Hyperspectral Image Compression with the 3D Dual-Tree Wavelet Transform

Joseph B. Boettcher, Qian Du, and James E. Fowler
Department of Electrical and Computer Engineering
GeoResources Institute
Mississippi State University, Mississippi State, MS 39762

Abstract— The complex dual-tree discrete wavelet transform is explored for the coding of hyperspectral imagery using a coder that has previously demonstrated efficient video-coding performance. A noise-shaping process increases the sparsity of the redundant transform-coefficient set, resulting in a high degree of regional coherency within the coefficient subbands. This coherency, as well as correlation across subbands, is exploited by a coding algorithm that performs set-partitioning using k-d trees. Prior experiments have shown that the proposed set-partitioning algorithm outperforms state-of-the-art JPEG2000 when coding video. However, experimental results indicate that the same coder fails to show similar gains when coding hyperspectral data, suggesting that hyperspectral data does not have properties that can be exploited by the increased directionality of the dual-tree transform.

I. INTRODUCTION

Due to the typically large size of hyperspectral imagery, compression is necessary to reduce data volumes to a size suitable for storage and transmission. Recent research has explored the use of compression schemes based on 3D wavelet transforms for the coding of hyperspectral data. Such schemes exploit the high degree of spatial and spectral correlation found in hyperspectral images by applying wavelet transforms in both the spatial and spectral directions. Afterward, a coding algorithm operates on the resulting transform coefficients to produce an embedded bitstream, allowing progressive transmission of the hyperspectral data. See [1] for an overview of such 3D wavelet-based coders for hyperspectral imagery.

While most wavelet-based compression techniques employ the traditional critically sampled discrete wavelet transform (DWT), alternative wavelet transforms have recently been proposed. Specifically, the complex dual-tree discrete wavelet transform (DDWT) [2–4] has undergone investigation in 3D video-coding systems [5–9]. The DDWT is a redundant, or overcomplete, transform that, in the 3D case, produces four times as many subbands as the DWT, with each subband oriented in a different direction. The additional subbands motivated the use of the DDWT in video compression, since the different orientations help isolate signal features moving in a number of different directions. Although the 3D DDWT is four times redundant, an iterative projection-based noise-shaping process [10] can increase the sparsity of the coefficient set so that the number of significant coefficients needed to accurately describe the signal is greatly reduced.

Prior research has shown that video-coding systems based on the DDWT outperform similar DWT systems [6, 8, 9]. Since video sequences and hyperspectral images can both be represented and coded as 3D volumes, an investigation of the DDWT in hyperspectral image compression is warranted to determine if similar gains are possible. In this paper, we extend our DDWT-BISK algorithm [9] to code hyperspectral imagery. Because of the increased directionality inherent to the 3D DDWT, the DDWT-BISK algorithm does not perform explicit motion compensation or any other process specific to video, making an extension to hyperspectral data straightforward.

In the following sections, we first overview complex wavelet transforms before describing the DDWT-BISK coder in detail. Then we present experimental results concerning the coding of hyperspectral data with DDWT-BISK, followed by an analysis of those results.

II. COMPLEX WAVELET TRANSFORMS

In order to overcome shortcomings of the traditional critically sampled DWT, Kingsbury [2] introduced the DDWT consisting of two trees of real wavelet filters operating on the same data in parallel, with the filters designed such that the two trees produce the real and imaginary parts of the complex-valued coefficients. While the DWT lacks shift invariance, the DDWT is approximately shift invariant and offers higher directional selectivity. However, 2^m:1 redundancy is added for an m-dimensional signal.

Selesnick and Li [3] developed a 3D version of the DDWT to provide a useful representation for video. It turns out that the degree of redundancy can be reduced without sacrificing perfect reconstruction by simply discarding the complex parts of the coefficients, resulting in 4:1 redundancy. For this real-valued transform, four separable transforms based on Hilbert pairs are applied to the original signal, and only the real parts of the coefficients are retained. The four sets of transform data are then combined with linear operations to produce subbands that isolate edges in a variety of orientations. The resulting DDWT subbands are arranged in four separate *transform combinations* with each combination having the same subband organization as would a 3D DWT of the original data but with each combination containing subbands of different orientation. For example, a 3D dyadic DWT consists of 7 highpass subbands at each resolution level, plus a single baseband at the lowest resolution. Consequently, at each resolution level, the 3D DDWT consists of $4 \times 7 = 28$ highpass subbands, except

at the lowest resolution, which contains 4 baseband subbands. This 3D DDWT decomposition is illustrated in Fig. 1.

