A CLASSIFICATION METHOD FOR ADAPTIVE TRANSFORM IMAGE CODING

Jianping Pan
Wireless Systems, MS: 506-108
Rockwell Telecommunications
Newport Beach, CA 92658-8902, USA
email: panj@nb.rockwell.com

ABSTRACT

This paper introduces a new classification method, termed vector-scalar classification, where each DCT (ac) vector is indexed using vector quantization (VQ) on the energy vector of the corresponding DCT subvectors, and the subvectors each are then classified according to the class assignment for the VQ. Vector-scalar classification achieves an average classification gain about 6.5 dB higher than block classification proposed by Chen and Smith [1], since VQ has high clustering performance and classification of subvectors can exploit compactness of DCT coefficients more efficiently. The complexity of this method is relatively low. For coding small images, a fixed-rate DCT coding system developed using vector-scalar classification is competitive with some best variable-rate coding systems reported in the literature, and the classification can capture most of the performance gain obtained by entropy coding.

1. INTRODUCTION

Classification is widely used for adaptive image coding in discrete cosine transform (DCT), subband, and wavelet transform coding systems. Chen and Smith [1] proposed a classification method for DCT coding. In this method, each DCT block is classified into one of the four equally populated classes based on its ac energy, and an adaptive quantizer is used in each class being coded. Woods and O'Neil [2] applied a classification method (similar to [1]) to subband coding. Tse [3] improved the classification performance by having different-size classes. Joshi, Fischer and Bamberger [4] recently investigated this type of classification techniques. They optimized the classification method by maximizing classification gain [4], and applied this technique to subband coding. Jafarkhani, Farvardin and Lee [5] modified the Chen-Smith classification method for the discrete wavelet transform coding of images.

The Chen-Smith classification method [1] and its modified versions mentioned above belong to block classification; a whole block of DCT coefficients or data in each subband is classified into one of predesigned classes according to block energy. Four classes are usually used. The simple formulation of block classification actually restrains its classification performance. Mohdyusof and Fischer's simulation results [6] showed that the four-class block classification for DCT image coding can only achieve minimal improvement of the peak signal-to-noise ratio (PSNR) in their entropy-coded lattice vector quantization systems. Entropy coding is often needed to exploit most of the advantage of non-stationary nature of images in the variable-rate coding systems with block classification.

This paper introduces vector-scalar classification for DCT-based image coding. It is a great improvement over the Chen-Smith block classification method [1]. A fixed-rate DCT coding system is developed using this classification. For coding small-size (256×256) images, the fixed-rate system is superior or comparable to some best variable-rate coding systems found in the literature, and the vector-scalar classification achieves a performance gain similar to that obtained by variable-rate entropy coding.

This paper is organized as follows. The vector-scalar classification method for DCT-based image coding is described in Section 2. A fixed-rate trellis-coded quantization DCT image coding system using the classification is presented in Section 3. Simulation results and PSNR comparisons are then provided in Section 4. Finally, conclusions are made in Section 5.

2. VECTOR-SCALAR CLASSIFICATION

The classification in DCT image coding is to track the local activity of imagery. It seems difficult to meet this objective by using a simple block classification like in [1], since there are a variety of local characteristics in natural images, e.g., edges, texture, smooth areas, etc. [7]. One better way is to perform classification on the basis of DCT subvectors rather than DCT vectors. This is the basic idea of the vector-scalar classification presented here.

