GENERALIZED VARIABLE DIMENSIONAL SET PARTITIONING FOR EMBEDDED WAVELET IMAGE COMPRESSION*

Debargha Mukherjee and Sanjit K. Mitra
Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106.
Email: (debu, mitra)@iplab.ece.ucsb.edu

ABSTRACT

A vector enhancement of Said and Pearlman's Set Partitioning in Hierarchical Trees (SPIHT) methodology, named VSPIHT, has recently been proposed for embedded wavelet image compression. While the VSPIHT algorithm works better than scalar SPIHT for most images, a common vector dimension to use for coding an entire image may not be optimal. Since statistics vary widely within an image, a greater efficiency can be achieved if different vector dimensions are used for coding the wavelet coefficients from different portions of the image. We present a generalized methodology for developing a variable dimensional set partitioning coder, where different parts of an image may be coded in different vectoring modes, with different scale factors, and upto different number of passes. A Lagrangian rate-distortion criterion is used to make the optimum coding choices. Coding passes are made jointly for the vectoring modes to produce an embedded bitstream.

1. INTRODUCTION

The wavelet transform, over the last few years, has grown to be a very effective means for transform coding of images [1]-[11]. Using the conceptual foundations of zerotree prediction laid by Shapiro's EZW [3] algorithm, Said and Pearlman [4] recently developed a very efficient wavelet image compression scheme, called *Set Partitioning in Hierarchical Trees* (SPIHT). In both schemes efficient scans are used to partially order the scalar wavelet coefficients by magnitude, followed by progressive refinement on a bit-plane by bit-plane basis. The bitstream generated is perfectly embedded. Xiong *et al* [5] developed a complex space-frequency quantization scheme based on [3] that uses a rate-distortion criterion to jointly optimize zerotree quantization and scalar frequency quantization. Several modifications (e.g. [6]) has been attempted on both [3] and [4] for improved efficiency. Inspired by the success of these scalar schemes, several researchers proposed vector extensions of these algorithms. While the lattice VQ based schemes of Da Silva et. al. [7], Knipe et. al. [8], and Mukherjee and Mitra [9], are considerably generic, and have fast algorithms, the trained VQ schemes [10], [11] are usually superior in rate-distortion performance.

Although a trained VQ based VSPIHT coder is more efficient than a scalar SPIHT coder for most images, the performance for an arbitrary image depends heavily on the distribution of the wavelet coefficients in it. For example, if the distribution of coefficients in a particular portion of an image is such that only a few high magnitude coefficients exist, a large number of vectors will have only one or two high magnitude coefficients. VQ will therefore be unnecessarily wasting too many bits on insignificant coefficients, and as such, the coding performance for the same portion with VSPIHT is likely to be worse than that with scalar SPIHT. In this work we propose a generalized image adaptive variable dimensional vector SPIHT coding paradigm where different parts of an image can be coded as vectors of different sizes and different scales, with different number of set-partitioning passes, based on performance. This ensures that the algorithm works better than both scalar and vector algorithms taken separately. The coding choices for each portion is transmitted to the decoder as side information.

In the next section we present a brief overview of vector set partitioning, with particular emphasis on the aspects relevant to this work. In Section 3 we introduce the coding paradigm of the current work. Section 4 presents the methodology used for making the various coding choices, such as dimension, scale and number of passes. In section 5, the implementation details and coding results for a scalar-vector coder are presented. Finally Section 6 concludes the paper.

2. OVERVIEW OF VECTOR SPIHT

We present a brief review of the aspects of VSPIHT relevant to this work, in particular, variable vector scaling and adaptive arithmetic coding.

2.1 Review of VSPIHT

In VSPIHT, first a dyadic wavelet decomposition of an image is performed. Then, the wavelet transform coefficients in each $H \times V$ window in each subband are grouped to form a single vector of dimension HV, which forms the elemental quantization unit. In the course of multiple set partitioning passes that follow, these vectors are classified into several classes using ordered lists QLIP, QLIS and QLSP, based on their vector magnitude in relation to certain decreasing thresholds R_0, R_1, R_2, etc. All vectors with magnitude higher than R_i, but less than R_j, $j = 0,1,...,R_{i-1}$, constitute Class i. Each successive vector set-partitioning pass is associated with one of these vector magnitude thresholds and yields a new set of vectors which have magnitudes higher than the threshold associated with the pass. The thresholds decrease from one pass to the next, usually by a factor of 2. Vectors thus classified are gradually refined using class-specific successive refinement VQ systems. Multistage VQ or tree-structured VQ, or a combination of both, may be used for designing the progressive refinement VQ systems. In this implementation we use VQ systems whose first stage is tree-structured, and is followed by stages of multistage VQ.