Although the 3D DDWT produces four times the data that the 3D DWT does, the DDWT requires fewer critical coefficients to efficiently represent the underlying signal [5]. In fact, Reeves and Kingsbury [10] proposed deliberately reducing the number of DDWT coefficients by discarding small-magnitude coefficients and refining the remaining coefficients to compensate. This "noise-shaping" procedure is an iterative projection between the original-signal domain and the DDWT domain. On each iteration, the signal is transformed into the DDWT domain, wherein it undergoes thresholding to remove small coefficients. The signal is then transformed back into its original domain, and the amount of error induced by the thresholding is computed. This error is then transformed into the DDWT domain and added back to the thresholded signal. The threshold decreases with each iteration until the desired number of coefficients or level of reconstruction quality is reached. This noise-shaping method increases the sparsity of the representation to the point that the 3D DDWT typically requires fewer non-zero coefficients than the 3D DWT to achieve the same level of reconstruction quality, as was observed for video in [5]. The video coder proposed in [6, 8] extends this observation to actual coding results that include quantization and entropy coding.

III. DDWT-BISK

In [9], we introduced the DDWT-BISK coder for video coding. This coder improved upon prior 3D-DDWT coding attempts by successfully exploiting the spatiotemporal (or in the hyperspectral case, spatiospectral) coherency of insignificant-coefficient regions within a given subband. In order to efficiently code coherent regions of DDWT coefficients, the DDWT-BISK coder uses a modified version of the BISK algorithm [11–13]. BISK performs bitplane coding in which significant coefficients are located by recursive partitioning of the 3D dataset. Specifically, k-d trees [14] are used to split sets of coefficients into two subsets of roughly equal size. Once a significant coefficient is located, its sign information is coded, and its magnitude is refined on successive passes. Significance, sign, and magnitude-refinement information are all coded with adaptive arithmetic coding.

The DDWT-BISK coder begins by applying the noise shaping of [10] to a 3D dataset to produce the sparse DDWT coefficients. Then, a modified version of the 3D-BISK algorithm [12, 13] operates on the transform coefficients to produce the final coded bitstream. First, coefficients are grouped into 4-dimensional vectors, where each vector consists of the four coefficients at the same spatiospectral location in the same subband from each of the four DDWT transform combinations, as illustrated in Fig. 1. These coefficient vectors are then assembled into sets spanning the entire 3D subbands, producing 7 sets of vectors at each resolution level (4 sets of vectors at the baseband level), assuming a dyadic decomposition structure. All the sets are placed in the list of insignificant sets (LIS).

The algorithm then performs bitplane coding with *sorting* and *refinement* passes. In the sorting pass, sets in the LIS are tested against the current threshold to determine the significance of the set as a whole—if the magnitude of any coefficient in the set is above the threshold, the set is significant. Significant sets are split in two along the longest dimension of the set. The resulting subsets are added back to the LIS as two new sets to be recursively tested and split if necessary. Eventually, a significant set will be reduced to a single four-coefficient vector in which at least one of the four coefficients will be significant. At this point, the vector is removed from the LIS, and a significance symbol is output to denote which coefficients in the vector are significant and which are not. The significant coefficients from the vector are then added to the list of significant pixels (LSP), while the insignificant coefficients are added to the list of insignificant pixels (LIP). After each sorting pass, the LIP is processed by comparing each coefficient to the current threshold and outputting the significance state. If a coefficient in the LIP becomes significant, it is transferred to the LSP. The refinement pass then processes each coefficient in the LSP and outputs the current bitplane value of the coefficient magnitude. Sorting and refinement passes continue until the target bitstream length has been reached.

For the DDWT-BISK system, we consider two types of decomposition structure for the 3D DDWT as illustrated in Figs. 2(a) and (b). Fig. 2(a) gives a traditional dyadic decomposition wherein the wavelet transform is applied to only the lowpass band at each successive level of decomposition. Alternatively, we also consider a wavelet-packet decomposition, shown in Fig. 2(b) as the "anisotropic" structure, in which a full J-scale 1D wavelet transform is applied to each dimension of the 3D dataset separately. This anisotropic transform structure generates a greater number of subbands than does a dyadic structure for the same number of decomposition levels; in the context of the 3D DDWT, these additional subbands can provide additional directional orientations and can thus increase the degree of directional selectivity. We note that a dyadic decomposition was used in the original 3D DDWT development [3] and in the video coder of [6, 8] while the anisotropic structure was initially investigated in [7].

