EMBEDDED CODING OF THE MOTION-COMPENSATED 3-D WAVELET COEFFICIENTS BY CONCATENATING SPATIAL AND TEMPORAL ORIENTATION TREES

Liang Zhang

Communications Research Centre Canada
3701 Carling Avenue, Ottawa, Ontario, K2H 8S2, Canada
email: liang.zhang@crc.ca

ABSTRACT

This paper focuses on the issue of coding of the spatiotemporal 3-D wavelet coefficients. Experimental examination found that there are temporal dependencies among the wavelet coefficients at different temporal decomposition levels. Temporal orientation trees are introduced to exploit these dependencies to improve the coding efficiency. Statistical properties of the wavelet coefficients are measured using test video sequences. 3-D orientation trees that are built by spatial orientation trees followed by temporal orientation trees are exploited to magnitude-order the spatiotemporal 3-D wavelet coefficients. The experimental results confirmed that the coder with temporal orientation trees outperformed the coder without them. The coding gain could be up to 0.5 dB.

1. INTRODUCTION

Interframe wavelet video coding system removes temporal redundancy by motion-compensated temporal filtering (MCTF) technique. In the so-called *t+2-D* wavelet transform scheme, motion estimation is carried out on the input video signal. With the motion fields, original frames are temporally filtered along motion trajectories, resulting in lowpass and highpass temporal subbands, for each pair of frames in a group of pictures (GOP). This MCTF process described above is the first level of the multi-level temporal wavelet decomposition. After that, this temporal decomposition process is repeated on the lowpass temporal subbands, building a temporal wavelet tree. Subsequently, the resulting temporal subbands go through a spatial wavelet analysis, resulting in the motion-compensated 3-D (spatiotemporal) wavelet coefficients. Finally, all motion fields and the motion-compensated spatiotemporal wavelet coefficients are encoded and transmitted.

The MCTF technique has been investigated recently by many researchers [1]~[5]. The actual MCTF technique is described in [2], in which a so-called MC-EZBC coder was reconstructed with bi-directional MCTF of up to 1/8-pixel accuracy. However, less research on methodologies to encode the spatiotemporal wavelet coefficients has been conducted until now [1].

In a 2-D wavelet-coding scheme such as EZW [6] and SPIHT [7], a 2-D spatial orientation tree, which exploits the spatial cross-subband similarity of the image wavelet transform across different spatial subbands, has proved very successful. Motivated by its success in image compression, researchers have extended it directly from 2-D to 3-D for video coding. The 3-D SPIHT [8] employs the SPIHT coding algorithm in 3-D spatiotemporal orientation trees in video coding, which are analogous to the 2-D orientation trees in image coding, and requires that the transform stages along all the dimensions are equal. However, the properties of the video signal and the wavelet coefficients are not equal along all three dimensions. To address this issue, the authors in [9] proposed an asymmetric 3-D orientation tree structure that is designed by attaching all subbands together for constructing a longer subband tree. The proposed coding algorithm in [9] outperforms 3-D SPIHT and has no limitation on the levels of transform along each direction without motion compensation. The state-of-the-art MC-EZBC coder encodes the 3-D spatiotemporal wavelet coefficients using the embedded image-coding scheme EZBC [10], in which no temporal cross-subband similarity of the spatiotemporal wavelet coefficients are applied.

In this paper, we focus on the issue of efficient coding of spatiotemporal wavelet coefficients by additionally exploiting their temporal statistical properties. Correlations among different spatiotemporal wavelet subbands are measured using testing video sequences. Based on the measured statistical properties, temporal orientation trees are proposed to exploit the temporal cross-subband similarity of the spatiotemporal wavelet coefficients between different temporal levels. Combining with spatial orientation trees, the spatiotemporal wavelet coefficients could be encoded and transmitted even more efficiently.

This paper is organized as follows. After the introduction, temporal orientation trees are introduced in Section 2 based on the measurement of temporal statistical properties of spatiotemporal wavelet coefficients. Section 3 describes the coding algorithm using the concatenation of spatial and temporal orientation trees. Experimental results with video sequences are presented in Section 4 for demonstrating the performance of the proposed algorithm. Section 5 concludes this paper.

