wavelet trees of images into two classes: low frequency tree and high frequency tree, which is given as follows:

$$Arg\left\{\max_{\ell} \quad \alpha_\ell \cdot Avg\left\{\left|D_\ell^d(m,n)\right|\right\}\right\} < L_{tr} \quad \rightarrow \quad \text{high frequency tree} \tag{4}$$

where $\left|D_\ell^d(m,n)\right|$ is the wavelet coefficient magnitude at tree node coordinate (m,n), $\ell = 1, 2, \ldots, L$ denotes the resolution level with larger meaning coarser, L is the number of decomposition levels, $d = 1, 2, 3$ denotes the wavelet subband orientation in the horizontal, vertical and diagonal directions, respectively, α_ℓ is a weighting factor with respect to the resolution level, L_{tr} is a threshold value, and $Avg\{\cdot\}$ is an average operator.

3.2 MVQ coding for high frequency wavelet trees

Even though it is noted that wavelet transform provides decorrelation property, i.e. most of the correlation between image pixels can be removed in the wavelet domain, there may still be some residual correlation between neighbouring coefficients across subbands of the same orientation. In order to get a good quality of the reconstructed images at relatively low bit rates, the residual correlation between wavelet coefficients must be exploited. According to Shannon's theory, VQ can significantly reduce the coding bits of signals over scalar quantisation. The VQ approach is therefore suitable for coding the high frequency wavelet trees of images.

Our strategy is as follows. First, group the high frequency wavelet trees into three categories according to their respective subband orientations: horizontal, vertical or diagonal. Second, partition each category of high frequency wavelet trees into small vectors based on the standard deviation distribution. Third, encode the small vectors of high frequency wavelet trees by using multistage VQ (MVQ). Figure 3 shows the MVQ structure with

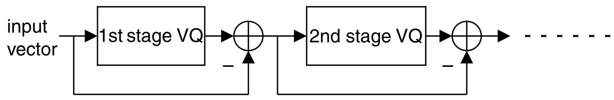


Fig. 3 *Multistage VQ structure*

successive refinements. Here, the input vector is quantised at the first stage, and the residual information is quantised at the following stages in a recursive manner.

3.3 MVQ codebook generation

A representative collection of images is utilised as training images for codebook generation. After 2-D DWT, the high frequency wavelet trees will be partitioned into small vectors to alleviate the computation complexity. The partitions chosen for each of the three categories of high frequency trees are determined in such a manner that tree nodes that have similar standard deviations are grouped into a single vector. Thereafter, a unique codebook is constructed for each vector because the intrinsic statistics and dimensions of vectors are different.

By taking into account that one of the key issues of the proposed hybrid image coding system, which is presented in Section 3.4, is the bit allocation between two different coding procedures, the codebook size for each MVQ stage is 2. In other words, each vector will be coded in a progressive manner by using MVQ with one code bit per stage. The MVQ codebooks are constructed by using the LBG algorithm [17], stored in tables on both encoder and decoder sides, and therefore not transmitted along with the bit stream header.

3.4 Hybrid image coding

After wavelet tree classification, the low frequency trees can be coded efficiently by SPIHT, and the high frequency trees are to be coded by MVQ. A hybrid coding system that combines SPIHT and MVQ is then proposed to improve the overall compression performance. Figure 4 shows the block diagram. The input image is decomposed into a set of subbands with orientation selectivity using 2-D DWT. The scaling coefficients at the coarsest resolution are coded by using the differential pulse code modulation algorithm. The corresponding wavelet coefficients taken from all the subbands of the same orientation (i.e. horizontal, vertical or diagonal) are rearranged to form wavelet trees. While SPIHT coding is suitable only for wavelet trees with a large amount of significant nodes in the lower frequency subbands, the MVQ approach seems promising for coding

