A Comparative Study of DCT- and Wavelet-Based Image Coding

Zixiang Xiong, Kannan Ramchandran, Michael T. Orchard, and Ya-Qin Zhang

Abstract—**We undertake a study of the performance difference of the discrete cosine transform (DCT) and the wavelet transform for both image and video coding, while comparing other aspects of the coding system on an equal footing based on the state-of-the-art coding techniques. Our studies reveal that, for still images, the wavelet transform outperforms the DCT typically by the order of about 1 dB in peak signal-to-noise ratio. For video coding, the advantage of wavelet schemes is less obvious. We believe that the image and video compression algorithm should be addressed from the overall system viewpoint: quantization, entropy coding, and the complex interplay among elements of the coding system are more important than spending all the efforts on optimizing the transform.**

Index Terms—**Discrete cosine transform (DCT), image coding, video coding, wavelet transform.**

I. Introduction

TRANSFORM coding has become the de facto standard paradigm in image (e.g., JPEG [1], [2]) and video coding (e.g., MPEG-2 [3] and H.263 [4]), where the discrete cosine transform (DCT) is used because of its nice decorrelation and energy compaction properties [5]. In recent years, much of the research activities in image coding have been focused on the discrete wavelet transform. While the good results obtained by wavelet coders (e.g., the embedded zerotree wavelet (EZW) coder [6] and the set partitioning in hierarchical trees (SPIHT) coder [7]) are partly attributable to the wavelet transform, we emphasize that much of the performance gain is obtained by carefully designing quantizers (e.g., zerotree quantizer) that are tailored to the transform structure. We have seen in publications where many authors compare their best wavelet-based coding scheme with the worst DCT-based scheme (e.g., baseline JPEG). This often gives readers a distorted perspective of the issues involved in image coding.

In this paper, we highlight the coding gain of the wavelet transform over the DCT while comparing the other aspects of the system design on an equal footing. This allows us to address the real issues involved in image coding, which are quantization and entropy coding rather than the difference caused by the DCT and the wavelet transform. We feel that this viewpoint is not very well represented in the image-coding community.

Manuscript received April 2, 1998; revised January 14, 1999. This paper was recommended by Associate Editor J. Brailean.

Z. Xiong is with the Department of Electrical Engineering, University of Hawaii, Honolulu, HI 96822 USA (e-mail: zx@lena.eng.hawaii.edu).

K. Ramchandran is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA.

M. T. Orchard is with the University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA.

Y.-Q. Zhang is with Microsoft Research, Beijing 100080 China.

Publisher Item Identifier S 1051-8215(99)07275-4.

We first point out that the baseline JPEG coding results that many people use as the performance benchmark are far from the best that JPEG offers. Much better performance can be obtained with JPEG by optimal quantization matrix (Q-matrix) design [8], [9] and coefficient thresholding [10], while being compatible with the JPEG syntax. Note that, for the standard 512 $\times$ 512 *Lena* image at a moderate bit rate, the optimal JPEG coder in [11] gives even better peak signal-to-noise ratio (PSNR) than the original EZW coder proposed by Shapiro [6].

We then compare DCT- and wavelet-based image coding based on results in the literature for the 512 $\times$ 512 *Lena* and *Barbara* images. For a fair comparison, we have to weigh in complexity and performance. It is well acknowledged that hardware (or software) implementation of the DCT is less expensive than that of the wavelet transform. For example, the most efficient algorithm for the two-dimensional (2-D) 8 $\times$ 8 DCT requires only 54 multiplications [12], while the complexity of calculating the discrete wavelet transform depends on the length of the wavelet filters, which is at least one multiplication per coefficient. We hence focus on the performance comparison of DCT- and wavelet-based coders in this paper.

To accurately characterize the coding performance difference due to the transform (wavelet versus DCT), we have to keep the quantizer (and entropy coder) the same. Two representative quantizers are used: the uniform quantizer in baseline JPEG [1] and the zerotree quantizer in SPIHT [7]. We pick the SPIHT coder because it is an improved version of Shapiro's original EZW coder. We will explain in detail key differences between the two embedded coders.

The actual performance comparison is carried out by citing results from two coders: the wavelet-based JPEG-like coder proposed by de Queiroz *et al.* [13], where only the DCT in baseline JPEG is replaced by a three-level wavelet transform, and the DCT-based embedded image coder described in [14], where a zerotree quantizer is used to quantize 8 $\times$ 8 DCT coefficients. Results from these coders show that, when the same quantizer (and entropy coder) is used, the wavelet transform (using the 7/9 biorthogonal wavelet filters [15]) gives only 0.6–1.0-dB gain over the DCT at the same bit rate.