1-4244-0367-7/06/$20.00

2. TEMPORAL ORIENTATION TREES

It is well known that there is a spatial cross-subband similarity among spatial subbands of 2-D spatial wavelet decomposition. Since the temporal decomposition is similar to the spatial wavelet decomposition, it is straightforward to consider that a temporal cross-subband similarity might also exist among the spatiotemporal wavelet coefficients. To confirm this extension, statistical properties of spatiotemporal wavelet coefficients among different temporal levels have to be measured. To this end, an interframe video coding system is constructed using unidirectional MCTF technique, in which hierarchical variable size block matching (HVSBM) is used for motion estimation of quarter subpixel accuracy [4] and a Haar transform using lifting implementation is used for temporal decomposition along motion trajectories. For the 2-D spatial wavelet transform, the 9/7 wavelet filters using lifting implementation are used. The measurements are performed on two video sequences, *Foreman* in CIF resolution (352×288, 30fps) with 144 frames and *Mobile* in SIF resolution (352×240, 30fps) with 240 frames, with a GOP length of 8 frames for T=3 temporal decomposition levels (Fig. 1).

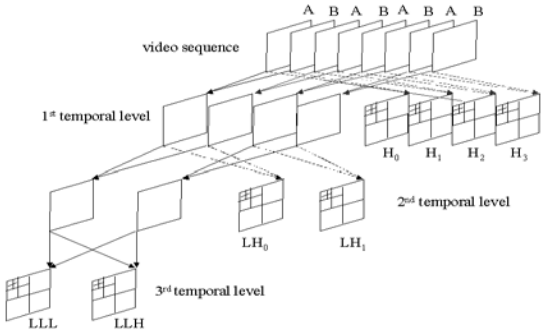

Fig. 1. Spatiotemporal wavelet decomposition using T=3 temporal levels.

Table I. Average maximum values of spatiotemporal wavelet coefficients in each temporal subband in a group of pictures measured over all GOPs of test sequences, *Foreman* and *Mobile*.

	LLL	LLH	LH_0	LH_1	H_0	H_1	H_2	H_3
Foreman	4239	286	182	188	116	112	124	118
Mobile	2566	618	370	361	196	200	195	196

To confirm the temporal cross-subband similarity existing among the spatiotemporal wavelet coefficients, maximum values of spatiotemporal wavelet coefficients for each temporal subband within a group of pictures are measured. Table I shows the results that are averaged over all GOPs of test sequences *Foreman* and *Mobile*. We can see that the temporal lowest lowpass subband, denoted by LLL, has the largest maximum value, while the temporal highest highpass subbands, denoted by H_0, H_1, H_2, and H_3, have the smallest maximum value (Fig. 1). These results indicate that there is a temporal cross-subband

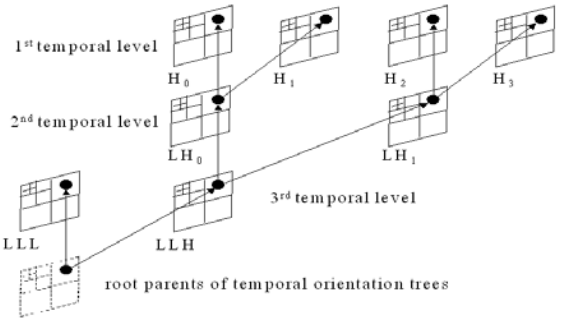

Fig. 2. Examples of parent-child dependencies in temporal orientation trees for T=3 temporal levels.

Table II. Normalized cross-correlation coefficients between spatiotemporal wavelet coefficients along temporal orientation trees for test sequence *Foreman*, measured for each spatial subband.

	temporal subband pairs along temporal orientation trees						
spatial subbands	LLL,LLH	LLH,LH_0	LLH,LH_1	LH_0, H_0	LH_0, H_1	LH_1, H_2	LH_1, H_3
LL_4	0.07343	0.31940	0.42094	0.54919	0.52816	0.51129	0.55042
HL_4	0.04293	0.19941	0.26501	0.37454	0.38337	0.31984	0.36793
LH_4	0.06934	0.21769	0.25825	0.35366	0.40338	0.31657	0.36464
HH_4	0.04688	0.16863	0.20515	0.28144	0.31200	0.25736	0.30325
HL_3	0.02429	0.14616	0.19586	0.26897	0.29268	0.29999	0.32356
LH_3	0.03909	0.13846	0.17134	0.31450	0.26871	0.23199	0.30359
HH_3	0.03107	0.09216	0.10739	0.22628	0.20262	0.19764	0.24429
HL_2	0.02561	0.08753	0.12743	0.20202	0.21154	0.17951	0.25585
LH_2	0.02795	0.09398	0.12896	0.21879	0.18101	0.17964	0.22853
HH_2	0.03236	0.06073	0.07317	0.12893	0.12007	0.10295	0.16018
HL_1	0.01943	0.05241	0.08972	0.11560	0.13612	0.10587	0.16668
LH_1	0.15037	0.07129	0.14315	0.10315	0.17504	0.11860	0.19512
HH_1	0.03285	0.04751	0.05679	0.07912	0.09580	0.06637	0.09202

