array exhibits advantages in terms of latency. The latency of the proposed structure is $(2N-1)T$, whereas that of Chang and Chen's systolic array is $(4N-2)T'$.

IV. Conclusions

In this paper, we propose unified systolic arrays for computation of the 1-D and 2-D DCT/DST/DHT. Compared to the conventional methods, the proposed systolic arrays exhibit advantages in terms of the number of PE's and latency. Also, the unified systolic arrays can be employed for computation of the 1-D and 2-D IDCT/IDST/IDHT. Further research will focus on development of the wavefront array architecture to cope with a clock skew problem.

Acknowledgment

The authors would like to thank the anonymous reviewers for their valuable comments.

References

[1] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE Trans. Commun.*, vol. COM-23, pp. 90–93, Jan. 1974.
[2] K. R. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, and Applications*. New York, NY: Academic, 1990.
[3] R. N. Bracewell, *The Hartley Transforms*. Oxford: Oxford University Press, 1986.
[4] H. S. Hou, "A fast recursive algorithm for computing the discrete cosine transform," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 1455–1461, Oct. 1987.
[5] H. V. Sorensen, D. L. Jones, C. S. Burrus, and M. T. Heideman, "On computing the discrete Hartley transform," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 1231–1238, Oct. 1985.
[6] H. S. Malvar, "Fast computation of the discrete cosine transform and the discrete Hartley transform," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 1484–1485, Oct. 1987.
[7] J.-L. Wu and S.-C. Pei, "The vector split-radix algorithm for 2-D DHT," *IEEE Trans. Signal Processing*, vol. 41, pp. 960–965, Feb. 1993.
[8] N. I. Cho and S. U. Lee, "DCT algorithms for VLSI parallel implementations," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 121–127, Jan. 1990.
[9] L.-W. Chang and M.-C. Wu, "A unified systolic array for discrete cosine and sine transforms," *IEEE Trans. Signal Processing*, vol. 39, pp. 192–194, Jan. 1991.
[10] M. H. Lee, "On computing 2-D systolic algorithm for discrete cosine transform," *IEEE Trans. Circuits Syst.*, vol. 37, pp. 1321–1323, Oct. 1990.
[11] C. Chakrabarti and J. JáJá, "Systolic architectures for the computation of the discrete Hartley and the discrete cosine transforms based on prime factor decomposition," *IEEE Trans. Comput.*, vol. 39, pp. 1359–1368, Nov. 1990.
[12] A. S. Dhar and S. Banerjee, "An array architecture for fast computation of discrete Hartley transform," *IEEE Trans. Circuits Syst.*, vol. 38, pp. 1095–1098, Sept. 1991.
[13] J.-H. Hsiao, L.-G. Chen, T.-D. Chiueh, and C.-T. Chen, "Novel systolic array design for the discrete Hartley transform with high throughput rate," in *Proc. Int. Conf. Circuits Systems*, Minneapolis, MN, May 1993, pp. 1567–1570.
[14] J.-I. Guo, C.-M. Liu, and C.-W. Jen, "A CORDIC-based VLSI array for computing 2-D discrete Hartley transform," in *Proc. Int. Conf. Circuits Syst.*, Minneapolis, MN, May 1993, pp. 1571–1574.
[15] I. Gertner and M. Shamash, "VLSI architectures for multidimensional Fourier transform processing," *IEEE Trans. Comput.*, vol. C-36, pp. 1265–1274, Nov. 1987.
[16] L. W. Chang and M. Y. Chen, "A new systolic array for discrete Fourier transform," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-36, pp. 1665–1666, Oct. 1988.
[17] J.-F. Yang, B.-L. Bai, and S.-C. Hsia, "An efficient two-dimensional inverse discrete cosine transform algorithm for HDTV receivers," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, pp. 25–30, Feb. 1995.
[18] P. Pirsch, N. Demassieux, and W. Gehrke, "VLSI architectures for video compression- A survey," *Proc. IEEE*, vol. 83, pp. 220–246, Feb. 1995.
[19] H. T. Kung, "Why systolic architectures?," *IEEE Comput.*, vol. 15, pp. 37–46, Jan. 1982.
[20] S. Y. Kung, *VLSI Array Processors*. Englewood Cliffs, NJ: Prentice-Hall, 1988.