We also compare the DCT and the wavelet transform for video coding. Both 2-D and three-dimensional (3-D) transform-based approaches are considered. Conventional 2-D transform-based approaches apply the DCT or the wavelet transform on the motion compensated residuals. We compare the difference of the DCT and the wavelet transform based on the MPEG-4 research activities, where both DCT- and wavelet-based schemes are proposed for low-bit-rate video coding. We benchmark the performance of the wavelet-based video coder proposed by Sarnoff Corp. [16] against that of the

TABLE I
PERFORMANCE COMPARISON OF BASELINE JPEG [2] AND ITS OPTIMIZED VERSION [11]

Rate	PSNR (dB)			
	Baseline JPEG [2]		Optimized JPEG [11]	
(b/p)	Lena	Barbara	Lena	Barbara
0.25	31.60	25.20	32.30	26.70
0.50	34.90	28.30	35.90	30.60
0.75	36.60	31.00	38.10	33.60
1.00	37.90	33.10	39.60	35.90

TABLE II
PERFORMANCE COMPARISON OF SHAPIRO'S EZW CODER [6] AND THE SPIHT CODER [7] PROPOSED BY SAID AND PEARLMAN

Rate	PSNR (dB)			
	EZW [6]		SPIHT [7]	
(b/p)	Lena	Barbara	Lena	Barbara
0.125	30.23	24.03	31.09	24.85
0.25	33.17	26.77	34.11	27.58
0.50	36.28	30.53	37.21	31.39
0.75			39.04	34.25
1.00	39.55	35.14	40.40	36.41

current MPEG-4 verification model (VM), which is a slightly modified DCT-based H.263 coder. Sarnoff's wavelet-based video coder gives comparable performance to the VM.

In 3-D subband video coding [17]–[21], in addition to spatial domain subband filtering of each frame, temporal domain subband filtering is also used to exploit redundancies in consecutive frames with or without motion compensation (MC). Kim and Pearlman recently extended the SPIHT algorithm from 2-D images to 3-D video and developed a 3-D SPIHT video coder [22], where segments of 16 consecutive frames are grouped together for temporal domain wavelet filtering. The 3-D SPIHT video coder also has an option for MC. At the same bit rate and frame rate, the average PSNR's given by the coder in [22] without MC are about 1.4–1.8 dB less than those given by the H.263 coder, which requires much more computation. For comparison purposes, we replace the wavelet transform in the 3-D SPIHT video coder with the 8×8 DCT and build a 3-D DCT-based SPIHT video coder. Experiments indicate that the average PSNR difference between using the 3-D wavelet transform and the 3-D DCT is about 0.5 dB. We believe that more research is needed in both 2-D and 3-D wavelet-based video coding, especially in the areas of optimal MC and transform structure.

II. DCT-BASED JPEG IMAGE CODING

The basic components of the JPEG standard [2] are the DCT transform, scalar quantization, zig-zag scan, and Huffman coding. It has long been realized that the current JPEG standard does not provide state-of-the-art coding performance. Several methods have been proposed to improve upon JPEG, including optimal Q-matrix design [8], [9], optimal thresholding [10], and joint optimization [11]. In Table I, we tabulate the coding results in PSNR of the baseline JPEG for *Lena* and *Barbara*. To highlight the best results one can get while being compatible with the JPEG standard, we include in Table I results from the optimal JPEG coder described in [11]. The gain from optimizing JPEG can be as much as 1.7 dB at the same bit rate. This clearly indicates that the baseline JPEG is far from optimal. Results in Table I also show that the optimal JPEG coder is capable of surpassing Shapiro's embedded zerotree wavelet (EZW) coder in performance.

III. WAVELET IMAGE CODING

Recent years have witnessed explosive growth in research activities involving wavelet image coding [15]. Earlier related work in subband image coding showed the potential coding gain of subband/wavelet image coding, which depends on the spectrum flatness of the input image [23]. An example of wavelet image coding is Shapiro's EZW coder [6]. The main contribution of Shapiro work is zerotree quantization of wavelet coefficients, which works by efficiently predicting the children nodes based on the significance/insignificance of their parent. An embedded zerotree quantizer refines each input coefficient sequentially using a bitmap type of coding scheme, and it stops when the size of the encoded bitstream reaches the exact target bit rate.