2.1. Classification

The DCT transform is applied to 16×16 blocks of images. A simple edge detection, which can be modified from Ho and Gersho's method [8], is used to divide each DCT block into one of the non-edge, horizontal-edge, vertical-edge, and diagonal-edge blocks. For the n-th block, a DCT (ac) vector $\mathbf{f}_n = \{f_1, f_2, ..., f_{255}\}$ is formed by zigzag-scanning or directionally scanning the ac coefficients according to the result of the edge detection. 2 bits/block side information is needed for the edge detection, which results in a better compactness of coefficients in the DCT vectors. The n-th DCT vector $\mathbf{f}_n$ is then divided into subvectors, $\mathbf{x}_{n1}$, $\mathbf{x}_{n2}$, ..., $\mathbf{x}_{nL}$, where $\mathbf{x}_{nl} = \{x_{nl1}, x_{nl2}, x_{nl3}, ...\}$, that can be of different dimensions. For example, sixteen 15-dimensional subvectors

Proceedings of the 1995 International Conference on Image Processing (ICIP '95)
0-8186-7310-9/95 $10.00 © 1995 IEEE

have been constructed from the lower-order coefficients of each DCT vector (the remaining higher-order DCT coefficients are truncated due to their lower energies). Energy of each subvector is then calculated. A 16-dimensional energy vector, $\mathbf{e}_n = \{e_{nl}, l = 1, 2, ..., L\}$, consists of these energy elements.

A vector quantizer (VQ) is designed using the Linde, Buzo and Gray (LBG) design algorithm [9] for energy vectors, based on a large training database of images. The 6-bit codebook has been used in this VQ. The codewords in the VQ codebook are denoted as $\mathbf{w}_k$, $k = 1, 2, ..., K$, and $\mathbf{w}_k = \{w_{kl}, l = 1, 2, ..., L\}$. The first stage of this classification is to index (or cluster) the DCT vector $\mathbf{f}_n$ by encoding the corresponding energy vector $\mathbf{e}_n$ using the VQ. This is accomplished by searching an optimal VQ codeword that has a minimum squared-error distance to the energy vector. The index of the n-th DCT vector is then

$$k^*(n) = \arg\min_k ||\mathbf{e}_n - \mathbf{w}_k||^2. \tag{1}$$

An optimal scalar quantizer (SQ) [10] is designed for the components of the VQ codewords, w_{kl}, $k = 1, 2, ..., K$, and $l = 1, 2, ..., L$. The elements of the SQ are arranged in the decreasing order of values (energies), $s_1 > s_2 > ... > s_m > ... > s_M$. The size M of the SQ can be chosen differently from or as the same as that of the VQ codebook. Thus, each component of the VQ codewords has a SQ index, $m(w_{kl})$, and many of the components may share the same SQ index. Each element s_m of the SQ actually represents the 'centroid' of a class in the domain of subvector energy. There are M classes available for classifying the DCT subvectors. At the second stage of the classification, each subvector of the n-th DCT vector, $\mathbf{x}_{nl}$, is automatically mapped into one of the M classes, indicated by $m(w_{k^*(n),l})$, according to the index of the optimal VQ codeword and the SQ index assignment of the components of the VQ codewords.

The sequence in each class is constructed by collecting all the DCT subvectors belonging to the class. The coefficients in each class have similar statistical properties although they are from different DCT subvectors and different blocks. The sequences are varying in size. For the ease of optimal bit rate allocation and quantization, these sequences can be made the same length, for example, of 128 or 256 coefficients, along with a slight influence on encoding efficiency. One simple way is to concatenate all the sequences, from a highest-energy sequence to a lowest-energy one, into a long sequence, and then divide it into equal-length sequences. The sequences being coded can be either the energy-decreasing equal-length sequences or the classified varying-length sequences.

2.2. Classification Gain

Let the total coding bit rate be R (bits per sample) and the side information bit rate for classification be R_s, then the bit rate for the classified sources is $R_c = R - R_s$. The classification gain [4] is

$$G_c = \frac{\epsilon_x^2 \sigma_x^2 2^{-2R}}{\prod_{j=1}^{J} (\epsilon_j^2 \sigma_j^2)^{p_j} 2^{-2R_c}}. \tag{2}$$

where, σ_x^2 is the variance of source X, and σ_j^2, $j = 1, 2, ..., J$, is the variance of source X_j in the j-th class. p_j is the probability that any given sample belongs to source X_j. The factors ϵ_x and ϵ_j, $j = 1, 2, ..., J$, depend on densities of the sources and coding schemes, and $\epsilon_x = \epsilon_j$, $j = 1, 2, ..., J$, can be reasonably assumed, especially for the comparison below between the vector-scalar classification and the Chen-Smith classification [1] (since we assume the same coding scheme is applied to the coding systems that use these two classification methods).