Table III. Normalized cross-correlation coefficients between spatiotemporal wavelet coefficients along temporal orientation trees for test sequence *Mobile*, measured for each spatial subband.

	temporal subband pairs along temporal orientation trees						
spatial subbands	LLL, LLH	LLH, LH_0	LLH, LH_1	LH_0, H_0	LH_0, H_1	LH_1, H_2	LH_1, H_3
LL_3	0.08890	0.60470	0.61692	0.63958	0.65449	0.6511	0.66069
HL_3	0.06548	0.31092	0.37576	0.49947	0.52649	0.47437	0.51629
LH_3	0.06240	0.40002	0.38279	0.41253	0.39512	0.39714	0.39911
HH_3	0.02638	0.18508	0.23576	0.30327	0.29176	0.28510	0.31088
HL_2	0.02255	0.06920	0.11982	0.23899	0.28895	0.21914	0.28889
LH_2	0.04099	0.23234	0.25638	0.29552	0.30396	0.28509	0.31851
HH_2	0.01979	0.05211	0.04745	0.17005	0.20422	0.16233	0.22131
HL_1	0.04391	0.03726	0.07587	0.08366	0.16221	0.07360	0.17606
LH_1	0.01847	0.05411	0.09031	0.17493	0.34848	0.18115	0.35414
HH_1	0.02928	0.03187	0.08415	0.08768	0.13394	0.08175	0.14255

similarity between spatiotemporal wavelet coefficients, which is analogous to the spatial cross-subband similarity of the 2-D spatial wavelet decomposition. The temporal cross-subband similarity states that with respect to a given threshold if a spatiotemporal wavelet coefficient at a higher temporal level is insignificant, then all spatiotemporal wavelet coefficients in the

same spatial location at the lower temporal level are likely to be insignificant.

A tree structure, called *temporal orientation tree*, is used to represent this temporal cross-subband similarity between different temporal levels. Fig. 2 shows how the temporal orientation tree is defined with recursive two-subband temporal splitting. Each node of the tree corresponds to a pixel and is identified by the pixel coordinate and the series of temporal subbands in a temporal level. Its direct descendants (offspring) correspond to the pixels of the same spatial location in the temporal highpass subbands of a lower temporal level. The tree is defined in such a way that each node has either no offspring (the leaves) or two offspring. In Fig. 2, the arrows are oriented from the parent node to its two offspring. The tree roots have also two offspring. However, only one of them further has offspring that correspond to the temporal highpass subband, while the other node has no offspring that corresponds to the temporal lowpass subband.

The spatiotemporal wavelet coefficients can be magnitude-ordered via either temporal or spatial orientation trees. They can also be better magnitude-ordered using the concatenation of temporal and spatial orientation trees. In the later case, the concatenating order should be determined. Do spatial orientation trees follow temporal orientation trees or *vice versa*?

We use the concatenating order of spatial orientation trees followed by temporal orientation trees to magnitude-order the spatiotemporal wavelet coefficients. The reason for this concatenating order choice is that good dependency of spatiotemporal wavelet coefficients between different temporal levels along temporal orientation trees only exists for the lower spatial subbands.

The dependencies of the spatiotemporal wavelet coefficients between different temporal levels along temporal orientation trees are measured by the values of the normalized cross-correlation coefficients (NCC). Tables II and III show the NCC values that are measured on test sequences *Foreman* and *Mobile*. The rows of Tables II and III are spatial subbands at different spatial decomposition levels, which are denoted by LL_n, HL_n, LH_n, and HH_n with the spatial decomposition level n, while the columns of Tables II and III stand for the pair of temporal subbands along temporal orientation trees. Not surprisingly, the NCC values between the temporal LLL (lowpass) and LLH (highpass) subbands are very small for every spatial subband (see the second column in Tables II and III) because these two temporal subbands, LLL and LLH, represents total different signals. For the remaining temporal subband pairs, it is also notable from Tables II and III that the NCC values decrease from the lowest spatial subband (LL subband) to the highest spatial subbands (LH_1, HL_1, and HH_1). For example, the NCC value, measured for the temporal pair of LH_0 and H_0, decreases from a value of 0.55 for the spatial LL_4 subband to the value of 0.1 for the spatial subbands, LH_1, HL_1 and HH_1, (see the fifth column in Table II). This indicates that the temporal cross-subband similarity between these two temporal subbands is most likely to be violated for the spatial highpass subbands, e.g., LH_1, HL_1 and HH_1, since they all represent residual signals. Only the spatial LL (LL_4 in Table II and LL_3 in Table III) subband is likely to follow the temporal cross-subband similarity, because they have relative good correlations (see the second rows in Tables II and III). Based on this finding, we first magnitude-order the spatiotemporal wavelet coefficients using spatial orientation trees for each temporal subbands and then magnitude-order the root nodes of these spatial orientation trees along the temporal levels using temporal orientation trees to build 3-D orientation trees.