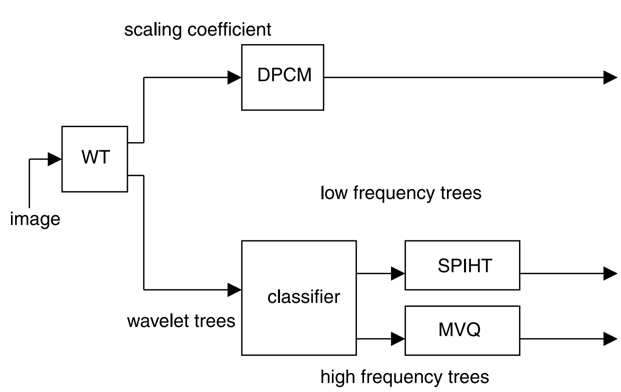


Fig. 4 *Block diagram of the proposed hybrid image coder by combining SPIHT and MVQ for coding the low and high frequency wavelet trees, respectively*

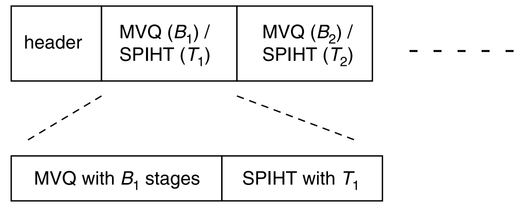


Fig. 5 *Bitstream structure*

wavelet trees with many nodes of significance in the higher frequency subbands.

In MVQ, the high frequency wavelet trees are partitioned into small vectors. These partitions have been determined previously in the training process for codebook generation. Initially, each vector is coded progressively using the MVQ algorithm with B_1 stages, which is determined in such a manner that the norm of the residual quantisation error will not be greater than the initial threshold T_1 of SPIHT. After one MVQ pass followed by one scan pass of SPIHT (for coding the high frequency wavelet trees and low frequency wavelet trees, respectively), the threshold of SPIHT is halved, and based on which the corresponding parameter of MVQ is determined similarly for the next MVQ pass. The alternate coding of high frequency wavelet trees and low frequency wavelet trees will go on until the expected bit rate (or the quality of the reconstructed image) is reached.

Bit allocation between the SPIHT and MVQ coding procedures needs to be truly adaptive in order to generate an improved, embedded bit stream. During hybrid coding, both sequences of SPIHT thresholds, T_k, and MVQ parameters, B_k, are adapted to the input image since T_k is determined by the wavelet coefficient magnitudes of the input image, and B_k is determined by T_k. In our experiments, the parameter B_k is often about 1 or 2 for $k = 2, 3, \ldots$ if the sequence of successively smaller thresholds T_k is obtained by using (3), therefore, they can be set constant. The side information that is required to be transmitted along with the bit stream will be the number of wavelet decomposition levels, the tree classification threshold, one bit per wavelet tree to indicate the tree class, the initial SPIHT threshold T_1, and the MVQ parameter B_1. Figure 5 shows the embedded bit stream structure, where the side information is stored in the header portion.

4 Experimental results

The proposed hybrid coding system is evaluated on natural 256×256 greyscale images. A set of nine training images is utilised to determine the partitions of the high frequency wavelet trees into small vectors, and to construct the MVQ codebooks for encoding these vectors. The partition strategy is as follows. For each of the three categories of high frequency wavelet trees, the standard deviation values are uniformly quantised with nine quantisation levels. All the wavelet coefficients that have the same standard deviation level are grouped into a single vector. Consequently, the high frequency wavelet trees are partitioned into nine small vectors. The test 256×256 greyscale images; Mandrill (shown in Fig. 2*a*), Bridge, and Lena, which represent natural images with a large amount of high frequency, middle frequency, and low frequency components, respectively, are outside the training set.