IMAGES	CLASSIFICATION METHODS	
	BC-CS	VSC
Lenna	3.87	9.90
Boat	3.86	11.29
Photographer	9.84	12.95
Fruits	3.08	13.16
Average (20 images)	4.57	11.25

Table 1: Classification gain (in dB). BC-CS indicates the Chen-Smith block classification method, and VSC the vector-scalar classification method.

Table 1 reports the classification gains of the Chen-Smith classification method [1] and vector-scalar classification method. 16 classes are used in the Chen-Smith classification method, and side information rate is 4 bits/block. 64 classes are used in the vector-scalar classification method, and an edge detector described above is also utilized prior to the classification, so side information rate is (6+2) bits/block. The total bit rate used for coding images is 0.5 bits/pixel for all the cases in the table.

The vector-scalar classification has an average classification gain about 6.5 dB higher than the Chen-Smith block classification. The block classification for 16 classes and 64 classes is almost the same and the classification for 4 classes has a lower gain. There are only slight differences of the classification gain under different testing conditions of total coding bit rates since side information rate is relatively small. In an image coding system, how to exploit the classification gain depends on an adaptive encoding scheme, but more gain in the classification stage will likely lead to a higher total encoding gain in most cases. Hence the improvement of the classification gain over the basic block classification method [1] demonstrates the advantage of the vector-scalar classification formulation. The complexity of this classification is relatively low (the main computation is the 6-bit VQ encoding of the 16-dimensional energy vectors in our cases).

3. A TRELLIS-CODED QUANTIZATION CODING SYSTEM USING VECTOR-SCALAR CLASSIFICATION

Optimum bit rate allocation is needed for adaptive quantization, and is performed over the sequences in the classes or the re-organized equal-length sequences. Assume the length of each sequence being coded is L_j, $j = 1, 2, ..., J$, where J is the number of the sequences being coded (equal to M if they are the sequences in the classes). Denote R_j as the bit rate for the j-th sequence, then

$$\sum_{j=1}^{J} L_j R_j = L_T R_0, \tag{3}$$

where R_0 is the total bit rate for coding these sequences and $L_T = \sum_{j=1}^{J} L_j$. Assume the distortion-rate function for encoding source X_j is [10]

$$D_j(R_j) = \epsilon_j^2 \sigma_j^2 2^{-2R_j}, \tag{4}$$

then the total distortion is

$$D = \frac{1}{L_T} \sum_{j=1}^{J} L_j \epsilon_j^2 \sigma_j^2 2^{-2R_j}. \tag{5}$$

Using Lagrange multiplier techniques to minimize (5) subject to (3) (assuming $R_j \geq 0$, $j = 1, 2, ..., J$) results in the optimal rates

$$R_j = \frac{1}{2} \log_2 \epsilon_j^2 \sigma_j^2 - \frac{1}{2L_T} \sum_{i=1}^{J} L_i \log_2 \epsilon_i^2 \sigma_i^2 + R_0. \tag{6}$$

There are two reasons for using trellis-coded quantization (TCQ) [11] in this image coding system. First, the vector-scalar classification method has provided the long sequences being coded. Compared to vector quantizers and multistage vector quantizers that are widely used in image coding, the trellis-coded quantizer is suitable to encode the long sequences with a reasonable complexity. Second, the performance of vector-scalar quantization in this TCQ coding system can be compared with that of entropy coding in the TCQ coding systems proposed by Marcellin, Sriram and Tong [12].

The encoding procedure of the present fixed-rate coding system is summarized as follows:

(1) The edge detector is used to divide each DCT block. The DCT ac vector is formed by zigzag-scanning or directional-scanning the DCT block, depending on the result of the edge detection. The DCT vector is then divided into subvectors and an energy vector is calculated from the subvectors.

