AN INVESTIGATION OF 3D DUAL-TREE WAVELET TRANSFORM FOR VIDEO CODING

Beibei Wang[1], Yao Wang[1], Ivan Selesnick[1] and Anthony Vetro[2]

[1] Polytechnic University, Electrical and Computer Engineering Dept, Brooklyn, NY
[2] Mitsubishi Electric Research Laboratories, Cambridge, MA
Email: (bb_w, yao)@vision.poly.edu, selesi@duke.poly.edu, avetro@merl.com

ABSTRACT

This paper examines the properties of a recently introduced 3-D dual-tree discrete wavelet transform (DDWT) for video coding. The 3-D DDWT is an attractive video representation because it isolates motion along different directions in separate subbands. However, it is an overcomplete transform with 8:1 redundancy. We examine the effectiveness of the iterative projection-based noise shaping scheme proposed by Kingsbury [3] on reducing the number of coefficients. We also investigate the correlation between subbands at the same spatial/temporal location, both in the significance map and in actual coefficient values.

1. INTRODUCTION

The standard separable discrete wavelet transform (DWT) provides a multi-resolution representation of a signal. The DWT can be used for a variety of applications such as denoising, enhancement, and compression. Several recently proposed DWT-based video coders have achieved coding efficiency similar to or slightly better than block-based hybrid video coders [1]. In addition, such coders provide a scalable representation of the video in spatial resolution, temporal resolution and quality. However, the multidimensional DWT mixes orientations in its subbands, which can lead to checkerboard artifacts at the low bit rate range.

An important recent development in wavelet-related research is the design and implementation of 2-D multiscale transforms that represent edges more efficiently than does the DWT. Kingsbury's complex dual-tree wavelet transform (DT-CWT) is an outstanding example [2]. The DT-CWT is an overcomplete transform with limited redundancy (2^m :1 for m-dimensional signals). This transform has good directional selectivity and its subband responses are approximately shift-invariant. The 2-D DT-CWT has given superior results for image processing applications compared to the DWT [2,3]. Recently, Selesnick and Sendur introduced a 3-D version of the dual-tree wavelet transform and showed that it has superior motion selectivity [4]. In this paper, we explore the suitability of the 3-D DDWT for video coding.

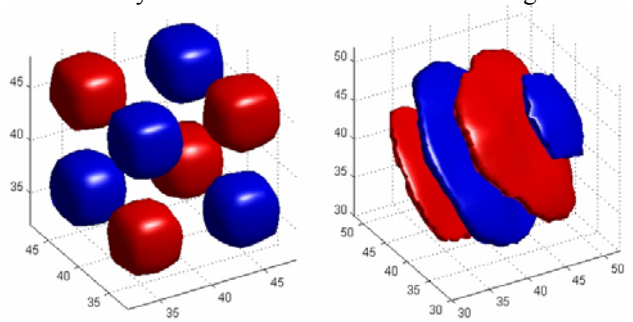


Fig.1 Isosurfaces of a typical 3-D DWT (left) and a typical 3-D DDWT (right). For the 3-D DDWT, each subband corresponds to motion in a specific direction.

We start by briefly introducing the 3-D DDWT and its properties for video representation. Section 3 describes how to select significant coefficients for video coding. Section 4 investigates the correlation between coefficients for different wavelet bases for both the significance map and the actual coefficients. The final section summarizes our work and discusses options for video coding using 3-D DDWT.

2. 3-D DUAL-TREE WAVELET TRANSFORM

The design and the motion-selectivity of 3-D dual-tree complex wavelet transform are described in [4]. A Daubechies-like algorithm for the construction of Hilbert pairs of short orthonormal (and biorthogonal) wavelet bases yields pairs of bases, which can be used to efficiently implement the motion-selective wavelet transform [5]. The dual-tree wavelet transform is implemented by first applying separable transforms and then combining subband signals with simple linear operations. So even though it is non-separable (and therefore free of some of the limitations of separable transforms), it inherits the computational efficiency of separable transforms.