Wavelet Transforms in a JPEG-Like Image Coder

Ricardo de Queiroz, C. K. Choi, Young Huh, and K. R. Rao

***Abstract*—The discrete wavelet transform (DWT) is incorporated into the JPEG baseline system for image coding. The discrete cosine transform (DCT) is replaced by an association of two-channel filter banks connected hierarchically. JPEG block-scanning and quantization schemes are adopted while we use JPEG's entropy coder. The changes in scanning can be incorporated into the transform block in such a way that the only part that needs to be changed in a JPEG framework is to replace the DCT by the DWT. Objective results and reconstructed images are presented demonstrating that the proposed coder outperforms JPEG and approaches the performance of more sophisticated and complex wavelet coders. However, it does not require full-image buffering nor imposes a large complexity increase.**

***Index Terms*—Image coding, JPEG, wavelet transforms.**

I. Introduction

The discrete cosine transform (DCT) [1] plays a major role in popular image data compressors, and DCT-based algorithms and hardware are widely available nowadays. In still image compression, the JPEG baseline system (JPEG) [2] is a "de facto" standard based on the DCT and there are several chips and software available to perform JPEG compression and decompression. Recently, much attention has been devoted to the dyadic discrete wavelet transform (DWT) [3], which has a versatile time-frequency localization due to a pyramid-like multiresolution decomposition. Several authors have studied the DWT in image coding, obtaining a performance superior to JPEG or to most of other DCT-based coders [4]–[10]. However, often, the improved performance carries along a large increase in coding complexity. In this work, we explore a JPEG-like coder which uses the main building blocks of JPEG and the DWT to achieve a versatile system that outperforms JPEG with a small complexity penalty. We refer to it as DWT-JPEG.

For large resolution images,[1] compression is necessary, and it is desirable to find a compression scheme which avoids: 1) buffering the image, 2) performing multiple passes, and 3) very complex processing. Most of the efficient wavelet coders [5]–[10] require

Manuscript received February 28, 1995; revised February 12, 1996. This paper was recommended by Associate Editor C. A. Gonzales.

R. de Queiroz was with the Electrical Engineering Department, University of Texas at Arlington, Arlington, TX 76019 USA. He is now with Xerox Corp., Webster, NY 14580 USA.

C. K. Choi, Y. Huh, and K. R. Rao are with the Electrical Engineering Department, University of Texas at Arlington, Arlington, TX 76019 USA.

Publisher Item Identifier S 1051-8215(97)01134-8.

[1] For example, an 8.5″ by 11″ document, scanned or generated at 600 pixels-per-inch, yields a 5100×6000-pels bitmap image, and requires roughly 30 MB of storage space per color separation.

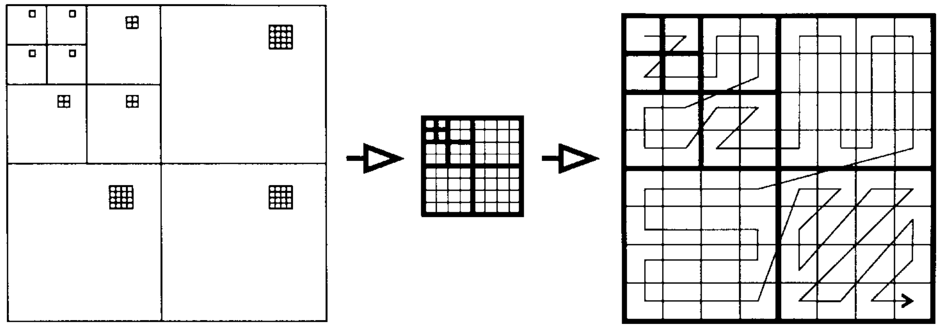

Fig. 1. Illustration of the block construction procedure for a three-level DWT. The DWT coefficients are usually displayed and grouped by subbands. In this example, coefficients in the same location, but different subbands are grouped together in a block. The resulting block is scanned into a vector as shown.

buffering,[2] are much more complex than JPEG, and may require multiple passes through the image. JPEG possesses computational advantages but suffers from blocking and ringing artifacts at higher compression ratios. We attempt to reduce these problems by replacing JPEG's transform. As a remark, some JPEG chipsets have the DCT implemented externally, easing the replacement task.

