3DSLCCA – A HIGHLY SCALABLE VERY LOW BIT RATE SOFTWARE-ONLY WAVELET VIDEO CODEC

Jozsef Vass, Bing-Bing Chai,[†] and Xinhua Zhuang
Multimedia Communications and Visualization Laboratory
Department of Computer Engineering and Computer Science
University of Missouri-Columbia
Columbia, MO 65211, USA

[†]Sarnoff Corporation
Princeton, NJ 08543, USA

Abstract - In the paper, our high performance significance-linked connected component analysis (SLCCA) image compression algorithm is extended to three-dimensional wavelet coding resulting in a highly efficient and scalable video codec termed 3DSLCCA. Advantageous features of 3DSLCCA include low computational complexity, rate, spatial, temporal, and SNR scalability, and prevention of error propagation at the price of acceptable degradation in both objective and subjective quality when compared to the state-of-the-art H.263 standard.

INTRODUCTION

Very low bit rate video coding has triggered intensive research in both academia and industry. The adopted ITU-T H.263 Recommendation which is specifically tailored for very low bit rate video coding [1] is the first standard to break the 64k bits-per-second (bps) barrier in audio-visual communications. The main application areas of H.263 are videoconferencing and videophony. Although H.263 provides acceptable visual quality, its widespread use is limited due to the high computational complexity of the encoder, requiring hardware implementation.

As an alternative to the most popularly used hybrid video coding scheme, three-dimensional (3-D) wavelet video coding algorithms were proposed [2]. The extension of 2-D subband image coding to include the temporal domain naturally leads to 3-D (temporal $\times$ spatial $\times$ spatial) subband video coding algorithms. The advantages of 3-D wavelet video coding schemes include scalability, low computational complexity, and prevention from error propagation. These advantages over traditional hybrid video coding algorithms have triggered intensive research, resulting in several 3-D subband video coding algorithms being published [3, 4, 5, 6]. The main problem with the 3-D approach is correctly recognized that the temporal filtering has not been as efficient as the time domain block-based motion estimation/compensation algorithms in terms of the exploitation of the temporal redundancy in video sequences. To alleviate this problem motion adaptation schemes have been proposed earlier in [4, 5] that the temporal filtering is performed along the motion trajectory. Due to motion estimation which is obviously needed in motion adaptation schemes, the low computational complexity is compromised.

DATA ORGANIZATION AND REPRESENTATION

Significance-linked connected component analysis (SLCCA) [7, 8] represents a high performance wavelet image coding algorithm. As opposed to zerotree-based algorithms such as embedded zerotree wavelet (EZW) [9] and set partitioning in hierarchical trees (SPIHT) [10], it follows morphological representation of wavelet data (MRWD) [11] in spirit to directly deal with significant coefficients organized as irregularly shaped clusters. But SLCCA exploits not only the *within subband* clustering property but also the *cross-scale* dependency among clusters. As was implied by the *cross-scale* statistical decaying of magnitude of wavelet coefficients [9, 8], a significant cluster at a coarse scale likely can be linked to a significant cluster at a finer scale. This significance-linked connected component analysis or clustering renders the highest peak signal-to-noise ratio (PSNR) with SLCCA in comparison to other three wavelet image coders.

Despite the striking success in still image coding, there have been only a few successful attempts to use innovative data organization and representation strategies for wavelet video coding. Bhutani and Pearlman proposed to use Shapiro's EZW algorithm to encode error frames obtained by recursive motion compensation [12]. Kim and Pearlman applied SPIHT to 3-D subband video coding [6]. In zerotree entropy (ZTE) [13] coder, after block-based motion estimation and compensation similar to that of H.263, an EZW variant algorithm was used for the representation and encoding of motion-compensated error frames. Recently, Vass *et al.* [14] successfully applied SLCCA for hybrid video coding in which a fine-tuned time domain motion estimation and exhaustive overlapped block motion compensation were used to ensure coherency of motion-compensated residual error frames, and results superior to H.263 were reported. Most of the recently developed wavelet video coding algorithms use some kind of motion estimation technique to maintain a high objective performance. Unfortunately, they do not offer scalability, low computational complexity, and prevention from error propagation.

3DSLCCA CODING ALGORITHM

In 3DSLCCA, two kinds of 3-D wavelet decomposition are available. In coding-delay-critical applications, such as videophony and videoconferencing, to ensure a minimal coding delay the Haar wavelet is used to obtain two temporal subbands, and each subband is further spatially decomposed. When the coding delay is not of prime importance, such as in video-on-demand or digital video broadcast, we adopt the 3-D subband decomposition structure shown in Fig. 1, which is the natural extension of 2-D subband decomposition to include the temporal domain, in which only the lowpass subband is split further into two temporal or four spatial subbands. In 3DSLCCA, different filters can be specified for the temporal and spatial filtering. For the spatial decomposition, we use the Daubechies 9/7 biorthogonal filter, and for the temporal decomposition either the 9/7 filter or the Haar wavelet is used. The parent-child relationship is naturally extended from the 2-D case to 3-D as described in [6].