2.2 Variable Vector Scaling

In order to bring about a certain amount of uniformity in the way images with varying dynamic range of wavelet coefficients are coded with a common set of VQ systems, the wavelet vectors formed by grouping are each scaled by a factor γ, before the set partitioning passes start. The factor γ is given by:

$$\gamma = \frac{R_{-1}}{max(\|v\|)} \quad (1)$$

where R_{-1} is a coding parameter greater than R_0, and $max(\|v\|)$ denotes the maximum vector magnitude in an image [11]. The fac-

* This work was supported by ONR grant N00014-95-1-1214.

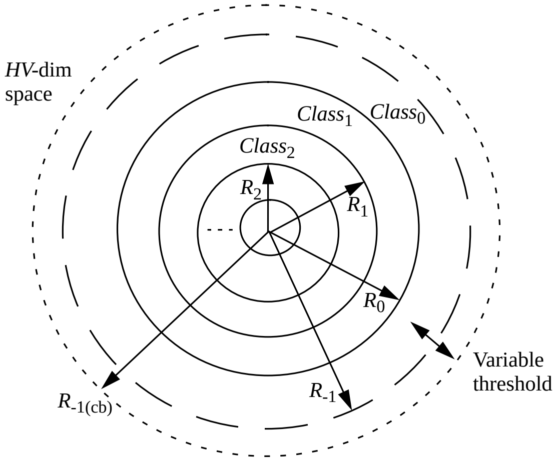


Figure 1. Decreasing magnitude thresholds to determine significance of vectors, and the corresponding classes.

tor γ is transmitted to the decoder with high precision for reconstruction. After scaling, all vectors are guaranteed to lie within a HV-dimensional shell of radius R_{-1}. We find that the factor R_{-1} has a significant impact on the rate distortion performance for a particular image, and therefore, must be optimized for best performance.

In the light of the above scaling mechanism, it is appropriate to consider the procedure used for generating the training vectors for designing the successive refinement VQs. A large set of training images are wavelet transformed, and the coefficients are grouped appropriately to form vectors. For each image, the wavelet vectors are scaled by a factor $\gamma_{(cb)}$, given by:

$$\gamma_{(cb)} = \frac{R_{-1\,(cb)}}{max\,(\|v\|)}\,, \tag{2}$$

where $R_{-1\,(cb)}$ is a parameter. The scaled vectors for all images are partitioned into classes based on thresholds R_0, R_1, R_2, etc., and the candidates in each class are then used to design the corresponding VQ system. Note that the factor R_{-1} used during coding, and the factor $R_{-1\,(cb)}$ used for training set generation are not necessarily the same, although the thresholds R_0, R_1, R_2, etc. used for classification remain the same. Usually, R_{-1} is chosen as less than $R_{-1\,(cb)}$, but greater than R_0 during coding. Figure 1 shows a typical classification scenario in HV-dimensional space. Note that the threshold R_{-1} is variable during coding.

2.3 Adaptive Arithmetic Coding

To enhance the rate distortion performance of the VSPIHT coder, two different kinds of adaptive arithmetic coding can be performed [11]. The first is aimed at exploiting repetitive patterns in images. When patterns repeat in an image, similar wavelet vectors recur within the same subband. Similar vectors, when vector quantized coarsely using the first stage VQ, are likely to yield the same encoding index. Adaptive arithmetic coding of the first stage VQ index for each class and each subband is used to exploit this redundancy. The adaptive arithmetic coder progressively assigns smaller and smaller codelengths to repeating indices. In order to allow the models to adapt fast enough to the underlying statistics, it is necessary that the first stage VQ, which is also tree-structured, be designed with relatively few codevectors.

The second kind of adaptive arithmetic coding is aimed at reducing the significance information bits associated with set partitioning, in a manner similar to scalar SPIHT [4]. The vectors in the lists are maintained in groups of 2×2, and the significance information for the group is transmitted jointly using multiple adaptive context models.