Our prior experiments with DDWT-BISK have shown that it can outperform JPEG2000 for the 3D coding of video [9]. Table I illustrates PSNR results that were obtained from coding video with DDWT-BISK and JPEG2000; in these results, DDWT-BISK uses the anisotropic decomposition of Fig. 2(b) which performed uniformly better than the dyadic decomposition. As the results show, DDWT-BISK achieves significantly higher PSNR values for each video sequence. Additionally, we have seen similar results for a 2D version of DDWT-BISK developed in [**?**] for the coding of still images. In the 2D case, DDWT-BISK was competitive overall with JPEG2000 but offered better rate-distortion performance on several images with highly directional content. For the purposes of this paper, we now turn our attention to the coding of hyperspectral imagery with the 3D version of the DDWT-

BISK coder.

IV. RESULTS

In our experiments, we code the AVIRIS datasets Cuprite, Jasper Ridge, and Moffett which have 224 spectral bands and which have been cropped spatially to a size of 256×256. The thresholds used in the noise-shaping procedure were optimized for each given bitrate. We apply the DDWT-BISK algorithm with three levels of wavelet decomposition in each dimension. As in the video-coding results of [9], we observed that a 3D DDWT using the anisotropic decomposition of Fig. 2(b) significantly outperformed a 3D DDWT using the dyadic structure of Fig. 2(a); consequently, the results presented here use the anisotropic DDWT structure exclusively.

We compare our results against the state-of-the-art JPEG2000 coder, wherein the JPEG2000 results are obtained through two different approaches to spectral decorrelation. The first, and more common, method consists of applying a DWT spectrally followed by a spatial dyadic DWT ("DWT+JPEG2000," see [1, 15]). The second approach couples the spatial DWT of JPEG2000 with principle component analysis (PCA) for spectral decorrelation as described in [16] ("PCA+JPEG2000"). All DWTs used with JPEG2000 use the popular biorthogonal 9/7 filters.

Table II presents SNR results for the hyperspectral datasets at a fixed rate of 1.0 bits per pixel per band (bpppb). The results indicate that, for hyperspectral data, DDWT-BISK does not exhibit the same performance gains over JPEG2000 that were witnessed on video data in [9]. Fig. 3 shows the full rate-distortion performance of DDWT-BISK and the two JPEG2000 approaches over a range of bitrates for the Moffett dataset. As these results show, DDWT-BISK falls short of both JPEG2000 approaches.

V. CONCLUSIONS

Prior work has demonstrated that the 3D DDWT provides significant performance advantage when deployed in video-coding systems. In particular, the DDWT-BISK coder we proposed in [9] achieved coding results that surpassed those of JPEG2000 which is widely considered to be the state of the art in 3D coding. Given that JPEG2000 is widely used in hyperspectral coding, an evaluation of the potential of the DDWT-BISK coder for hyperspectral imagery is warranted.

For video data, the primary benefit the 3D DDWT provides is an inherent ability to capture motion of salient signal features as they move from one spatial location at a given instant in time to another spatial location at another time. The directionally-oriented subbands of the DDWT capture such signal features moving within the spatiotemporal image volume of the video sequence; thus, video coders using the 3D DDWT do not need to perform explicit motion compensation.

However, hyperspectral data does not appear to benefit from the directionally-oriented subbands in the way video data does. In short, we have concluded that the failure of DDWT-BISK to code hyperspectral data efficiently is largely due to the fact that hyperspectral data does not contain "motion" in the sense

that video data does. That is, salient signal features within the spatiospectral volume of a hyperspectral image do not "move" spatially from band to band. As a consequence, the highly spatiospectrally directional nature of the 3D DDWT remains largely unexploited when applied to hyperspectral imagery. As such, the 3D DDWT is not as an effective model for hyperspectral data, and the primary advantage of DDWT-based coding is lost. For these reasons, it appears that the DDWT is not a suitable transform for use in hyperspectral image compression.

REFERENCES

[1] J. E. Fowler and J. T. Rucker, "3D wavelet-based compression of hyperspectral imagery," in *Hyperspectral Data Exploitation: Theory and Applications*, C.-I. Chang, Ed. Hoboken, NJ: John Wiley & Sons, Inc., 2007, ch. 14, pp. 379–407.

[2] N. G. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Journal of Applied Computational Harmonic Analysis*, vol. 10, pp. 234–253, May 2001.