Some attempts were made to replace the DCT by a cascade of filter banks in JPEG, for example using *ad-hoc* wavelet packets [11] or a full-tree [12]. In the proposed DWT-JPEG, each transform block will not only contain coefficients in different subbands but also coefficients in different spatial locations.

II. Proposed Coder

In the DWT, the coefficients are generated by applying a cascade of two-channel filter banks to the input image [3]. In any subband/transform coding procedure, the coefficients can be grouped according to the subbands or according to spatial position (block). When using the DCT, it is common to group coefficients into blocks (common spatial location, different subbands) while using the DWT, it is common to have subband oriented grouping (common subband, different spatial locations). We group the DWT coefficients into blocks as illustrated in Fig. 1. For S-levels DWT, blocks of $2^S \times 2^S$ samples are constructed. The resulting block is scanned into a vector in order to be processed by the remaining parts of JPEG. The subbands are scanned from low- to high-frequency, obeying the following subband scan sequence: horizontal, then vertical, then diagonal. Vertical subbands are scanned horizontally, and vice-versa. The diagonal subbands are scanned in zigzag.

In baseline JPEG, DCT coefficients are scanned in zigzag ordering, quantized, and encoded as shown in Fig. 2(a). In DWT-JPEG, the DCT is replaced by the DWT. Using the DWT, there are coefficients which belong to different subbands, but also multiple coefficients in the same subband for a given block. Hence, the scanning process as well as the quantizer selection may be changed. Quantizer tables are downloadable in JPEG and only a new scanning sequence has to be used (see Fig. 1). The flow-graph for implementing DWT-JPEG is shown in Fig. 2(b).

Comparing Fig. 2(a) and Fig. 2(b) we can see that the differences reside on the transform and scanning components. It is easy to see that one can reorder the output DWT coefficients in such a way that after undergoing a zigzag rescanning, the resulting coefficients will be arranged as in the sequence depicted in Fig. 1. Therefore, one can implement DWT-JPEG as in Fig. 2(c) using existing parts of JPEG (except for the transform).[3]

[2]The computation of an FIR-filter-based DWT does not require full-image buffering.

[3]For $S \neq 3$ JPEG may have to be adjusted to handle blocks with $2^{2S} (\neq 64)$ samples.

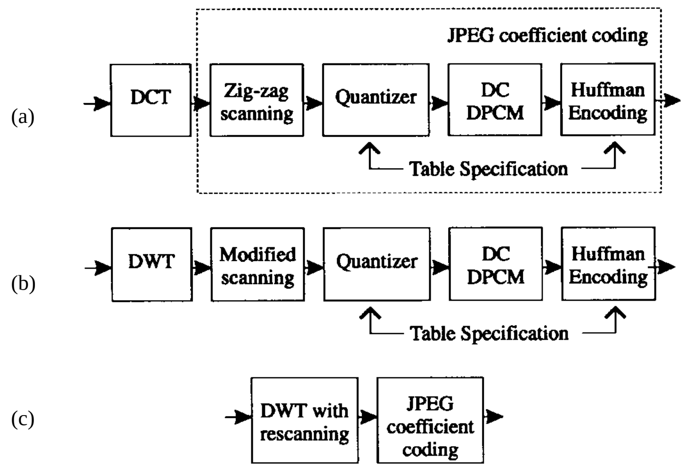

Fig. 2. (a) Baseline JPEG basic encoding flow diagram. (b) The proposed coder (DWT-JPEG) based on a JPEG structure. (c) Optional implementation of DWT-JPEG by rearranging the DWT coefficients in such a way that after a regular zigzag rescanning the coefficients are displaced in the desired order.

