EFFICIENT THREE-DIMENSIONAL WAVELET CODECS FOR NETWORKED VIDEO COMMUNICATIONS

Jozsef Vass and Xinhua Zhuang

Multimedia Communications and Visualization Laboratory
Department of Computer Engineering & Computer Science
University of Missouri-Columbia
Columbia, MO 65211, USA

ABSTRACT

Based on our previously developed significance-linked connected component analysis (SLCCA) data organization and representation strategy, two three-dimensional wavelet video coding techniques are proposed covering a wide range of applications areas. In three-dimensional SLCCA (3D-SLCCA), both the wavelet decomposition and SLCCA data organization and representation strategy are extended to 3-D case. In Haar wavelet SLCCA (H-SLCCA), the SLCCA data organization and representation strategy is separately applied on each of the two temporal subbands obtained by Haar wavelet-based temporal decomposition. The developed codecs not only provide comparable or superior performance in comparison to state-of-the-art video coding standards, but also have several desirable properties such as scalability, low computational complexity, and error resilience, well suited for networked video communications.

1. INTRODUCTION

The increasing importance of video communications is evidenced by extensive research in both academia and industry resulted in a large number of standards adopted. Clearly, the so-called *hybrid* scheme (motion estimation and motion compensation followed by discrete cosine transform (DCT)) is the most widely used approach in today's video technology. Although provides high coding performance, its drawbacks include high computational complexity, blocking artifacts and mosquito noise at low bit rates, and inefficient support for scalable representation and coding. As a possible alternative, three-dimensional (3-D) wavelet video coding has been proposed, where the wavelet decomposition is extended to include the time domain. The advantages of 3-D wavelet video coding include (1) symmetric low computational complexity of both the encoder and decoder due to replacing computationally expensive motion estimation with temporal filtering, (2) scalability supporting multiresolution transmission and display, and (3) robustness against transmission error propagation due to the lack of both spatial prediction and recursive loop in the coder architecture as is the case with hybrid coders.

In most of the 3-D wavelet video coding algorithms, 2-D spatial decomposition is separately carried out on each of the two temporal subbands obtained by Haar wavelet-based temporal decomposition [1, 2, 3] and wavelet coefficients are encoded by traditional methods. Most of the wavelet video codecs applying *advanced data organization and representation* strategies belong to hybrid structure, where 2-D wavelet transform and data organization and representation strategy replace DCT [4, 5, 6, 7]. There were only few wavelet video codecs reported in the literature [8, 9, 10], which applied advanced data organization and representation strategies directly to 3-D subband decomposition.

2. PROPOSED VIDEO CODING ALGORITHMS

In the paper, our previously developed significance-linked connected component analysis (SLCCA) [11, 12, 13] data organization and representation strategy is extended to wavelet video coding. In the followings, the two proposed video coding schemes are detailed.

2.1. Three-Dimensional Significance-Linked Connected Component Analysis

In three-dimensional SLCCA (3D-SLCCA) (Fig. 1), 3-D wavelet decomposition is carried out on N frames. The applied decomposition is natural extension of the

---

This work was supported in part by NASA Applied Information Systems Research award NAG-2573, NASA Innovative Research award NAGW-4698, and NASA Remote Sensing Center Award NAG-13.

2-D wavelet decomposition to include the time domain, i.e., iteratively the lowpass spatial subband is split into two temporal subbands and the lowpass temporal subband is decomposed into four spatial subbands. Based on the block size N either Haar wavelet or Daubechies 9/7 biorthogonal filter is used for temporal decomposition. After 3-D wavelet decomposition, conditioned dilation operation is applied to progressively segment the significance map into 3-D connected components by using 3-D symmetric structuring elements. As a result, 3D-SLCCA provides high coding performance and scalability with moderate robustness against transmission error propagation. Thus it is well applicable for both low bit rate and high bit rate wired video distribution, where the same video must be represented at different resolution, frame rate, and quality depending on both the networking capabilities and hardware resources of receivers.

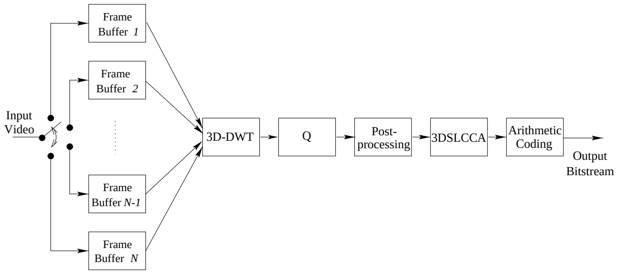


Figure 1: Block diagram of 3D-SLCCA encoder.