Said and Pearlman described an SPIHT coder in [7] that achieves about 1 dB gain in PSNR over Shapiro's original coder at the same bit rate for typical images (see Table II). The better performance of SPIHT coder is due to the following three reasons:

- better wavelet filters (7/9 biorthogonal wavelet filters instead of length-9 QMF filters);
- special symbol for the significance/insignificance of child nodes of a significant parent;
- separation of the significance of child (direct descendant) nodes from that of the grandchild nodes.

IV. COMPARISON OF DCT- AND WAVELET-BASED IMAGE CODING

We now proceed to compare the DCT and the wavelet transform for image coding. To this end, we fixed the quantizer (and entropy coder) while allowing the transform to vary, as this is the only way to provide accurate assessment of the coding efficiencies of wavelet versus DCT.

A. Wavelet-Based JPEG-Like Image Coding

When the wavelet transform is coupled with the baseline JPEG quantizer, the resulting coder becomes the one described in [13], where only the DCT in baseline JPEG is replaced by a three-level wavelet transform. The wavelet coefficients are rearranged into wavelet blocks and scanned into vectors before scalar quantization and Huffman coding (see Fig. 1). A gain of about 1 dB was reported in [13] for *Lena* with the wavelet-based JPEG-like coder over the baseline JPEG.

B. DCT-Based Embedded Image Coding

If we fix the SPIHT quantizer and use it to quantize the DCT coefficients, we will have a DCT-based embedded image coder. Such a coder is described in [14]. The observation in [14] is that an 8×8 DCT image representation can be thought of as a 64-subband decomposition, and that we can treat each 8×8 DCT block as a depth-three tree of coefficients. After

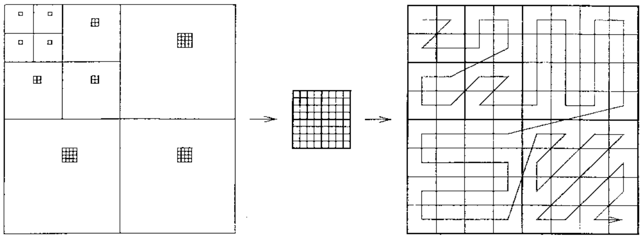

Fig. 1. In the wavelet-based JPEG-like image coder [13], the three-level wavelet transform coefficients are rearranged into blocks and scanned into vectors before scalar quantization and Huffman coding.

0	1	4	5	16	17	20	21
2	3	6	7	18	19	22	23
8	9	12	13	24	25	28	29
10	11	14	15	26	27	30	31
32	33	36	37	48	49	52	53
34	35	38	39	50	51	54	55
40	41	44	45	56	57	60	61
42	43	46	47	58	59	62	63

Fig. 2. In the DCT-based embedded image coder [14], an 8 $\times$ 8 DCT block is treated as a depth-three tree of coefficients.

labeling the 64 DCT coefficients in each block as in Fig. 2, we can identify the parent–children relationships between DCT coefficients as follows: the parent of coefficient i is $\lfloor i/4 \rfloor$ for $1 \leq i \leq 63$, while the set of four children associated with coefficient j is $\{4j, 4j+1, 4j+2, 4j+3\}$ for $1 \leq j \leq 15$. The dc coefficient "0" is the root of the DCT coefficient tree, which has only three children: coefficients 1, 2 and 3.

In the DCT-based embedded coder [14], no more processing (DPCM, DCT, or wavelet transform) is applied to the dc image. For fair comparison, we set the number of wavelet transform levels to three in the SPIHT coder so that the size of the lowest band for both coders is the same. Coding results from both coders are given in Table III. One way to improve the coder in [14] is to apply an 8 $\times$ 8 DCT to the dc image. The coefficient tree structure for this case can be similarly constructed as for the wavelet case. The advantage of further transforming the dc image is better compression at lower bit rate due to energy compaction into fewer dc coefficients. The two-layer 8 $\times$ 8 DCT transform structure can be thought of as an equivalent of six-level wavelet transform, which is used in [7] for the 512 $\times$ 512 *Lena* and *Barbara* images. Results from the improved DCT-based embedded coder and those reported in [7] are given in Table IV.