3. VARIABLE DIMENSIONAL SET PARTITIONING

The variable dimensional set partitioning methodology is conveniently explained by means of the diagrams in Figure 2. After a dyadic wavelet decomposition of an image, the low-low subband, where all the roots of the spatial orientation trees reside, is further divided into L *superblocks* of size $M \times N$. Each such superblock has a *subimage* in its *region of support*, consisting of itself, the $M \times N$ superblocks in the same position in the lowest subbands of the LH, HL and HH orientations, along with all their descendants (see Figure 2). Each of these superblock subimages can be further divided in K ways into blocks of size $H_i \times V_i$, $i = 0, 1, \ldots, K-1$, as shown in Figure 2. A decision mechanism is used to decide for each subimage, how, among the available K ways, the coefficients in it will be grouped into vectors, for subsequent VSPIHT coding. The encoder makes a decision based on rate distortion performance, and transmits the decision map to the decoder as side information. Additionally, for each of the subimages, a different scaling parameter γ_i, $i = 0,1,...,L-1$, is transmitted to the decoder for best results. Furthermore, the set partitioning passes on each subimage may be executed upto different stages, as we see later in this section. Note that different successive refinement systems are required to code the subimages mapping to different vectoring modes.

Each subimage rooted at the low-low superblock is essentially encoded or decoded by set-partitioning independent of others. To this end, a parent-child relationship is defined for each subimage with the elemental coding units in each being the vectors obtained by grouping the coefficients therein in one of the K ways. Depending on whether adaptive arithmetic coding for reducing the significance information is used or not, two types of variable dimensional SPIHT must be considered. If adaptive arithmetic coding for significance information is not used, there is little restriction on the possi-

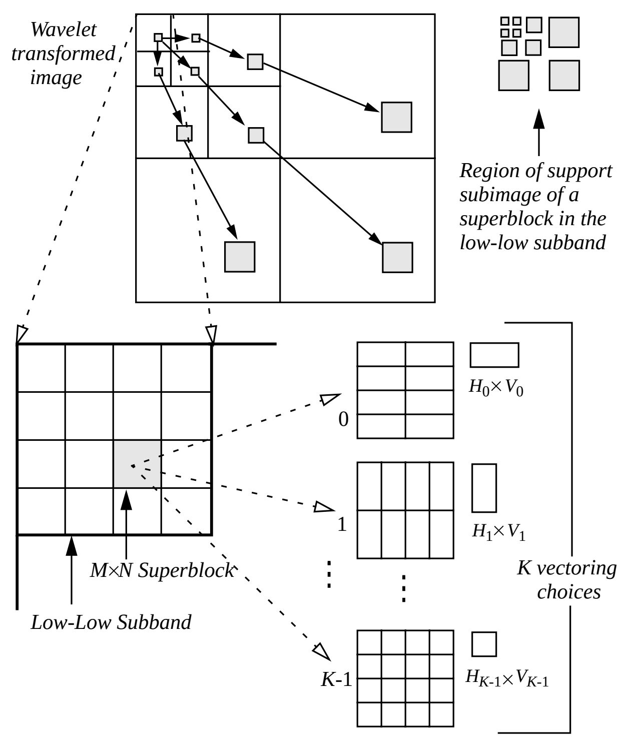


Figure 2. Region of support of a block in low-low Subband, and various vectoring choices for it.

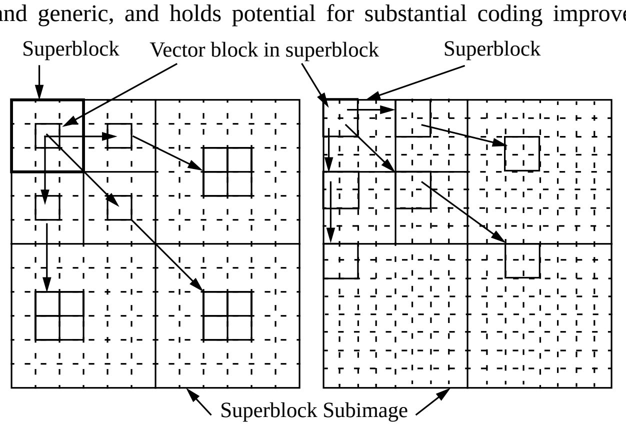
ble ways in which the $M \times N$ superblocks may be divided into smaller vector blocks. In this case, the parent child relationship between vector blocks in the superblock subimage may be defined as in Figure 3(a) similar to that in [3]. On the other hand, if adaptive arithmetic coding is used to reduce significance information, groups of 2×2 vectors must be maintained together. As such, in the superblock, there must be an even number of vector blocks in both horizontal and vertical directions, for each of the K possible vectoring modes. In this case, the parent child relationship is defined as in Figure 3(b) similar to that in [4]. Note that because of this constraint, with adaptive arithmetic coding, the superblocks cannot be too small. If arithmetic coding is not used, smaller superblocks can be used, thus allowing finer vectoring mode decisions.