JPEG uses a fixed quantizer table with 64 entries, representing steps of uniform quantization for each coefficient in a block. JPEG also provides an example (default) table for the luminance and chrominance components, designed after extensive tests [2]. Let the block size be $M \times M (M = 2^S)$ and the step sizes be Δ_{ij} for $0 \leq (i, j) \leq M - 1$. The step sizes used in our tests are found as

$$\Delta_{ij} = \frac{A}{H\left(\frac{\sqrt{i^2+j^2}}{M}\right)} \tag{1}$$

where A is a scaling to control the bit rate. We then define a model for $H(x)$ as

$$H(x) = (a_0 + a_1 x + a_2 x^2 + a_3 x^3) e^{(a_4 x + a_5 x^2)} \tag{2}$$

and we optimized the parameters for $S = 3$ in such a way as to minimize the error between the achieved table and the table provided by JPEG. The parameters found are

$$\begin{aligned} a_0 &= 0.794 \quad a_1 = -1.639 \quad a_2 = 0.614 \\ a_3 &= 0.470 \quad a_4 = 4.277 \quad a_5 = -4.892. \end{aligned} \tag{3}$$

Because we use the DWT, in a block there will be several coefficients in the same band, and the model (developed for the DCT) will have

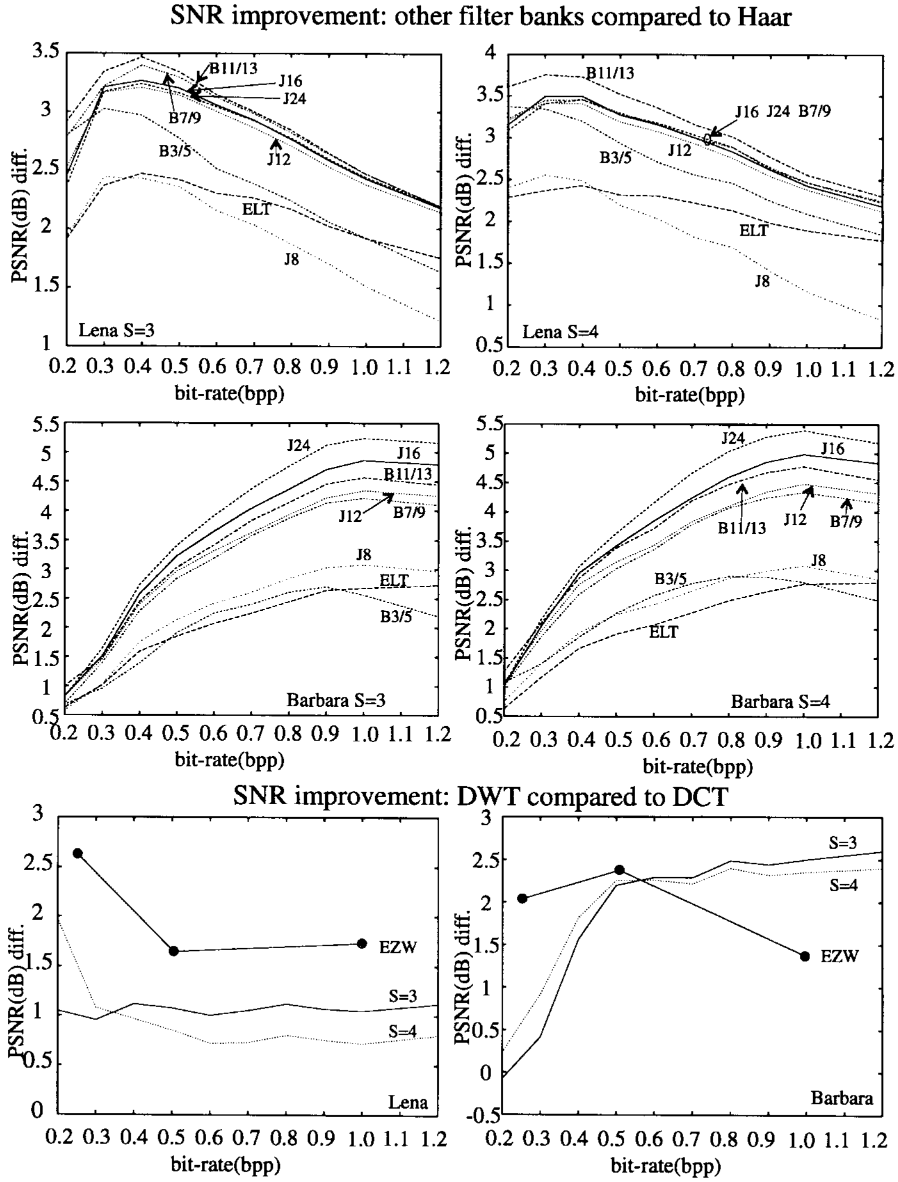

Fig. 3. Objective evaluation of DWT-JPEG performance using plots of PSNR differences.