Figure 1 illustrates the difference between the standard 3-D DWT and the 3-D DDWT. The figure

0-7803-8554-3/04/$20.00 ©2004 IEEE.

depicts the wavelets (i.e. the basis functions) associated with the 3-D DWT and the 3-D DDWT respectively. As illustrated, the 3-D DWT mixes different orientations in one wavelet basis, but the 3-D DDWT is free of this effect. The 3-D DDWT has many more subbands (a subband refers to the coefficients associated with one wavelet basis) than the 3-D DWT (28 high subbands instead of 7, 4 low subbands instead of 1). Only a subset of the high subbands is drawn in Fig.1 for DDWT. The 28 high subbands isolate 2-D edges with different orientations that are moving in different directions. Because of this motion selectivity, the 3-D DDWT is likely to be substantially more effective for the representation of video than the 3-D DWT.

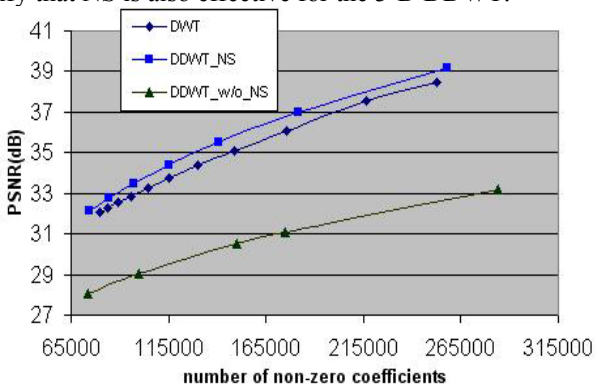
A core element common to all state-of-the-art video coders is motion-compensated temporal prediction, which is the main contributor to the complexity as well error-sensitivity of a video encoder. Because the subband coefficients associated with the 3-D DDWT directly capture moving edges in different directions, it may not be necessary to perform motion estimation explicitly. This is our primary motivation for exploring the use of 3-D DDWT for video coding.

For all the results in this paper we use the Daubechies (9, 7)-tap filters for the DWT implementation, and use the DDWT described in [4]. Each transform uses three levels of wavelet decomposition. All results are tested on two sequences "Foreman (QCIF)" and "Mobile-Calendar (CIF)".

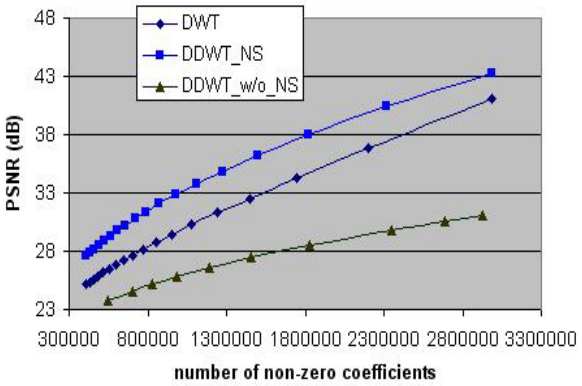
3. ITERATIVE SELECTION OF COEFFICIENTS

The major challenge to apply the 3-D complex DDWT for video coding is it is an overcomplete transform with 8:1 redundancy. In our current study, we chose to retain only the real parts of the wavelet coefficients, which can still lead to perfect reconstruction, while retaining the motion selectivity. This reduces the redundancy to 4:1 [4].

An overcomplete transform is not necessarily ineffective for coding, because a redundant set provides flexibility in choosing which basis functions to use in representing a signal. Even though the transform itself is redundant, the critical coefficients that must be retained to represent a video signal accurately can be smaller than that obtained with standard non-redundant transform. In addition, motion-selective oriented basis functions are likely to lead to better visual quality especially at lower bit rates. The matching pursuit algorithm is a well-known technique for video coding with the overcomplete representations [6]. With matching pursuit, larger coefficients are chosen iteratively to represent a signal, but once the largest coefficient is chosen from the remaining ones, its associated basis is deleted from the set of bases to be considered in the following iterations. Kingsbury proposed an iterative projection-based noise shaping (NS) scheme [3], which modifies previously chosen large coefficients to compensate for the loss of small coefficients. It was shown that noise shaping applied to 2-D DT-CWT can yield a more compact set of coefficients than from the 2-D DWT. In this section, we verify that NS is also effective for the 3-D DDWT.