[3] I. W. Selesnick and K. Y. Li, "Video denoising using 2D and 3D dual-tree complex wavelet transforms," in *Wavelets: Applications in Signal and Image Processing X*, M. A. Unser, A. Aldroubi, and A. F. Laine, Eds. San Diego, CA: Proc. SPIE 5207, August 2003, pp. 607–618.

[4] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 123–151, November 2005.

[5] B. Wang, Y. Wang, I. Selesnick, and A. Vetro, "An investigation of 3D dual-tree wavelet transform for video coding," in *Proceedings of the International Conference on Image Processing*, vol. 2, Singapore, October 2004, pp. 1317–1320.

[6] ——, "Video coding using 3-D dual-tree discrete wavelet transforms," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, Philadelphia, PA, March 2005, pp. 61–64.

[7] F. Shi, B. Wang, I. W. Selesnick, and Y. Wang, "A new structure of 3-D dual-tree discrete wavelet transform and applications to video denoising and coding," in *Visual Communications and Image Processing*, J. G. Apostolopoulos and A. Said, Eds. San Jose, CA: Proc. SPIE 6077, January 2006, p. 60771C.

[8] B. Wang, Y. Wang, I. Selesnick, and A. Vetro, "Video coding using 3-D dual-tree wavelet transform," *EURASIP Journal on Image and Video Processing*, vol. 2007, 2007, article ID 42761, 15 pages.

[9] J. B. Boettcher and J. E. Fowler, "Video coding using a complex wavelet transform and set partitioning," *IEEE Signal Processing Letters*, vol. 14, no. 9, pp. 633–636, September 2007.

[10] T. H. Reeves and N. G. Kingsbury, "Overcomplete image coding using iterative projection-based noise shaping," in *Proceedings of the International Conference on Image Processing*, vol. 3, Rochester, NY, September 2002, pp. 597–600.

[11] J. E. Fowler, "Shape-adaptive coding using binary set splitting with k-d trees," in *Proceedings of the International Conference on Image Processing*, vol. 2, Singapore, October 2004, pp. 1301–1304.

[12] J. T. Rucker and J. E. Fowler, "Coding of ocean-temperature volumes using binary set splitting with k-d trees," in *Proceedings of the International Geoscience and Remote Sensing Symposium*, vol. 1, Anchorage, AK, September 2004, pp. 289–292.

[13] ——, "Shape-adaptive embedded coding of ocean-temperature imagery," in *Proceedings of the 40th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, October 2006, pp. 1887–1891.

[14] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Communications of the ACM*, vol. 18, no. 9, pp. 509–517, September 1975.

[15] J. T. Rucker, J. E. Fowler, and N. H. Younan, "JPEG2000 coding strategies for hyperspectral data," in *Proceedings of the International Geoscience and Remote Sensing Symposium*, vol. 1, Seoul, South Korea, July 2005, pp. 128–131.

[16] Q. Du and J. E. Fowler, "Hyperspectral image compression using JPEG2000 and principal component analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 2, pp. 201–205, April 2007.

TABLE I

VIDEO CODING RESULTS—DISTORTION AVERAGED OVER ALL FRAMES OF
THE SEQUENCE FOR A RATE OF 0.5 BITS/PIXEL.

| | PSNR (dB) | |
	DDWT-BISK	JPEG2000
Stefan	31.5	30.7
Mobile	30.5	27.7
Foreman	38.8	37.9
Coastguard	33.9	33.1
Table Tennis	35.6	35.2

Sequences are CIF (352 × 288) with 80 frames at 30 Hz.

TABLE II

SNR PERFORMANCE AT A RATE OF 1.0 BPPPB.

| | SNR (dB) | | |
	DDWT-BISK	DWT+JPEG2000	PCA+JPEG2000
Moffett	42.4	45.9	50.7
Jasper Ridge	42.2	45.4	49.8
Cuprite	47.9	51.3	53.8

(a)

(b)

Fig. 2. Three-level wavelet decomposition for (a) dyadic and (b) anisotropic decompositions. The decomposition structure is applied in each transform combination shown in Fig. 1.