The compression performance is compared with the SPIHT coding algorithm. The compression rate is measured in bpp. The distortion is measured by peak-signal-to-noise-ratio (PSNR), which is given by

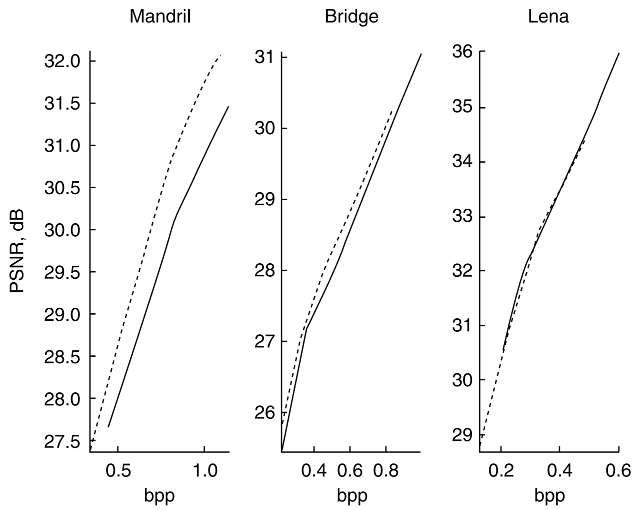


Fig. 6 *Rate-distortion curves of the test images Mandrill, Bridge and Lena by using the proposed hybrid coder (dotted lines) and SPIHT (solid lines)*

$$\mathrm{PSNR}[dB] = 20\mathbf{log}_{10}\frac{255}{RMSE} \tag{5}$$

where *RMSE* is the root mean squared error between the original and reconstructed images. The computed compression rates and PSNR values are collected to generate the rate distortion curves. The linear phase, biorthogonal wavelet with 9/7-coefficient filter set is utilised. The number of wavelet decomposition levels is 4. The tree classification threshold L_{tr} is 3. The weighting factors are empirically obtained by $\alpha_{\ell+1} = 0.5\alpha_\ell$ with $\alpha_1 = 1$. The infinite norm is used to compute the norms of the residual quantisation error vectors. The maximum wavelet coefficient magnitude is halved and then used for the initial SPIHT threshold T_1, and the successively smaller thresholds are obtained by using (3).

Figure 6 shows the rate-distortion curves. The horizontal and vertical axes are the compression rates (in bpp) and PSNR values (in dB), respectively. For the Lena image that is relatively smooth and most of the significant wavelet coefficients are in the lower frequency subbands, only a small number of wavelet trees are classified into the high frequency class. The performances of the hybrid coder and SPIHT coder are comparable, as expected. For the texture-rich images, e.g. Bridge and Mandrill that contain a large amount of significant wavelet coefficients in the middle and high frequency subbands, the hybrid coder is superior to the SPIHT algorithm in general. As indicated by the simulation results, the hybrid coder outperforms SPIHT by 0.38 dB at 0.5 bpp for the Bridge image and by 0.74 dB at 0.5 bpp for the Mandrill image.

It has been shown that when the textured images are encoded, 2-D DWT is unlikely to yield many large zero trees owing to lack of homogeneous regions. Thus, the advantage of encoding zero trees of insignificant wavelet coefficients by using SPIHT is weakened. Alternatively, the high frequency wavelet trees can be efficiently sought out by using the proposed tree classifier, and then can be encoded by using MVQ to improve the overall compression performance.

5 Conclusion

Wavelet transform provides an efficient multi-resolution analysis. It decomposes images into subbands with orientation selectivity in addition to joint space-frequency localisation. Many efficient wavelet coders, e.g. EZW and its improved version SPIHT have been developed by taking advantage of the following hypothesis: if a wavelet coefficient is insignificant with respect to a given threshold, then all the corresponding wavelet coefficients at the finer resolutions are likely to be insignificant with respect to the same threshold. However, natural images are often composed of textures with rapid variations in greyscale. For such images, there are many significant wavelet coefficients scattered in the higher frequency subbands. Consequently, the coding performances of both EZW and SPIHT are usually not adequate since groups of insignificant wavelet coefficients in the tree structure are very rare. A hybrid image coder by combining SPIHT and MVQ with their respective advantages is proposed, in which trees with a large amount of significant wavelet coefficients in the higher frequency subbands are to be identified by using a simple tree classifier, and then coded by using a different, more suitable method instead of SPIHT. Experimental results show that the proposed hybrid coding is superior to SPIHT coding for images with textures composed of the middle and high frequency components. It improves the overall compression performance at the cost of additional computations, i.e. the computation of (4) for classifying wavelet trees into two classes, and the computation of norms of MVQ error vectors for determining the MVQ parameters B_k.