(a) Foreman (QCIF)



(b) Mobile-Calendar (CIF)

Fig. 2 PSNR (dB) vs. number of non-zero coefficients for the DDWT using noise shaping (DDWT_NS, upper curve), the DWT (middle curve), and the DDWT without noise shaping (lower curve).

Figure 2 compares the reconstruction quality (in terms of PSNR) using the same number of retained coefficients with DWT, DDWT with noise shaping (DDWT_NS) and DDWT without noise shaping (DDWT_w/o_NS). For a given number of coefficients to retain, *N*, the results for DWT and DDWT_w/o_NS are obtained by simply choosing the *N* largest ones from the original coefficients. With DDWT_NS, the coefficients are obtained by running the iterative projection algorithm with a preset initial threshold, and gradually reducing it until the number of remaining coefficients reaches *N*. Figure 2 shows that although the raw number of coefficients with 3-D DDWT is 4 times more than DWT, this number can be reduced substantially by noise shaping. In fact, with the same number of retained coefficients,

DDWT_NS yields higher PSNR than DWT. For "foreman", 3-D DDWT_NS has a slightly higher PSNR than the DWT (0.3~0.7 dB), and is 4~6 dB better than DDWT_w/o_NS. For "Mobile-Calendar", the DDWT_NS is 1.5~3.4 dB better than the DWT. Subjectively, DDWT_NS preserves edge and motion information better than DWT, while DWT exhibits blurs in some regions and when there are a lot of motions.

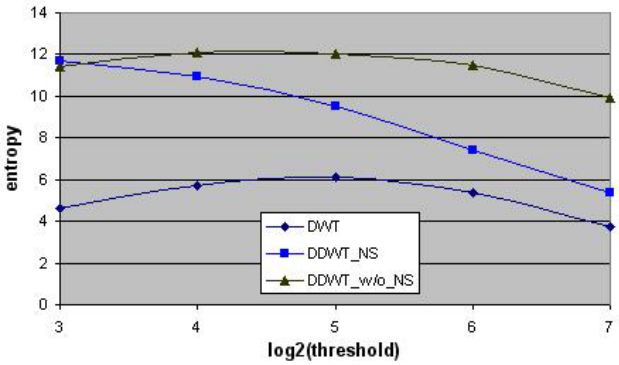
4. CORRELATION BETWEEN SUBBANDS

4.1. Correlation in Significance Maps

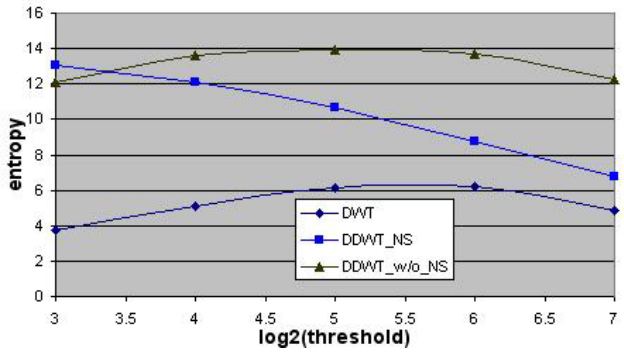
Figure 2 shows that with DDWT_NS, we can use fewer coefficients to reach a desired reconstruction quality than DWT. However, this does not necessarily mean that DDWT_NS will require fewer bits. This is because DDWT coefficients are spread over more subbands than DWT, and specifying the location of a DDWT coefficient may require more bits than specifying the location of a DWT coefficient. The success of a wavelet-based coder critically depends on whether the location information can be coded efficiently.