different step sizes for different coefficients in a block. Then, we have chosen to average the quantizer steps for all coefficients in the same subband and to replace the respective stepsizes by this average. Also, if we use orthonormalized filter banks (not unity gain) the subbands will have different gains. In order to apply the quantizer, one may scale up the coefficients in high-pass bands or scale down the quantizer steps. For $A = 6.7$, the quantizer table for three-stage DWT is given by

$$Q = \begin{bmatrix} 8 & 7 & 8 & 8 & 34 & 34 & 34 & 34 \\ 7 & 7 & 8 & 8 & 34 & 34 & 34 & 34 \\ 8 & 8 & 12 & 12 & 34 & 34 & 34 & 34 \\ 8 & 8 & 12 & 12 & 34 & 34 & 34 & 34 \\ 34 & 34 & 34 & 34 & 55 & 55 & 55 & 55 \\ 34 & 34 & 34 & 34 & 55 & 55 & 55 & 55 \\ 34 & 34 & 34 & 34 & 55 & 55 & 55 & 55 \\ 34 & 34 & 34 & 34 & 55 & 55 & 55 & 55 \end{bmatrix}. \quad (4)$$

One can also "stretch" and "bias" the quantizer table using $H(f_r\sqrt{(\alpha i)^2 + (\beta j)^2}/M)$, where f_r is a relative sampling frequency, while α and β can be used to bias the quantizer table in either direction.

III. Tests

Tests were carried to select DWT parameters (number of levels and filter bank) and to evaluate the coder performance. We tested a family of Johnston's quadrature mirror filters (QMF's) with 8, 12, 16, and 24 taps [13] (denoted by J8 through J24, respectively). We also used biorthogonal symmetric filter banks with 3/5-taps (B3/5) [14], 7/9-taps (B7/9) [15], and 11/13-taps (B11/13) [15]. For completeness, we also tested a fast cosine modulated filter bank known as the extended lapped transform (ELT) [16]. The Haar transform (two-channel DCT) was used as a lower bound reference. Fig. 3 shows comparative results for all filters. We tested the DWT-JPEG using $S = 3$ and $S = 4$ (three and four levels) for 512 $\times$ 512-pixel images "Lena" and "Barbara." In order to improve the graphical presentation,

(a)

(b)

(c)

(d)

Fig. 4. Subjective evaluation of DWT-JPEG, using $S = 3$. Filter bank and bit rate are indicated. (a) Original, (b) 1.0 b/p—B11/13, (c) 0.5 b/p—B11/13, and (d) 0.5 b/p—B3/5.

it is shown plots of peak signal-to-noise ratio (PSNR) improvements led by replacing the Haar transform by each filter bank, for several bit rates. Therefore, the top plots in Fig. 3 refer to differences in PSNR relative to the performance of the Haar bases. We can see that the group composed by J24, J16, B7/9, and B11/13 consistently performs better than J8, ELT, and B3/5. Also, J24 consistently performs equal to or better than J12 and J16, while B11/13 performs better than B7/9. Thus, it may be a good guess to narrow the selection to J24 and B11/13 in terms of PSNR.