2.2. Haar Wavelet Significance-Linked Connected Component Analysis

The block diagram of Haar wavelet SLCCA (H-SLCCA) is shown in Fig. 2. As seen, the temporal lowpass and highpass subbands are obtained as the sum and difference of two consecutive input frames, respectively. Following dyadic spatial decomposition of each of the temporal subbands, significance wavelet coefficients are organized and represented by using the SLCCA technique. As is evidenced by our experiments, applying dyadic wavelet decomposition on both temporal subbands not only slightly improves the performance, but also decreases the computational complexity. More importantly, the SLCCA data organization and representation strategy can be directly applied without any modification. As a result, H-SLCCA provides high robustness against transmission error propagation and low coding delay thus being an excellent choice for both mobile video communications and interactive applications such as teleconferencing and videophony.

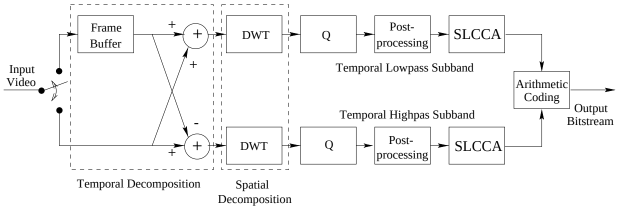


Figure 2: Block diagram of H-SLCCA encoder.

3. PROPERTIES OF 3D-SLCCA AND H-SLCCA

Both 3D-SLCCA and H-SLCCA provide video streams with the following properties desirable for networked video applications [14]:

High coding performance: High coding performance is necessary in order to provide the user with high video quality over low bandwidth connections. As we will demonstrate, 3D-SLCCA outperforms MPEG-2, and provides acceptable performance loss in comparison to the state-of-the-art H.263+ standard [15]. Among the same circumstances, H-SLCCA renders superior performance when compared to H.263+.

Scalable coding: Both 3D-SLCCA and H-SLCCA provide *resolution*, *quality*, *temporal*, and *rate* scalability as well. Resolution scalability is obtained by decoding only certain spatial subbands. Quality scalability is achieved by transmitting bit-planes in order of importance. Temporal scalability is implemented by decoding certain temporal subbands. Finally, rate scalability means that the bitstream is ordered in decreasing importance thus the truncated part is less important than the kept part.

The resolution, frame rate, and quality of the received video is decided depending on both the available bandwidth and hardware capabilities of the receiver. Clients with high bandwidth network connections can receive high quality color video at high frame rate while clients with low bandwidth connections may only receive low quality and/or grayscale video at low frame rate.

Due to its prime importance, scalability is also part of both MPEG-2 and H.263+, where spatial and quality scalability are achieved by encoding possibly several enhancement layer in addition to the base layer. Temporal scalability in H.263+ is implemented by dropping bidirectional predicted frames (B-frames). In the performance evaluation [16], however, it was shown clearly that even a two-layer spatial and quality scalable bitstream may be only achieved at the price of 1–2 dB drop in peak signal-to-noise ratio (PSNR). Although temporal scalability introduces only minor PSNR degradation, it requires B-frames, which are not admissible in coding delay critical applications. As seen, the developed video coding algorithms naturally support wide range of scalability without introducing any performance degradation or coding delay.

Well defined high priority information: To minimize the effects of transmission errors, it is desirable that the most important part of the bitstream be easily identifiable and separable so that it can be transmitted more reliably. In the proposed video coding algorithms, the importance of each bit-plane within each subband can be very easily determined by calculating the contained energy, which can be used to decide the proper transmission error protection strategy.

Low computational complexity: Both the encoder and decoder of the proposed video codecs have very low approximately symmetric computational complexity, which is due to that there is no optimal bit allocation needed and the computationally expensive block-based motion estimation and motion compensation are discarded.

Error resilience: Due to unavoidable transmission errors, techniques that limit the effect of channel errors in the smallest possible temporal and spatial neighborhood are necessary. In block-based hybrid video coding algorithms, an I-frame is followed by several predicted frames (P-frames). Due to the spatial prediction being used in motion estimation and motion compensation, the effects of an uncorrectable error may not only propagate for several consecutive frames till the next I-frame is received, but also spread out spatially in a larger neighborhood [17]. This undesirable error propagation is largely prevented in the proposed video codecs as follows. The spatial error propagation is significantly reduced in both 3D-SLCCA and H-SLCCA due to the lack of motion estimation and motion compensation. While in H-SLCCA two frames are coded as one unit which prevents temporal error propagation, in 3D-SLCCA, the temporal error propagation is limited to the block size N.