We hypothesize that although 3-D DDWT has many more subbands, only a few subbands have significant energy for an object feature. Specifically, an oriented edge moving in a particular direction is likely to generate significant coefficients only in the subbands with the same or adjacent spatial orientation and motion pattern. On the other hand, with the 3-D DWT, a moving object in an arbitrary direction that are not characterized by any specific wavelet basis will likely contribute to many small coefficients in all subbands. To validate this hypothesis, we compute the entropy of the vector consisting of the significance bits at the same spatial/temporal location across 28 high subbands. The significance bit in a particular subband is either 0 or 1 depending on whether the corresponding coefficient is below or above a chosen threshold. This entropy will be close to 28 if there is not much correlation between the 28 subbands. On the other hand, if the pattern that describes which bases are simultaneously significant is highly predictable, the entropy should be much lower than 28. Similarly, we calculate the entropy of the significance bits across the 7 high subbands of DWT, and compare it to the maximum value of 7.

Figure 3 compares the vector entropies of significance maps associated with DDWT_w/o_NS, DDWT_NS and DWT, for varying thresholds from 128 to 8. The results shown here are for the top scale only; other scales follow the same trend. We see that, with DDWT, even without noise shaping, the vector entropy is much lower than 28. Moreover, noise shaping helps reduce the entropy further. In contrast, with DWT, the vector entropy is close to 7 at some threshold values. Also noteworthy is that, at high thresholds, the entropy for DDWT_NS is quite close to that for DWT.



(a) Foreman (QCIF)



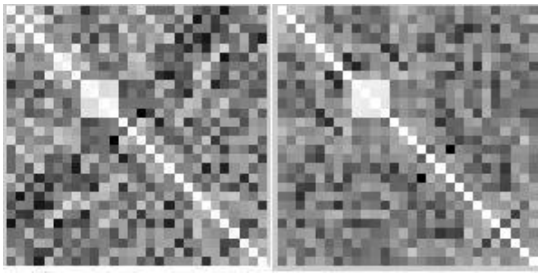
(b) Mobile-Calendar (CIF)

Fig. 3 The vector entropy of significance maps using the 3-D DWT (the lowest curve), the DDWT_NS (the middle curve) and the DDWT_w/o_NS (the upper curve), for the top scale.

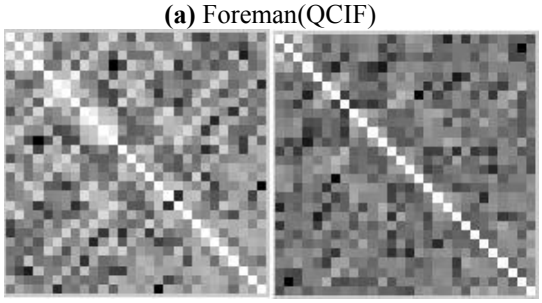
4.2. Correlation in coefficient values

In addition to the correlation among the significance maps of all subbands, we also investigate the correlation between the actual coefficient values. Strong correlation would suggest vector quantization or predictive quantization among the subbands. Towards this goal, we compute the correlation matrix and variance of the 28 high subbands. Figure 4 illustrates the correlation matrices for the top scale, for both the DDWT_w/o_NS and DDWT_NS. We note that the correlation patterns in other scales are similar to this top scale. From these correlation matrices, we find that only a few subbands have stronger correlation, and most other subbands are almost independent. After noise shaping, the correlation between subbands is reduced significantly. A greater number of subbands are almost independent from each other. It is interesting to note that, for the "Foreman" sequence (which has predominantly vertical edges and horizontal motion), bands 9-12 are highly correlated before and after noise shaping. The wavelets associated with these four bands have nearly vertical orientations but all moving in the horizontal direction.

Figure 5 illustrates the energy distribution among the 28 subbands for the top scale with and without noise shaping. The energy distribution pattern depends on the edge and motion patterns in the underlying sequence. For example, the energy is more evenly distributed between different subbands with "Mobile-Calendar". Furthermore, noise shaping helps to concentrate the energy into fewer subbands.