Given the decision map and the associated set of scaling parameters γ_i for each of the L superblocks, the encoder and decoder operates as follows. First, each subimage is scaled appropriately depending on the particular value of γ_i associated with it. A set of ordered lists - QLIP, QLIS and QLSP - is then created for each of the K vectoring modes. The QLIP and QLIS lists for each vectoring mode is initialized as in [4] with vectors of the appropriate size taken from those subimages that are to be coded in that mode. After initialization, the set partitioning passes commence to produce an embedded bitstream. Each full pass is actually an aggregate of K smaller passes, one for each coding mode. In practice, the K QLIPs are first processed one after another. Then the K QLISs are processed. Finally, the refinement passes are conducted using the K QLSPs. Therefore, each full pass can be viewed as consisting of 3 subpasses, the QLIP-subpass, the QLIS-subpass, and the QLSP-subpass. It is sometimes convenient to denote the progression of the algorithm in finer units of subpasses, rather than passes. For example, coding with 16 subpasses would mean coding with 5 full passes and only the QLIP-subpass of the 6th pass. Note that each subimage is essentially coded independently of the others in this approach, although the bit stream generated is mixed.

Additional encoding flexibility can be incorporated if the number of subpasses upto which each subimage is coded is varied based on performance, rather than executing all the subpasses for all subimages. The optimum number of subpasses for each subimage, p_i, $i = 0,1,...,L-1$, is also transmitted to the decoder as side-information.

4. DECISION MAKING

While the paradigm described above is considerably flexible and generic, and holds potential for substantial coding improvement when the vectoring modes, scaling parameters, and number of subpasses are appropriately chosen, the task of making the optimum decisions in the most general case, is by no means compuationally inexpensive. We present an approach to making these coding choices based on a Lagrangian rate-distortion optimization.



(a) No adaptive arithmetic coding of significance bits

(b) Adaptive arithmetic coding for significance bits used (Vector blocks grouped as 2x2)

Figure 3. Parent-child relationships for two possible implementations

In this approach, the best coding choice can only be based on rate-distortion performance upto the end of a specific number of subpasses. This is because, at the end of a subpass, all the choices reach a common state of completion, thereby providing a uniform platform for selecting the best. Let the L subimages, corresponding to the L superblocks the low-low band is divided into, be denoted as x_i, $i = 0, 1, \ldots, L-1$. The full image X, therefore, is an aggregate of the subimages: $X = \{x_0, x_1, \ldots, x_{L-1}\}$. Let the distortion and rate achieved when a subimage x_i is coded using vectoring mode m_i, $m_i \in \{0, 1, \ldots, K-1\}$, and scaling parameter γ_i, with p_i subpasses, be given as: $D(x_i, m_i, \gamma_i, p_i)$, and $R(x_i, m_i, \gamma_i, p_i)$, respectively. Since all the subimages are coded independently, neglecting the rate savings due to adaptive arithmetic coding for repetitive patterns, the overall distortion and rate obtained for the entire image X, given the set of modes $M = \{m_0, m_1, \ldots, m_{L-1}\}$, the set of scaling parameters $\Gamma = \{\gamma_0, \gamma_1, \ldots, \gamma_{L-1}\}$, and the number of subpasses used $P = \{p_0, p_1, \ldots, p_{L-1}\}$, are given by:

$$D(X, M, \Gamma, P) = \sum_{i=0}^{L-1} D(x_i, m_i, \gamma_i, p_i), \qquad (3a)$$

$$R(X, M, \Gamma, P) = \sum_{i=0}^{L-1} R(x_i, m_i, \gamma_i, p_i) \qquad (3b)$$

Our task then is to optimize the parameters M, Γ, and P for the lowest possible distortion $D(X, M, \Gamma, p)$ under a rate constraint R_d, i.e. $R(X, M, \Gamma, p) < R_d$. The constrained optimization problem can be readily transformed to an unconstrained problem using a Lagrangian parameter λ. The problem then becomes one of minimizing the lagrangian cost function $J(X, M, \Gamma, P)$, given by:

$$J(X, M, \Gamma, P) = D(X, M, \Gamma, P) + \lambda R(X, M, \Gamma, P) \qquad (4)$$