TABLE III
PERFORMANCE COMPARISON OF THE DCT-BASED EMBEDDED IMAGE CODER [14] AND THE SPIHT CODER [7] WHEN A THREE-LEVEL WAVELET TRANSFORM IS USED

Rate	PSNR (dB)			
	SPIHT with 3-level wavelet		Embedded DCT [17] (8 x 8 DCT only)	
(b/p)	Lena	Barbara	Lena	Barbara
0.125	30.13	24.16	28.50	24.07
0.25	33.53	27.09	32.27	26.93
0.50	36.90	31.07	35.98	30.87
0.75	38.86	34.00	38.04	33.73
1.00	40.23	36.17	39.60	36.08

TABLE IV
PERFORMANCE COMPARISON OF THE SPIHT CODER [7] AND THE DCT-BASED EMBEDDED IMAGE CODER [14] WITH ADDITIONAL 8 $\times$ 8 DCT OF THE DC IMAGE

Rate	PSNR (dB)			
	SPIHT [7]		Embedded DCT + DCT of the DC image	
(b/p)	Lena	Barbara	Lena	Barbara
0.125	31.09	24.85	29.32	24.51
0.25	34.11	27.58	32.63	27.26
0.50	37.21	31.39	36.13	31.10
0.75	39.04	34.25	38.14	33.93
1.00	40.40	36.41	39.69	36.21

The DCT-based coder has lower complexity than its wavelet-based counterpart. The loss in performance for using DCT instead of the wavelet transform is only about 0.7 dB for *Lena* at 1 b/p (see Tables III and IV), although the performance gap widens as the bit rate decreases. The remarkable thing about the DCT-based embedded coder is that it give better PSNR's over those from both JPEG and Shapiro's EZW coders.

V. COMPARISON OF DCT- AND WAVELET-BASED VIDEO CODING

Standard video coding methods are based on the framework of applying the DCT to the motion-compensated residual images [3], [4]. Wavelet-based video coding has been an intriguing research topic. One approach is to replace the DCT in standard methods by the wavelet transform. The key difference in this 2-D transform (DCT or wavelet)-based video coding approach is that the motion-compensated residue images have quite different statistics from most natural still images, namely, less spatial correlation after motion compensation. This renders the rationale behind using 2-D linear transformations for decorrelation and energy compaction less relevant. Other predominant issues involved in 2-D wavelet-based video coding have to do with what is a good paradigm for motion compensation (e.g., block based or hierarchically based). Driven by the huge potential applications of digital video and the MPEG-4 standardization effort in low-bit-rate video coding, many researchers have applied wavelets to the coding of motion-compensated residue images [16], [24]. In Table V, we quote results from the wavelet-based zerotree entropy (ZTE) video coder [16] proposed by Sarnoff Corp. to MPEG-4 and compare them with those from the MPEG-4 verification model [25], which is a slightly modified DCT-based H.263 coder. We see that the wavelet-based ZTE coder

TABLE V
PERFORMANCE COMPARISON OF SARNOFF'S WAVELET-BASED ZTE [16] CODER AND MPEG-4'S DCT-BASED CODER IN THE VM [25] FOR LOW-BIT-RATE VIDEO CODING

Sequence	Bitrate	Y/C	VM [20]	ZTE [19]
Akiyo	24 kb/s	Y	37.46	36.64
@ 10 f/s		C	42.15	44.02
Hall Monitor	24 kb/s	Y	34.46	34.11
@ 10 f/s		C	39.38	39.63
Coast	48 kb/s	Y	29.74	29.20
@ 7.5 f/s		C	40.78	40.88
News	48 kb/s	Y	35.10	35.17
@ 7.5 f/s		C	39.11	40.46

TABLE VI
PERFORMANCE COMPARISON OF 3-D SPIHT VIDEO CODERS USING THE WAVELET TRANSFORM AND THE DCT WITHOUT MOTION COMPENSATION

Sequence (QCIF)	Bitrate (kb/s)	H.263 (tmn2.0) (dB)	3D WT SPIHT (dB)	3D DCT SPIHT (dB)
Mother Daughter	16.41	34.06	32.27	31.87
News	35.49	32.40	31.02	30.60
Foreman	20.97	27.21	25.85	25.21
Container	22.08	32.28	30.50	30.28

produces comparable objective performance to the DCT-based coder in the VM at the same bit rate and frame rate.

Recently, there has been active research in 3-D subband/wavelet-based video coding [17]–[21]. Choi and Woods [21] reported better results than MPEG-1 using a 3-D subband approach together with a hierarchical variable-size block-based MC scheme. The 3-D version of the SPIHT video coder [22] was shown to give performance slightly inferior to the H.263 coder, but with much lower complexity when MC is not used. To compare the difference between the DCT and the wavelet transform for video coding, we extend the DCT-based embedded image coder described in Section IV-B to 3-D video and build a 3-D DCT-based SPIHT video coder, which can be thought of as a DCT-based variant of the SPIHT video coder in [22]. Objective PSNR results of the Y-component (averaged over 96 frames) given by these two video coders are tabulated in Table VI for coding four QCIF sequences at 10 frames/s. From Table VI, we see that the PSNR difference between DCT- and wavelet-based video coding is about 0.5 dB, which is less than the image case.