In order to decide number of levels, we compare DWT-JPEG with DCT-JPEG. On the bottom of Fig. 3 we show results for DWT-JPEG relating the PSNR obtained with DWT-JPEG to the PSNR obtained using regular DCT-JPEG. Thus, the graphics show PSNR differences, i.e., the improvements led by replacing the DCT with the DWT. Note that JPEG demands a minimal bit rate which is decided by the minimum number of bits encoded for each block. This yields a minimum bit rate slightly lower than 0.1 bit-per-pixel (b/p) for default luminance Huffman tables and 64-pixel blocks. Using four levels, DWT blocks have 16×16 coefficients. Hence, the minimum bit rate is a quarter of that using the 8×8 DCT or three-level DWT. Therefore, a better performance of a four-level DWT-JPEG is expected compared to a three-level DWT-JPEG at lower bit rates. For image "Lena," the three-level DWT-JPEG performs 1 dB better than DCT-JPEG at virtually all bit rates. For image "Barbara," the three-level DWT is slightly better than the four-level one above 0.5 b/p, and both are more than 2 dB superior to DCT-JPEG above 0.5 b/p.

Fig. 5. Subjective evaluation of DWT-JPEG, using $S = 3$. Filter bank and bit rate are indicated. (a) DCT-JPEG 0.5 b/p, (b) DCT-JPEG 1.0 b/p, (c) 0.5 b/p—B11/13, (d) 1.0 b/p—B11/13, (e) 0.5 b/p—B3/5, and (f) 1.0 b/p—B3/5.

In general, their performances are very close. Because of the block size, it is preferable to use three levels.

As mentioned earlier, the goal of this paper is not the development of a state-of-the-art coder at the cost of greater complexity, but to obtain improvement maintaining most of the JPEG structure. Embedded coders require buffering the whole image and Shapiro's EZW coder [7] is much more complex than the proposed DWT-JPEG. Nevertheless, we included results for the EZW coder along with plots for the DWT-JPEG in the bottom of Fig. 3. Although much simpler, the proposed coder performs slightly worse than EZW for image "Lena" and is comparable or superior to EZW for image "Barbara." We also present reconstructed images using the proposed

DWT-JPEG coder and image "Barbara" for subjective evaluation in Figs. 4 and 5. Other subjective and objective tests were carried but omitted due to space limitations.

IV. Conclusions

The DWT-JPEG was shown to outperform baseline JPEG approaching the performance of more sophisticated and complex DWT-based coders. It is worthwhile to mention that by eliminating encoding complexity constraints, JPEG can be optimized and still be decoder compatible [17], [18]. Both quantizer and Huffman tables can be optimized and optimal coefficient thresholding can be applied. In fact, all three processes can be jointly optimized [17], [18]. These techniques can largely improve JPEG performance and can also be applied to DWT-JPEG [18]. Studies in this sense are in an early stage and preliminary results show that optimized DWT-JPEG performs consistently better than an EZW coder [7]. Coder optimization is a topic for further research.

References

[1] K. R. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications.* San Diego, CA : Academic, 1990.

[2] W. B. Pennebaker and J. L. Mitchell, *JPEG: Still Image Compression Standard*. New York, NY: Van Nostrand Reinhold, 1993.

[3] M. Vetterli and J Kovacevic, *Wavelets and Subband Coding*. Englewood Cliffs, NJ: Prentice Hall, 1995.

[4] M. Antonini and M. Barlaud, "Image coding using wavelet transform," *IEEE Trans. Image Processing,* vol. 1, pp. 205–220, Apr. 1992.

[5] R. L. Joshi, V. J. Crump, and T. R. Fisher, "Image subband coding using arithmetic coded trellis coded quantization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, pp. 515–523, Dec. 1995.

[6] R. L. Joshi, T. R. Fisher, and R. H. Bamberger, "Optimal classification in subband coding of images," in *Proc. 1994 IEEE Int. Conf. Image Processing,* vol. II, Austin, TX, 1994, pp. 883–887.

[7] J. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing,* vol. 41, pp. 3445–3462, Dec. 1993.

[8] A. Said and W. A. Pearlman, "A new fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 243–250, June 1996

[9] C. Herley, Z. Xiong, K. Ramchandran, and M. Orchard, "An efficient algorithm to find a jointly optimal time-frequency segmentation using time-varying filter banks," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Detroit, MI, vol. II, 1995, pp. 1516–1519.

[10] P. Sriram and M. Marcellin, "Image coding using wavelet transforms and entropy-constrained trellis coded quantization," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Minneapolis, MN, vol. V, 1993, pp. 554–557.