4. PERFORMANCE EVALUATION

Extensive performance evaluation and comparison with state-of-the-art video coding standards are carried out on several test video sequences covering a wide range of application areas. All the reported results are computed from the decoded bitstream.

Performance comparison of 3D-SLCCA and 3D-SPIHT is carried out on the first 48 frames of the grayscale “Table Tennis” sequence in SIF resolution obtained from Rensselaer Polytechnic Institute. As seen from Table 1, 3D-SLCCA outperforms 3D-SPIHT in PSNR by 0.24 dB and 0.31 dB at 0.3 bits-per-pixel (bpp) and 1.0 bpp, respectively.

Bit Rate [bpp]	3D-SPIHT	3D-SLCCA
0.3	27.90	28.14
1.0	34.20	34.51

Table 1: Performance comparison of 3D-SLCCA and 3D-SPIHT on 48 frames of the grayscale “Football” sequence.

Performance comparison of 3D-SLCCA and H.263+ (using predefined Level 3 preferred combination without PB frames) on standard MPEG-4 test sequences in QCIF resolution for low bit rate coding is shown in Table 2. On average, H.263+ outperforms 3D-SLCCA by 1.96 dB, 1.22 dB, and 0.55 dB, for the Y, U, and V components, respectively.

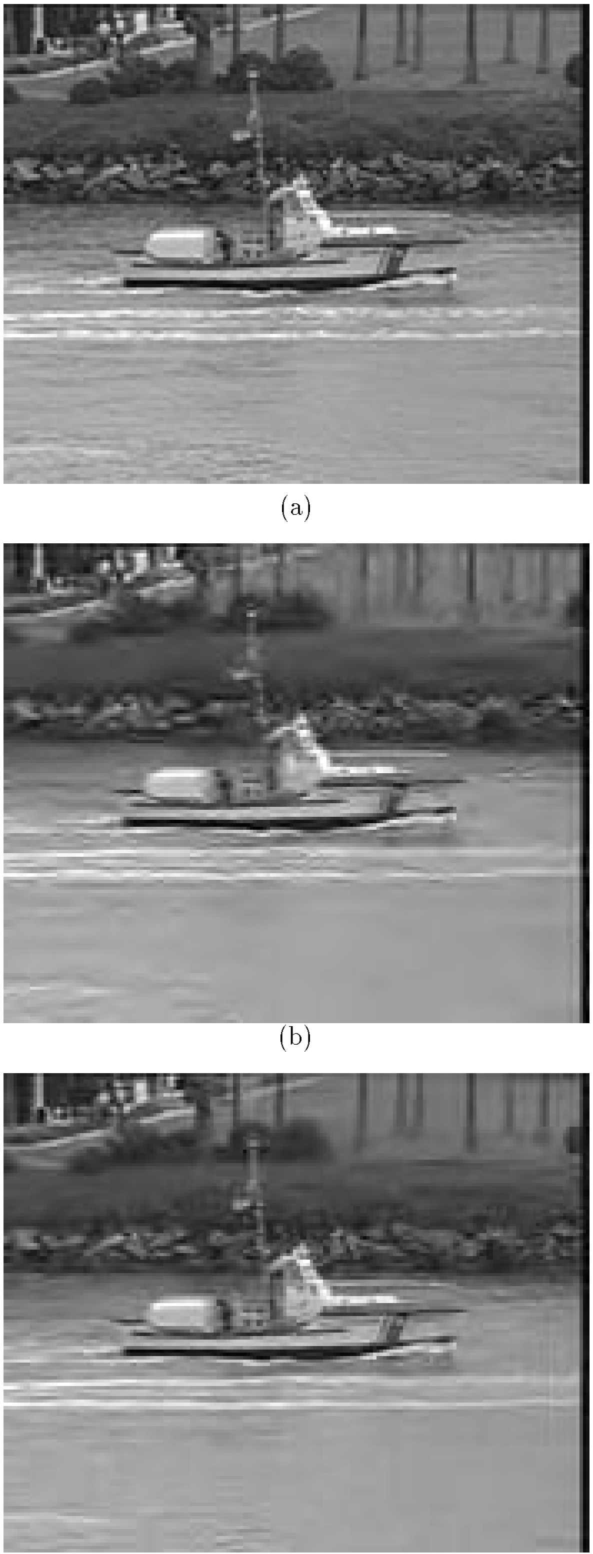
Sequence	Luminance Y, PSNR [dB] Chrominance U V, PSNR [dB]			
	H.263+		3D-SLCCA	
Akiyo	34.55		32.73	
5 fps, 10 kbps	39.57	41.31	37.40	40.35
Mother Daughter	33.41		31.96	
5 fps, 10 kbps	40.42	40.89	39.70	40.58
Foreman	31.87		28.92	
7.5 fps, 48 kbps	38.71	38.91	36.91	36.88
Coast Guard	29.69		27.97	
7.5 fps, 48 kbps	40.67	42.42	40.50	43.54