Combination 1   Combination 2
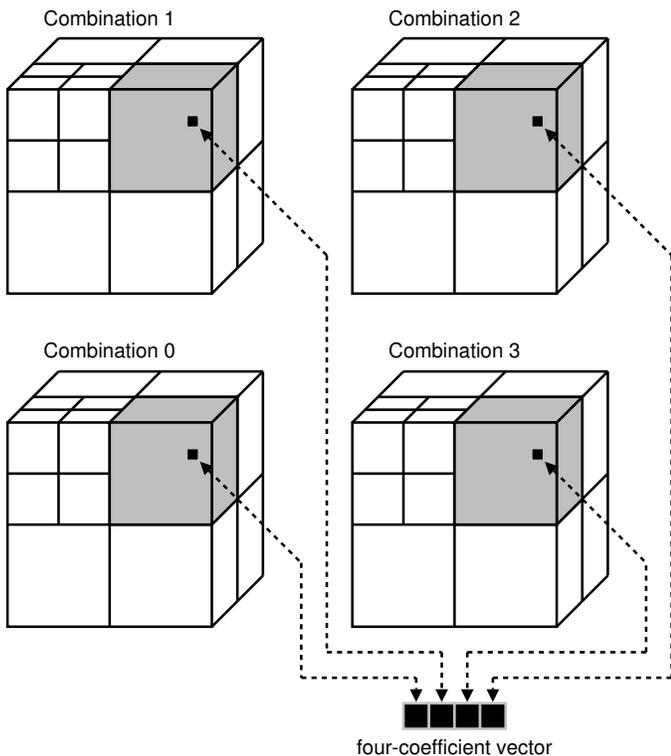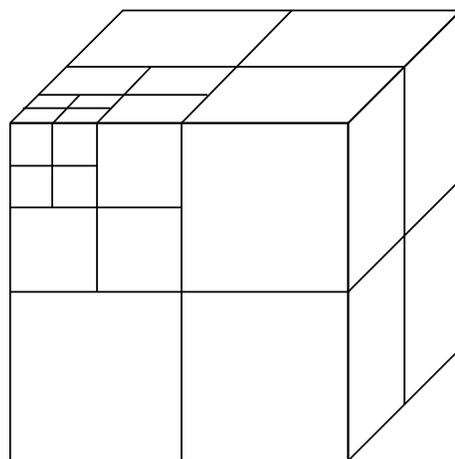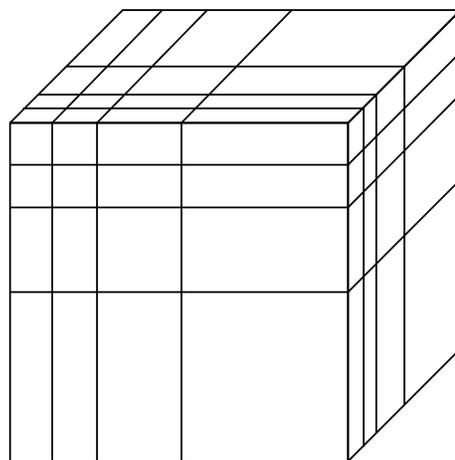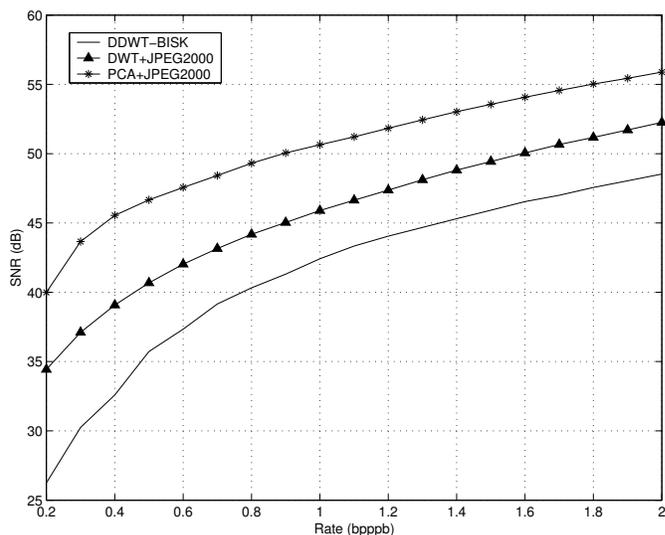
Combination 0   Combination 3

four-coefficient vector

Fig. 1. Co-located coefficients in each of the four transform combinations form a four-coefficient vector. The contents of the shaded regions form one subband set in DDWT-BISK. Illustrated for a dyadic decomposition structure (see Fig. 2).

Fig. 3. Rate-distortion performance for "Moffett"

In *Proceedings of the International Geoscience and Remote Sensing Symposium*, Barcelona, Spain, July 2007, pp. 1033-1036.