[11] J. Bradley, C. Brislawn, and T. Hopper, "The FBI wavelet/scalar quantization standard for gray-scale fingerprint image compression," in *Proc. SPIE VCIP,* Orlando, FL, vol. 1961, 1993, pp. 293–304.

[12] P. Y. Chen and C. J. Kuo, "Feature preserving wavelet scheme for low bit rate coding," presented at 1995 IS&T/SPIE Symp. Electronic Imaging: Science and Technology, San Jose, CA, Feb. 1995.

[13] J. D. Johnston, "A filter family designed for use in quadrature mirror filter banks," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Denver, CO, pp. 291–294, 1980.

[14] D. Le Gall and A. Tabatabai, "Sub-band coding of digital images using symmetric short kernel filters and arithmetic coding techniques" in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, 1988, pp. 761–764.

[15] G. Strang and T. Q. Nguyen, *Wavelets and Filter Banks.* Wellesley-Cambridge, 1996.

[16] H. S. Malvar, "Fast computation of wavelet transforms with the extended lapped transform," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, San Francisco, CA, vol. IV, 1992, pp. 393–396.

[17] M. Crouse and K. Ramchandran, "Joint thresholding and quantizer selection for decoder compatible baseline JPEG," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Detroit, MI, vol. 4, 1995, pp. 2331–2334.

[18] M. Crouse and K. Ramchandran, "Joint thresholding and quantizer selection for transform image coding: entropy constrained analysis and applications to baseline JPEG," under review for *IEEE Trans. Image Processing.*

Very Low Bit-Rate Color Video Coding Using Adaptive Subband Vector Quantization with Dynamic Bit Allocation

Stathis P. Voukelatos and John J. Soraghan

***Abstract*— In this correspondence a novel adaptive vector quantization (VQ) based subband coding scheme for very low bit rate coding of video sequences is presented. Overlapped block motion estimation/compensation is employed to exploit interframe redundancy. A two-dimensional (2-D) wavelet transform (WT) is applied to the resulting displaced frame difference (DFD) signal. The WT coefficients are encoded using an adaptive subband vector quantization (ASBVQ) scheme in combination with a dynamic bit allocation strategy based on marginal analysis. Fixed rate coding is provided. Comparative experimental results of the ASBVQ codec with the H.261 and the recently defined H.263 video coding standards are given.**

***Index Terms*—Image coding, vector quantization, wavelet transforms.**

I. Introduction

The potential applications of very low bit-rate video coding in a number of forthcoming visual services has led to an increased research and standardization effort in the area. The well established ITU-T (former CCITT) H.261 [1] recommendation is aimed primarily at videophone and videoconferencing services at transmission rates of $p \times 64$ kb/s with $p = 1, \cdots, 30$. In very low bit-rate environments, below 30 kb/s, such as the public switched telephone network (PSTN), the performance of the H.261 coder is inadequate. The reconstructed video sequences contain annoying blocking artifacts, due to the block-based DCT transform coding employed in the H.261 codec. Recently, the ITU-T H.263 [2] recommendation was defined aiming at videotelephony applications over the PSTN.

The wavelet transform has been shown to be an efficient coding method for still images and video. Unlike block transform-based coding, it does not suffer from blocking artifacts and hence it is able to produce better subjective quality especially at low bit rates. In typical subband based video coding, a two-dimensional (2-D) wavelet transform is applied in the spatial domain and some form of motion-compensated prediction is performed to exploit redundancy between adjacent frames. In [3] the multiresolution motion estimation was introduced, where motion vectors are estimated in the wavelet transform (WT) domain using a hierarchical approach. This method is also used in [4] in combination with an adaptive bit plane run length coding scheme for the quantization of the subband error signals. Alternatively, full resolution motion estimation can be performed on

Manuscript received February 14, 1995; revised October 2, 1995. This paper was recommended by Associate Editor B. G. Haskell. This work was supported by the GEC-Marconi Hirst Research Center and a University of Strathclyde Studentship.

The authors are with the Signal Processing Division, Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow G1 1XW, U.K.

Publisher Item Identifier S 1051-8215(97)01135-X.