If the Lagrangian cost function for each subimage be denoted:

$$J(x_i, m_i, \gamma_i, p_i) = D(x_i, m_i, \gamma_i, p_i) + \lambda R(x_i, m_i, \gamma_i, p_i), \qquad (5)$$

the overall cost function can be written as the summation:

$$J(X, M, \Gamma, P) = \sum_{i=0}^{L-1} J(x_i, m_i, \gamma_i, p_i) \qquad (6)$$

Since the individual subimage cost functions are, for all practical purposes, independent of each other, minimizing the overall cost function in the LHS of Eq. (6) is equivalent to minimizing each of the subimage cost functions on the RHS. In other words, the optimization procedure chooses:

$$\{m_i, \gamma_i, p_i\} = \underset{(m, \gamma, p)}{\arg min} [J(x_i, m, \gamma, p)], i = 0, 1, \ldots, L-1. \qquad (7)$$

In practice, the scale factor is constrained to take on only certain discrete values from a small codebook. Each subimage is test coded in all the available coding modes, with all the available scale factors, upto all the possible number of subpasses. The rate and the distortion obtained for each combination is computed. The combination that yields the lowest Lagrangian among the candidates is eventually chosen as the optimum for that subimage. The decisions thus made for each subimage is subsequently used in the actual coding process to generate the bit stream. The value of λ determines the relative importance given to rate and distortion during the optimization procedure. The higher the value of λ, the lower the final bit rate obtained, and vice versa. By adjusting the value of λ, using techniques like binary search, bit rates close to the desired, R_d, can be obtained. Once the rate and distortions for each candidate $\{m_i, \gamma_i, p_i\}$ combination for each subimage x_i is computed and stored, the value of λ can be adjusted for the desired rate with little additional complexity.

5. IMPLEMENTATION AND RESULTS

We implemented a scalar-vector set partitioning coder based on the above principles. After a dyadic wavelet decomposition of an image, the low-low band is divided into superblocks of size 4×4. The corresponding subimages can be coded either as scalars or as vectors of dimension 4 in 2×2 blocks, yielding two possible vectoring modes. Since adaptive arithmetic coding for significance information is used, the scalars or vectors are maintained in groups of 2×2 in each subimage, with the parent child relation in each being given as in Figure 3(b). Additionally, adaptive arithmetic coding for repetitive patterns in the vector case is also used. The class codebooks used for 4-dimensional vector quantization are tree-structured at the first stage, followed by multistage VQ at the successive stages. The scalar mode is coded exactly as in [4], apart from variable scaling. The scale factor γ_i for each subimage is decomposed as follows: $\gamma_i = \gamma\eta_i$, where γ is given as in Eq. (1), and η_i is chosen from a small codebook. The codebook for the η_i's may look something like $\{\sigma^{-4}, \sigma^{-3}, \sigma^{-2}, \sigma^{-1}, 1, \sigma^1, \sigma^2, \sigma^3\}$, with σ = 1.25. A factor α, defined as the ratio of the largest scalar magnitude in the image, to the largest vector magnitude, is also computed. Note that if all $\eta_i = 1$, the maximum possible vector magnitude in the scaled subimage would be R_{-1}, while the maximum possible scalar magnitude would be αR_{-1}. γ is transmitted to the decoder with high precision, while α is transmitted after coarse quantization to $\hat{\alpha}$. Thereafter, for each subimage, the codebook scale η_i, the mode m_i, and the number of subpasses p_i are transmitted to the decoder. The coder scales all the subimages by $\gamma\eta_i$ before coding. All vector mode subimages after scaling are coded using decreasing vector magnitude thresholds R_0, R_1, R_2,..., while all scalar mode subimages after scaling are coded using octavely decreasing thresholds $\hat{\alpha} R_{-1}/2$, $\hat{\alpha} R_{-1}/2^2$, $\hat{\alpha} R_{-1}/2^3$,.... The decoder decodes the set partitioning bit stream, and performs the reverse scaling with the information provided to it, to reconstruct the true wavelet coefficients.