Table 2: Performance comparison (PSNR, [dB]) of 3D-SLCCA and H.263+.

Fig. 3 shows the original and reconstructed 192th frames by 3D-SLCCA and H.263+ of the “Coast Guard” sequence sampled at 7.5 frames-per-second (fps) and coded at 48k bits-per-second (bps). While H.263+ provides higher objective performance, the visual quality of the reconstructed frames by H.263+ and 3D-SLCCA is compatible with only a little loss in fine details by 3D-SLCCA.

Performance comparison of 3D-SLCCA and MPEG-2 is carried out on the first 144 frames of four standard MPEG-2 test sequences in ITU-T 601 resolution sampled at 30 fps and coded at 15 Mbps. As seen from Table 3, 3D-SLCCA on average outperforms MPEG-2 by 0.32 dB for the luminance component and 0.21 dB and 0.06 dB for the U and V chrominance components, respectively.

Finally, coding results for H.263+ and H-SLCCA are presented for low motion MPEG-4 test sequences in QCIF resolution sampled at 10 fps and encoded at 48 kbps. In both algorithms, two frames constitute a



(a)

(b)

(c)

Figure 3: Interframe coding results for the "Coast Guard" sequence at 7.5 fps, 48 kbps. (a) The original 192th frame of the sequence. Reconstructed frames by (b) 3D-SLCCA and (c) H.263+.

Sequence	Luminance Y, PSNR [dB] Chrominance U V, PSNR [dB]			
	MPEG-2		3D-SLCCA	
Flower Garden	36.97		37.31	
	38.24	38.82	39.03	38.52
Mobile Calendar	34.31		33.68	
	37.73	38.16	36.49	37.01
Football	36.99		37.70	
	38.90	39.30	39.23	39.87
Table Tennis	36.79		37.65	
	41.22	42.07	42.16	43.20

Table 3: Performance comparison (PSNR, [dB]) of 3D-SLCCA and MPEG-2 at 30 fps, 15 Mbps.

block. As seen form Table 4, for the luminance component, H-SLCCA outperforms H.263+ by 0.22 dB on average. For the chrominance components, H-SLCCA is superior by 0.35 dB and 0.18 dB on average for the U and V components, respectively.

Sequence	Luminance Y, PSNR [dB] Chrominance U V, PSNR [dB]			
	H.263+		H-SLCCA	
Akiyo	31.60		32.05	
	36.29	38.86	37.18	38.78
Hall Monitor	28.14		28.24	
	37.07	39.26	37.70	40.59
Mother Daughter	32.96		33.04	
	39.81	40.68	39.77	40.41
Container Ship	28.13		28.36	
	37.94	37.15	37.86	36.87

Table 4: Performance comparison (PSNR, [dB]) of H-SLCCA and H.263+ at 10 fps, 48 kbps.

Timing comparison between 3D-SLCCA and H.263+ for 144 frames of the "Coast Guard" sequence is done on a 195 MHz R10000 CPU of an SGI Octane workstation. As shown in Table 5, 3D-SLCCA requires 13.97 seconds and 10.6 seconds for the encoding and decoding, respectively. For the same sequence, H.263+ spends 108.34 seconds for encoding and 9.23 seconds for decoding.

Table 6 shows the timing results for H-SLCCA for the "Akiyo" sequence sampled at 10 fps and coded at 48 kbps. Encoding 100 frames takes 5.08 seconds for H-SLCCA vs. 34.61 seconds for H.263+. Decoding takes 3.51 seconds for H-SLCCA in comparison with 3.82 seconds for H.263+. If both the encoder and decoder are executed on the same CPU, 11.64 fps or CPU time of less than 85 ms per frame can be achieved by H-SLCCA.

Function	Encoder Time [sec]	Decoder Time [sec]
Wavelet Transformation	4.39	4.45
Quantization and Postprocessing	4.09	–
SLCCA Coding	5.49	6.15
Total	13.97	10.6

Table 5: 3D-SLCCA timing results (seconds) for 144 frames of the color "Coast Guard" sequence in QCIF resolution.