(2) Use the first stage of the vector-scalar classification algorithm to search an optimal VQ index for the DCT vector by VQ encoding the energy vector of the corresponding DCT subvectors.

(3) Use the second stage of the classification algorithm to determine the class index for each subvector, according to the optimal VQ index of the DCT vector and the SQ assignment of the components of the VQ codewords.

(4) The sequence in each class is collected from the DCT subvectors belonging to this class. The sequences being coded can be of varying-length or fixed-length (converting the varying-length sequences to fixed-length sequences if necessary). The optimal bit rate allocation algorithm is used to assign the encoding bit rates for the sequences being coded.

(5) Normalize the sequences being encoded. In this encoding system, the SQ elements are approximately energy centroids of the classes, and zero-means of the ac coefficients can be assumed. Thus the SQ elements have been efficiently utilized to normalize the sequences; no side information is needed for the normalization.

(6) The normalized sequences are encoded using the TCQ. Since the sequences being encoded are arranged in the decreasing order of energy, the fixed-rate coding system can be easily built by simply truncating the additional bits assigned to a portion of the last sequence that has the smallest variance.

(7) Use an optimal scalar quantizer to encode dc components of the DCT blocks.

In the decoder, the length of each class can be easily calculated from the optimal VQ indices transmitted. The varying-length sequence belonging to each class is recovered from the equal-length DCT sequences when the equal-length sequences encoded. The DCT vectors are then rebuilt by extracting the DCT subvectors from different-class sequences according to the optimal VQ indices and the class assignment for the VQ.

Rate	Encoding Systems		
	VSC-TCQ	TCQ-FR	TCQ-VR (Rate)
0.25	28.50	25.74	27.37 (0.26)
0.50	31.53	28.37	30.85 (0.51)
1.00	35.48	31.70	35.32 (1.00)

Table 2: Peak-signal-to-noise ratios (in dB) of different TCQ coding systems, for encoding the 256 × 256 Lenna image.

4. SIMULATION RESULTS AND PERFORMANCE COMPARISON

The 256 × 256 Lenna image is encoded by the 8-state TCQ coding systems using average encoding rates of 0.25, 0.50, and 1.00 bits/pixel, respectively. The vector-scalar classification and fixed-rate coding scheme described above are used in the present coding system which is indicated by VSC-TCQ. The optimal bit rate allocation algorithm and TCQ encoding work on the equal-length sequences (the length is 256) being coded.

Table 2 compares our fixed-rate VSQ-TCQ system with Marcellin, Sriram and Tong's fixed-rate TCQ coding system (indicated by TCQ-FR) and variable-rate entropy-coded TCQ coding system (indicated by TCQ-VR) [12], for encoding the 256×256 Lenna image. The comparison between the VSQ-TCQ system and the TCQ-FR system shows that the vector-scalar classification coding scheme has a much higher coding performance than non-classification coding scheme (there is 2.7-3.7 dB improvement on the PSNR performance). The comparison between the VSQ-TCQ system and the TCQ-VR system indicates that the present fixed-rate coding system performs better, especially in the lower-rate coding, than the variable-rate entropy coding system; the vector-scalar classification can achieve most of the performance gain that is obtained by using variable-rate entropy coding.

Figure 1: The 512 × 512 Lenna image is encoded using the fixed-rate TCQ coding system with vector-scalar classification. The encoding rate is 0.5 bits/pixel.

For encoding small-size images, the present coding system is superior to Jeong and Gibson's encoding systems with piecewise-uniform vector quantizers [13], and comparable to Ran and Farvardin's variable-rate entropy coding systems that were based on the three-component image model [14] (the vector-scalar classification has a much lower complexity than the three-component image modeling). However, there is still a performance gap of 1.5-2.0 dB for encoding larger images, when comparing the present fixed-rate coding system with the best variable-rate subband coding systems with entropy coding reported in the literature. It is probably due to the 'resolution' problem of the vector-scalar classification for large images and that DCT coding is worse on encoding performance than subband coding. Figure 1 is the 512 × 512 Lenna image encoded by the present coding system, using an average coding rate of 0.5 bits/pixel (PSNR = 35.52 dB).