6 Acknowledgments

The authors would like to thank the reviewers for their constructive comments. The National Science Council of Taiwan, under Grant NSC 93-2213-E-216-007, supported this work.

7 References

1 Musmann, H.G., Pirsch, P., and Grallert, H.J.: 'Advances in picture coding', *Proc. IEEE*, 1985, **73**, pp. 523–548
2 Clarke, R.J.: 'Transform coding of images' (Academic Press, New York, USA, 1985)
3 Kwon, O.J., and Chellappa, R.: 'Region adaptive subband image coding', *IEEE Trans. Image Process.*, 1988, **7**, (5), pp. 632–648
4 Rao, K.R., and Hwang, J.J.: 'Techniques and standards for image video and audio coding' (Prentice Hall, Upper Saddle River, NJ, USA, 1996)
5 Pennebaker, W.B., and Mitchell, J.L.: 'JPEG still image data compression standards' (Van Nostrand, New York, USA, 1993)
6 Strang, G., and Nguyen, T.: 'Wavelets and filter banks' (Wellesley-Cambridge, Wellesley, MA, USA, 1996)
7 Shapiro, J.M.: 'Embedded image coding using zero-trees of wavelet coefficients', *IEEE Trans. Signal Process.*, 1993, **40**, pp. 3445–3462
8 Said, A., and Pearlman, W.A.: 'A new, fast, and efficient image codec based on set partitioning in hierarchical trees', *IEEE Trans. Circuits Syst. Video Technol*, 1996, **6**, pp. 243–250
9 Servetto, S.D., Ramchandran, K., and Orchard, M.T.: 'Image coding based on a morphological representation of wavelet data', *IEEE Trans. Image Process.*, 1999, **8**, pp. 1161–1174
10 Hong, E.S., and Ladner, R.E.: 'Group testing for image compression', *IEEE Trans. Image Process.*, 2002, **11**, pp. 901–911
11 Hsin, H.-C.: 'Adaptive modulated wavelet subband image coding', *Pattern Recognit. Lett.*, in press
12 Frederick, F.W., and Pearlman, W.A.: 'Low-memory packetized SPIHT image compression', Asilomar Conf. *Signals Syst. Comput.*, 1999, **2**, pp. 1193–1197
13 JPEG 2000 Image Coding System, ISO/IEC CD15444-1: 1999 (Version 1.0)
14 Usevitch, B.E.: 'A tutorial on modern lossy wavelet image compression: foundations of JPEG 2000', *IEEE Signal Process. Mag.*, 2001, **18**, (5), pp. 22–35
15 Mukherjee, D., and Mitra, S.K.: 'Arithmetic coded vector SPIHT with classified tree-multistage VQ for color image coding'. IEEE Workshop on Multimedia Signal Processing, Los Angeles, California, Dec. 1998, pp. 444–449
16 Kim, B.-J., Xiong, Z., and Pearlman, W.A.: 'Low bit-rate scalable video coding with 3-D set partitioning in hierarchical trees (3-D SPIHT)', *IEEE Trans. Circuits Syst. Video Technol.*, 2000, **10**, (8), pp. 1374–1387
17 Linde, Y., Buzo, A., and Gray, R.: 'An algorithm for vector quantizer design', *IEEE Trans. Commun.*, 1980, **28**, (1), pp. 84–95