The principle to build 3-D orientation trees developed in this paper is different than that developed in [9]. In [9], a coefficient in the 3-D orientation tree has up to five child coefficients, while in the proposed 3-D trees the parent coefficients have two child coefficients in temporal orientation trees and four child coefficients in spatial orientation trees.

3. NEW CODING ALGORITHM

Using the new 3-D orientation tree structure introduced in Section 2, an algorithm for encoding the spatiotemporal wavelet coefficients is developed using the SPIHT technique. All significance information is stored in 3×L+1 ordered lists, where L is the GOP length. One is called list of temporal insignificant sets (TLIS) and the remaining 3×L lists are used for the spatial orientation trees. For one temporal subband, t, within the GOP, there are one list of spatial insignificant sets ($SLIS_t$), one list of spatial insignificant pixels ($SLIP_t$), and one list of spatial significant pixels ($SLSP_t$). In the temporal tree list TLIS, each entry is identified by (x,y,t), where (x,y) is a spatial coordinate and t is the series of the temporal subbands within the GOP. In all spatial tree lists, each entry is identified by a coordinate (x,y).

Below we present the new encoding algorithm in its entirety using the SPIHT technique.

Algorithm:

1) *Initialization*: output $n=\lfloor \log_2(\max_{(x,y,t)}\{|w_{x,y,t}|\})\rfloor$; set all spatial lists, $SLIS_t$, $SLIP_t$, and $SLSP_t$, as empty lists, and the coordinates (x,y) in the low-low spatial subband of the highest spatial wavelet decomposition level to the TLIS as root entries with t=0.
2) *Temporal Sorting Pass*:
   for each entry (x,y,t) in the TLIS do:
   - output $S_n(x,y,t)$;
   - if $S_n(x,y,t)=1$ then

   2.1) if the entry is a root entry (t=0) then
   - set (x,y) to the spatial list $SLIS_0$, add (x,y,1) to the end of TLIS, and remove entry (x,y,0) from the TLIS;

   2.2) if the entry is not a root entry (t>0) then
   - set (x,y) to the spatial list $SLIS_t$, add (x,y,2t) and (x,y,2t+1) to the end of TLIS, and remove entry (x,y,t) from the TLIS;
3) *SPIHT algorithm*: encode all spatial lists $SLIP_t$, $SLIS_t$, $SLSP_t$;
4) *Quantization-step Update*: decrement n by 1 and go to Step 2.

Please note that the entries added to the end of the TLIS in Step

2 are evaluated before that same temporal sorting pass ends. Compared to conventional 3-D wavelet video algorithms [8][9], the proposed coding algorithm introduces an additional temporal sorting pass. With this proposed algorithm, the rate can still be precisely controlled because the transmitted information is formed by one single bitstream. To increase the coding efficiency, groups of 2×2 pixels are kept together in the lists, and their significance values are coded as a single symbol by the arithmetic coding algorithm.

4. EXPERIMENTAL RESULTS

The proposed new coding algorithm is evaluated using the test sequences *Foreman* and *Mobile*, each having 128 frames, in a GOP length of 16 and 32 frames against the reference coding algorithm without using temporal orientation trees. Both algorithms use the same MCTF technique except for the magnitude ordering of the spatiotemporal wavelet coefficients. Therefore, they have the exact same bits for encoding the side information including all motion vectors.

The performance comparison results, shown in Tables IV and V, confirmed that the temporal orientation trees improved the efficiency of encoding the spatiotemporal wavelet coefficients. The gain could reach up to 0.5 dB for *Foreman* and 0.2 dB for *Mobile*, especially at the low bit rate. The results also show that the coder with the GOP length of 32 frames has better coding gain than the coder with the GOP length of 16 frames. Therefore, the proposed coding algorithm prefers long GOP so that the magnitude order using the temporal orientation trees becomes more effective than the coder with short GOP.