5. CONCLUSIONS

This paper introduces a classification method, called vector-scalar classification, where each DCT (ac) vector is indexed using a VQ encoding on the energy vector of the corresponding DCT subvectors, and the subvectors each are then classified according to the class assignment for the VQ codewords. The vector-scalar classification achieves a classification gain about 6.5 dB higher than the Chen-Smith block classification [1]. A fixed-rate TCQ coding system is built using the present classification. The preliminary simulation results show that for encoding small-size images, the vector-scalar classification has captured most of the performance gain that can be achieved by using entropy coding, and this coding system, although fixed-rate, is competitive with some best variable-rate entropy coding systems found in the literature. The system also has potential advantages of handling the problems in the variable-rate coding systems. So far, the performance of vector-scalar classification on larger images is not as good as on smaller images, and further work should be on the improvement of the classification for encoding large-size images and its applications to subband coding systems for images.

REFERENCES

[1] W. Chen and C. Smith, "Adaptive coding of monochrome and color images," *IEEE Trans. on Communications*, vol.25, no.11, pp.1285-1292, Nov. 1977.

[2] J. Woods and S. O'Neil, "Subband coding of images," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol.34, no.10, pp.1278-1288, Oct. 1986.

[3] Y. T. Tse, "Video coding using global/local motion compensation, classified subband coding, uniform threshold quantization and arithmetic coding," Ph.D. thesis, Univ. of California, Los Angeles, 1992.

[4] R. Joshi, T. Fischer, and R. Bamberger, "Optimum classification in subband coding of images," *First IEEE Conf. on Image Processing*, vol.2, pp.883-887, Nov. 1994.

[5] H. Jafarkhani, N. Farvardin, and C. Lee, "Adaptive image coding based on the discrete wavelet transform," *First IEEE Conf. on Image Processing*, vol.3, pp.343-347, Nov. 1994.

[6] Z. Mohdyusof and T. Fischer, "An entropy-coded lattice vector quantizer for transform and subband image coding," *First IEEE Conf. on Image Processing*, vol.1, pp.603-607, Nov. 1994.

[7] X. Ran and N. Farvardin, "A perceptually motivated three-component image model–Part I: Description of the model," *IEEE Trans. on Image Processing*, vol.4, no.4, pp.401-415, April 1995.

[8] Y. Ho and A. Gersho, "Classified transform coding of images using vector quantization," *Int. Conf. on Acoustics, Speech, and Signal Processing*, pp.1890-1893, May 1989.

[9] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. on Communications*, vol.28, no.1, pp.84-95, Jan. 1980.

[10] N. Jayant and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*. Prentice-Hall, Englewood Cliffs, 1984.

[11] M. W. Marcellin and T. R. Fischer, "Trellis coded quantization of memoryless and Gauss-Markov sources," *IEEE Trans. on Communications*, vol.38, no.1, pp.82-93, Jan. 1990.

[12] M. Marcellin, P. Sriram, and K. Tong, "Transform coding of monochrome and color images using trellis coded quantization," *IEEE Trans. on Circuits and Systems for Video Technology*, vol.3, no.4, pp.270-276, Aug. 1993.

[13] D. Jeong and J. Gibson, "Image coding with uniform and piecewise-uniform vector quantizers," *IEEE Trans. on Image Processing*, vol.4, no.2, pp.140-146, Feb. 1995.

[14] X. Ran and N. Farvardin, "A perceptually motivated three-component image model–Part II: Applications to image compression," *IEEE Trans. on Image Processing*, vol.4, no.4, pp.430-447, April 1995.

Proceedings of the 1995 International Conference on Image Processing (ICIP '95)
0-8186-7310-9/95 $10.00