Figure 4 compares for the Lena and the Barbara image, the peak-SNR vs. bits per pixel results obtained by a scalar SPIHT coder and a VSPIHT coder of dimension 4, against that obtained by the optimized multimode scalar-vector coder presented here. The data for the latter for various bit rates is generated by adjusting the λ values. Only full passes, and the η-codebook given in the above paragraph are used in the optimization. The 7/9 bi-orthogonal filters in [2] are used in each case. For Barbara, VSPIHT is better than SPIHT, while for Lena, SPIHT is slightly better than VSPIHT. For both images the PSNR obtained by variable dimensional SPIHT at each bitrate is found to be better than both the scalar and the vector SPIHT coders.

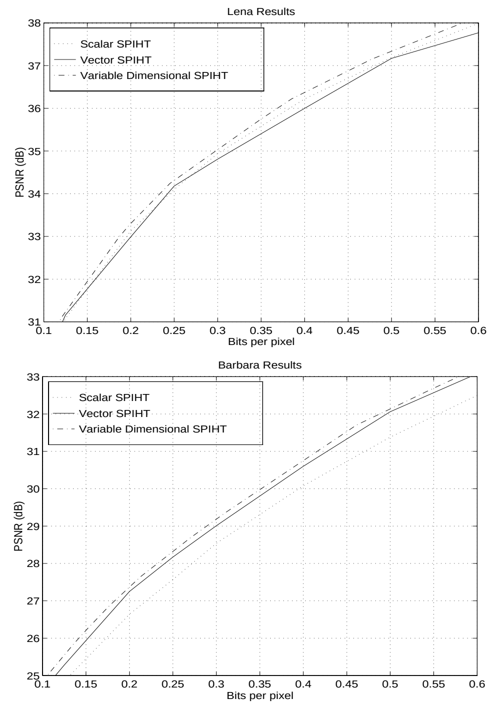


Figure 4. Coding results for SPIHT, VSPIHT and Variable Dimensional SPIHT for Lena and Barbara Images

6. CONCLUSIONS

A generalized framework for variable dimensional set-partitioning quantization of wavelet coefficients for image coding is presented. The coding decisions are based on minimizing a Lagrangian rate-distortion cost function. Improvement in coding efficiency is achieved, albeit at the expense of computational complexity.

7. REFERENCES

[1] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Processing*, vol. 1, pp. 205-220, April 1992.

[2] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3445-62, Dec. 1993.

[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. 243-50, June 1996.

[4] Z. Xiong, K. Ramachandran, and M. T. Orchard, "Space-Frequency Quantization for Wavelet Image Coding," *IEEE Trans. Image Processing*, vol. 6, no. 5, pp. 677-93, May 1997.

[5] H. J. Wang and C.-C. J. Kuo, "A multithreshold wavelet coder (MTWC) for high fidelity image compression," *Proc. International Conference on Image processing*, Santa Barbara, California, vol. 1, pp. 652-55, October 1997.

[6] E. A. B. Da Silva, D. G. Sampson, M. Ghanbari, "A successive approximation vector quantizer for wavelet transform image coding," *IEEE Trans. Image Processing*, vol. 5, no. 2, pp. 299-310, Feb. 1996.

[7] J. Knipe, X. Li, and B. Han, "An improved lattice vector quantization scheme for wavelet compression," *IEEE Trans. Signal Processing*, vol. 46, no. 1, pp. 239-43, Jan 1998.

[8] D. Mukherjee and S. K. Mitra, "Vector set partitioning with successive refinement Voronoi lattice VQ for embedded wavelet image coding," *Proc. International Conference on Image processing*, Chicago, Illinois, October 1998 - to be published.

[9] D. Mukherjee and S. K. Mitra, "Vector set partitioning with classified successive refinement VQ for embedded wavelet image coding," *Proc. International Symposium on Circuits and System*, Monterey, California, vol. 4, pp. 25-28, June 1998.

[10] D. Mukherjee and S. K. Mitra, "Vector set partitioning with classified successive refinement VQ for embedded wavelet image and video coding," *Proc. International Conference on Acoustics, Speech, and Signal Processing*, Seattle, Washington, vol. 5, pp. 2809-12, May 1998.

[11] D. Mukherjee and S. K. Mitra, "Arithmetic Coded Vector SPIHT with classified tree-multistage VQ for color image coding," *IEEE Workshop on Multimedia Signal Processing*, Los Angeles, California, Dec 1998 - to be published.