Table IV. Performance comparison of the new coder with the reference coder for video sequence *Foreman*.

kbps	PSNR (dB)					
	GOP=16			GOP=32		
	Ref	New	Diff	Ref	New	Diff
200	31.1812	31.6886	+0.5074	31.2701	31.7991	+0.5290
400	34.3198	34.5121	+0.1923	34.2855	34.5083	+0.2228
600	35.7587	35.9111	+0.1524	35.7284	35.8933	+0.1649
800	37.0892	37.1991	+0.1099	37.0440	37.1870	+0.1430
1000	37.8719	37.9494	+0.0775	37.7964	37.8952	+0.0988

Table V. Performance comparison of the new coder with the reference coder for video sequence *Mobile*.

kbps	PSNR (dB)					
	GOP=16			GOP=32		
	Ref	New	Diff	Ref	New	Diff
500	25.8749	26.0485	+0.1736	26.2599	26.4864	+0.2265
1000	28.7528	28.8679	+0.1151	28.9578	29.1259	+0.1681
1500	30.4992	30.5686	+0.0694	30.6109	30.7028	+0.0919
2000	32.2178	32.3226	+0.1048	32.3773	32.5239	+0.1466
2500	33.8235	33.8640	+0.0405	33.8613	33.9215	+0.0602

5. CONCLUSIONS

This paper investigated the issue of encoding the spatiotemporal 3-D wavelet coefficients using temporal orientation trees for interframe wavelet video coding. The experimental examination found that there is good temporal dependency among the spatiotemporal wavelet coefficients between different temporal levels, especially for the low-low spatial subband. Based on this finding, the spatiotemporal wavelet coefficients are magnitude-ordered by 3-D orientation trees, which are built by concatenating spatial orientation trees followed by temporal orientation trees. The principle to build 3-D orientation trees developed in this paper is different to those developed in [8] and [9]. The performance comparison results confirmed that the coding algorithm with temporal orientation trees outperforms the coder without them. In the future, the proposed coding algorithm will be evaluated against the coder presented in [9].

6. ACKNOWLEDGEMENTS

The authors would like to thank Dr. Demin Wang, Dr. Rosario Feghali, Dr. Filippo Speranza, and Mr. André Vincent for their valuable comments on an earlier version of this manuscript.

7. REFERENCES

[1] J.-R. Ohm, "Advances in scalable video coding," *Proceedings of The IEEE,* Vol. 93, No. 1, pp. 42-56, Jan. 2005.

[2] P. Chen, J. Woods, "Bidirectional MC-EZBC with lifting implementation," *IEEE Trans. on CSVT,* Vol. 14, No. 10, pp. 1183-1194, Oct. 2004.

[3] J.-R. Ohm, "Three-dimensional subband coding with motion compensation," *IEEE Trans. on IP*, Vol. 3, No.9, pp. 559-571, Sept. 1994.

[4] S. Choi, J. Woods, "Motion-compensated 3-D subband coding of video," *IEEE Trans. on IP*, Vol. 8, No. 2, pp. 155-167, 1999.

[5] S. Hsiang, J. Woods, J-R. Ohm, "Invertible temporal subband/wavelet filter banks with half-pixel-accurate motion compensation," *IEEE Trans. on IP*, Vol. 13, No. 8, pp. 1018-1028, Aug. 2004.

[6] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. on SP*, Vol. 41, No. 12, pp. 3445-3462, Dec. 1993.

[7] A. Said, W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. on CSVT,* Vol. 6, No. 6, pp. 243-250, 1996.

[8] B.-J. Kim, Z. Xiong, W. Pearlman, " Low bit-rate scalable video coding with 3-D set partitioning in hierarchical trees (3-D SPIHT)," *IEEE Trans. on CSVT*, Vol. 10, No. 8, pp. 1374-1387, Dec. 2000.

[9] C. He, J. Dong, Y. Zheng, and Z. Gao, "Optimal 3-D coefficient tree structure for 3-D wavelet video coding," *IEEE Trans. on CSVT*, Vol. 13, No. 10, pp. 961-972, Oct. 2003.

[10] S.-T. Hsiang, J. Woods, "Embedded image coding using zeroblocks of subband/wavelet coefficients and context modeling," *IEEE International Symposium on Circuits and Systems*, Geneva, Vol. III, pp. 662-665, 2000.