Function	Encoder Time [sec]	Decoder Time [sec]
Wavelet Transformation	1.93	1.88
Quantization and Postprocessing	1.71	–
SLCCA Coding	1.44	1.63
Total	5.08	3.51

Table 6: H-SLCCA timing results (seconds) for 100 frames of the color "Akiyo" sequence in QCIF resolution.

5. CONCLUSIONS

In the paper, both of the proposed video coding algorithms are built on the highly efficient SLCCA data organization and representation strategy covering a wide range of application areas ranging from very low bit rate videophony and teleconferencing through mobile video to high bit rate video broadcast. Although renders inferior performance when compared to H.263+, 3D-SLCCA provides highly scalable bitstream without performance loss or coding delay. For high bit rate coding, 3D-SLCCA outperforms MPEG-2. Among the same circumstances, H-SLCCA outperforms H.263+. In addition, both codecs provide scalability, low complexity, and error propagation immunity, that will play an increasingly important role in tomorrow's video communications.

6. REFERENCES

[1] G. Karlsson and M. Vetterli, "Three dimensional subband coding of video," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1988, pp. 1110–1113.

[2] C.I. Podilchuk, N.S. Jayant, and N. Farvardin, "Three-dimensional subband coding of video," *IEEE Transactions on Image Processing*, vol. 4, no. 2, pp. 125–139, Feb. 1995.

[3] C.-H. Chou and C.-W. Chen, "A perceptually optimized 3-D subband codec for video communication over wireless channels," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 2, pp. 143–156, Apr. 1996.

[4] G. Bhutani and W.A. Pearlman, "Image sequence coding using the zero-tree method," in *Proceedings of SPIE Conference Visual Communications and Image Processing*, 1993, vol. 2094, pp. 463–471.

[5] S.A. Martucci, I. Sodagar, T. Chiang, and Y.-Q. Zhang, "A zerotree wavelet video coder," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 1, pp. 109–118, Feb. 1997.

[6] J. Vass, B.-B. Chai, and X. Zhuang, "Significance-linked wavelet video coder," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Seattle, WA, May 1998, pp. 2829–2832.

[7] J. Vass, B.-B. Chai, K. Palaniappan, and X. Zhuang, "Significance-linked connected component analysis for very low bit rate wavelet video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 4, pp. 630–647, June 1999.

[8] Y. Chen and W.A. Pearlman, "Three-dimensional subband coding of video using the zero-tree method," in *Proceedings of SPIE Conference Visual Communications and Image Processing*, 1996, vol. 2727, pp. 1302–1312.

[9] B.-J. Kim and W.A. Pearlman, "An embedded wavelet video coder using three-dimensional set partitioning in hierarchical trees (SPIHT)," in *Proceedings of Data Compression Conference*, 1997, pp. 251–260.

[10] J. Vass, B.-B. Chai, and X. Zhuang, "3DSLCCA – A highly scalable very low bit rate software-only wavelet video codec," in *Proceedings of IEEE Workshop on Multimedia Signal Processing*, Los Angeles, CA, Dec. 1998, pp. 474–479.

[11] B.-B. Chai, J. Vass, and X. Zhuang, "Highly efficient codec based on significance-linked connected component analysis of wavelet coefficients," in *Proceedings of SPIE AeroSense*, 1997.

[12] B.-B. Chai, J. Vass, and X. Zhuang, "Significance-linked connected component analysis for low bit rate image coding," in *Proceedings of IEEE International Conference on Image Processing*, Oct. 1997, pp. 637–640.

[13] B.-B. Chai, J. Vass, and X. Zhuang, "Significance-linked connected component analysis for wavelet image coding," *IEEE Transactions on Image Processing*, vol. 8, no. 6, pp. 774–784, June 1999.

[14] M.R. Civanlar, "Protocols for real-time multimedia data transmission over the Internet," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1998, pp. 3809–3812.

[15] ITU-T Draft Recommendation H.263 Version 2 (H.263+), "Video coding for low bitrate communications," Sept. 1997.

[16] G. Côté, B. Erol, M. Gallant, and F. Kossentini, "H.263+: Video coding at low bit rates," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 7, pp. 849–866, Nov. 1998.

[17] Y. Wang and Q.-F. Zhu, "Error control and concealment for video communications: A review," *Proceedings of IEEE*, vol. 86, no. 5, pp. 974–997, May 1998.