VI. CONCLUSION

We have carried out a comparative study of DCT- and wavelet-based coding for both still images and video sequences. Based on empirical performance results, we illustrate that the main factors in image coding are the quantizer and entropy coder rather than the difference between the wavelet transform and the DCT. For still-image coding, the difference between the wavelet transform and the DCT is less than 1 dB, and it is even smaller for video coding.

REFERENCES

[1] G. K. Wallace, "The JPEG still-picture compression standard," *Commun. ACM*, vol. 34, pp. 30–44, Apr. 1991.
[2] W. B. Pennebaker and J. L. Mitchell, *JPEG Still Image Data Compression Standard*. New York: Van Nostrand Reinhold, 1992.
[3] "MPEG-2 video," ITU-T Recommendation H.262-ISO/IEC 13818-2, Jan. 1995.
[4] "Video coding for low bitrate communication," ITU-T Recommendation H.263, Dec. 1995.
[5] K. R. Rao and P. Yip, *Discrete Cosine Transform*. New York: Academic, 1990.
[6] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing*, vol. 41, pp. 3445–3463, Dec. 1993.
[7] 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.
[8] A. Hung and T. Meng, "Optimal quantizer step sizes for transform coders," in *Proc. ICASSP'91*, Apr. 1991, pp. 2621–2624.
[9] S. Wu and A. Gersho, "Rate-constrained picture adaptive quantization for JPEG baseline coders," in *Proc. ICASSP'93*, Apr. 1993, vol. 5, pp. 389–392.
[10] K. Ramchandran and M. Vetterli, "Rate-distortion optimal fast thresholding with complete JPEG/MPEG decoder compatibility," *IEEE Trans. Image Processing*, vol. 3, pp. 700–704, Sept. 1994.
[11] M. Crouse and K. Ramchandran, "Joint thresholding and quantizer selection for transform image coding: Entropy-constrained analysis and applications to JPEG," *IEEE Trans. Image Processing*, vol. 6, pp. 285–297, Feb. 1997.
[12] E. Feig, "A fast scaled DCT algorithm," in *Proc. SPIE Image Processing Algorithms and Techniques*, Feb. 1990, vol. 1244, pp. 2–13.
[13] R. de Queiroz, C. Choi, Y. Huh, and K. Rao, "Wavelet transforms in a JPEG-like image coder," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp. 419–424, Apr. 1997.
[14] Z. Xiong, O. Guleryuz, and M. T. Orchard, "A DCT-based embedded image coder," *IEEE Signal Processing Lett.*, vol. 3, pp. 289–290, Nov. 1996.
[15] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Processing*, vol. 1, pp. 205–221, Apr. 1992.
[16] S. A. Martucci, I. Sodagar, T. Chiang, and Y.-Q. Zhang, "A zerotree wavelet video coder," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp. 109–118, Feb. 1997.
[17] G. Karlsson and M. Vetterli, "Three dimensional subband coding of video," in *Proc. ICASSP'98*, Apr. 1988, pp. 1100–1103.
[18] J. R. Ohm, "Three-dimensional subband coding with motion compensation," *IEEE Trans. Image Processing*, vol. 3, pp. 559–571, Sept. 1994.
[19] D. Taubman and A. Zakhor, "Multirate 3D subband coding of video," *IEEE Trans. Image Processing*, vol. 3, pp. 572–588, Sept. 1994.
[20] C. I. Podilchuk, N. S. Jayant, and N. Farvardin, "Three-dimensional subband coding of video," *IEEE Trans. Image Processing*, vol. 4, pp. 125–139, Feb. 1995.
[21] S. J. Choi and J. W. Woods, "Motion-compensated 3D subband coding of video," *IEEE Trans. Image Processing*, vol. 8, pp. 155–167, Feb. 1999.
[22] B.-J. Kim and W. A. Pearlman, "An embedded wavelet video coder using three-dimensional set partitioning in hierarchical trees (SPIHT)," in *Proc. IEEE DCC'97*, 1997, pp. 251–260.
[23] J. W. Woods (Ed.), *Subband Image Coding*. Norwell, MA: Kluwer Academic, 1991.
[24] Y.-Q. Zhang and S. Zafar, "Motion-compensated wavelet transform coding for color video compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 2, pp. 285–296, Sept. 1992.
[25] "Very low bitrate video codec," Doc. ISO/IEC MPEG96/M0637, Munich MPEG Meeting, Jan. 1996.