(a) Foreman(QCIF)



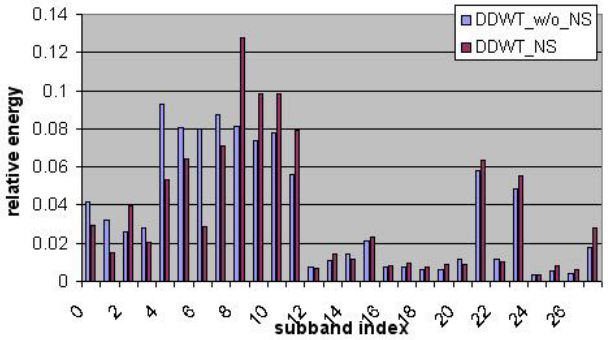
(b) Mobile-Calendar (CIF)

Fig. 4 The correlation matrices of the 28 subbands of 3-D DDWT_w/o_NS (left) and DDWT_NS (right). The grayscale is logrithmically related to the absolute value of the correlation. The brighter colors represent higher correlation.

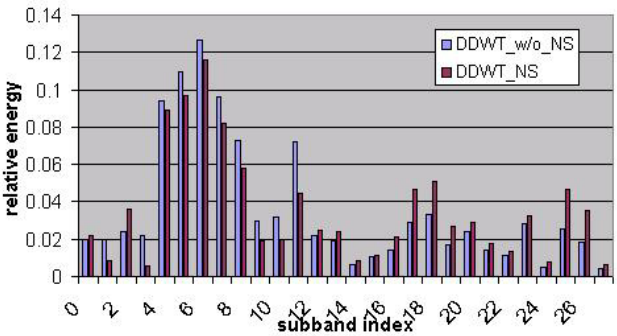
5. CONCLUSION

We demonstrated that 3-D DDWT has attractive properties for video representation. Although the 3-D DDWT is an overcomplete transform, the raw number of coefficients can be reduced substantially by applying noise shaping. The fact that noise shaping can reduce the number of coefficients to below that required by DWT (for the same video quality) is very encouraging. The vector entropy study validates our hypothesis that only a few bases have significant energy for an object feature. The relatively low vector entropy suggests that the whereabouts of significant coefficients may be coded efficiently by coding the significance bits across subbands jointly (through either vector Huffman coding or arithmetic coding). The fact that coefficient values do not have strong correlation among the subbands, on the other hand, indicates that the benefit from vector coding the magnitude bits across the subbands may be limited.

In terms of future work, the correlation among adjacent spatial and temporal coefficients still needs to be explored. Context-based arithmetic coding or quadtree coding may be used to explore such correlation if exists. Finally, with noise shaping, the optimal set of coefficients to be retained changes with the target bit rate. To design a rate-distortion optimized scalable video coder, we would like to have a scalable set of coefficients so that each additional coefficient offers a maximum reduction in distortion without modifying the previous coefficients. How to deduce such coefficient sets is a challenging open research problem.



(a) Foreman (QCIF)



(b) Mobile-Calendar (CIF)

Fig. 5 The relative energy of 3-D DDWT 28 subbands with (the right column in each subband) and without noise shaping (the left column).

6. REFERENCES

[1] S-T Hsiang and J. W. Woods, "Embedded video coding using invertible motion compensated 3-D subband/wavelet filter bank", *Signal Processing: Image Communications,* vol. 16, pp. 705-724, May 2001

[2] N.G. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals", *Applied Computational Harmonic Anal,* vol. 10, no. 3, pp. 234-253, May 2001

[3] T H Reeves and N G Kingsbury, "Overcomplete image coding using iterative projection-based noise shaping", *ICIP 02,* Rochester, NY, Sept 2002.

[4] I. W. Selesnick, and Ke Yong Li, "Video denoising using 2D and 3D dual-tree complex wavelet transforms", *Wavelet Appl Signal Image Proc. X (Proc. SPIE 5207)*, Aug 2003.

[5] I. W. Selesnick, "The design of approximate Hilbert transform pairs of wavelet bases", *IEEE Trans. on Signal Processing,* 50(5): 1144-1152, May 2002

[6] S. G. Mallat, Z. Zhang, "Matching pursuits with time frequency dictionaries", *IEEE Trans. Signal Processing,* vol. 41, pp. 3397-3415, Dec. 1993.