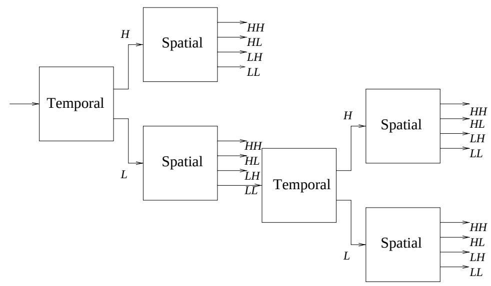

Figure 1: 3-D wavelet decomposition.

Both EZW and SPIHT provide a fully embedded bitstream, which means that information about significant coefficients is transmitted in the order of their contribution to the objective quality irrespective of the position. Therefore, when those embedded image coders are extended to video coding, they only provide rate scalability as described in [6]. The problem of temporal, spatial, and SNR scalability has been addressed recently in [15] via reordering the bitstream resulting in the loss of embeddedness.

In 3DSLCCA, the spatial and temporal scalability are obtained via transmitting only those subbands that are used in the reconstruction. SNR scalability is achieved by transmitting the most significant bit planes until the given bit budget is exhausted. Finally, rate scalability is ensured by transmitting the magnitudes of significant coefficients in subband-wise and in bit plane order. Based on the available bandwidth and computing power, the encoder and decoder can *automatically* negotiate the spatial, temporal, and SNR resolution as well.

Since scalability is an invaluable feature in multimedia communications, the temporal, spatial, and SNR scalability became part of the recently adopted H.263+ standard [16]. In H.263+, the temporal scalability is implemented by dropping B-pictures while the spatial and SNR scalability are achieved by encoding possibly several enhancement layers in addition to the base layer. In the performance evaluation [16], however, it was shown clearly that even a two-layer scalable bitstream could only be achieved at the price of 1–2 dB drop in PSNR. By contrast, 3DSLCCA provides scalability with fine granularity without sacrificing the coding performance.

PERFORMANCE EVALUATION

The performance of 3DSLCCA against H.263 is compared by using a low motion (Class A) test sequence "Mother & Daughter" coded at 24 kbps, and a high motion (Class B) test sequence "Coast Guard" coded at 48 kbps. Both sequences are in QCIF resolution sampled at 10 fps. For H.263, unrestricted motion vector mode and advanced prediction mode (annexes D and F, respectively) are used. The PSNR comparison of the above two sequences are shown in Fig. 2. On average, H.263 outperforms 3DSLCCA by 1.51 dB and 1.77 dB in PSNR for the "Mother & Daughter" and "Coast Guard" sequences, respectively. The visual comparison of the 87th frame of the "Mother & Daughter" sequence and the 111th frame of the "Coast Guard" sequence are shown in Fig. 3. While the visual quality of 3DSLCCA and H.263 for the "Mother & Daughter" sequence is compatible, 3DSLCCA introduces a little loss in fine details for the "Coast Guard" sequence.

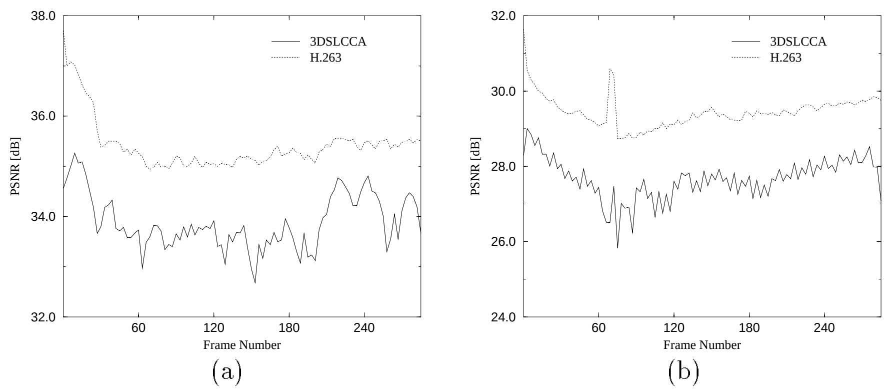

Figure 2: Performance comparison (PSNR, [dB]) between 3DSLCCA and H.263. (a) "Mother & Daughter" sequence at 10 fps, 24 kbps. Average PSNR are 33.88 dB and 35.39 dB for 3DSLCCA and H.263, respectively. (b) "Coast Guard" sequence at 10 fps, 48 kbps. Average PSNR are 27.66 dB and 29.43 dB for 3DSLCCA and H.263, respectively.

Timing experiments are performed on an SGI O_2 workstation with a single 150 MHz R10000 processor. 256 frames of the "Coast Guard" sequence are coded in QCIF resolution. As is shown in Table 1, the execution time of the encoder and decoder are nearly symmetric, and real time encoding and decoding are possible. When both the encoder and decoder are executed on a single CPU, 14 fps can be achieved, which is more than sufficient for videoconferencing or videophony applications. On the other hand, encoding and decoding of the same sequence under the same circumstances takes 63.91 seconds for the H.263 test model, which hinders real-time software-only implementation.

CONCLUSIONS

In the paper, a very efficient 3-D wavelet video coding algorithm termed 3DSLCCA is developed that provides invaluable features such as rate, temporal, spatial, and SNR scalability, prevention of error propagation, and real time software-only implementation with an acceptable overall performance degradation.

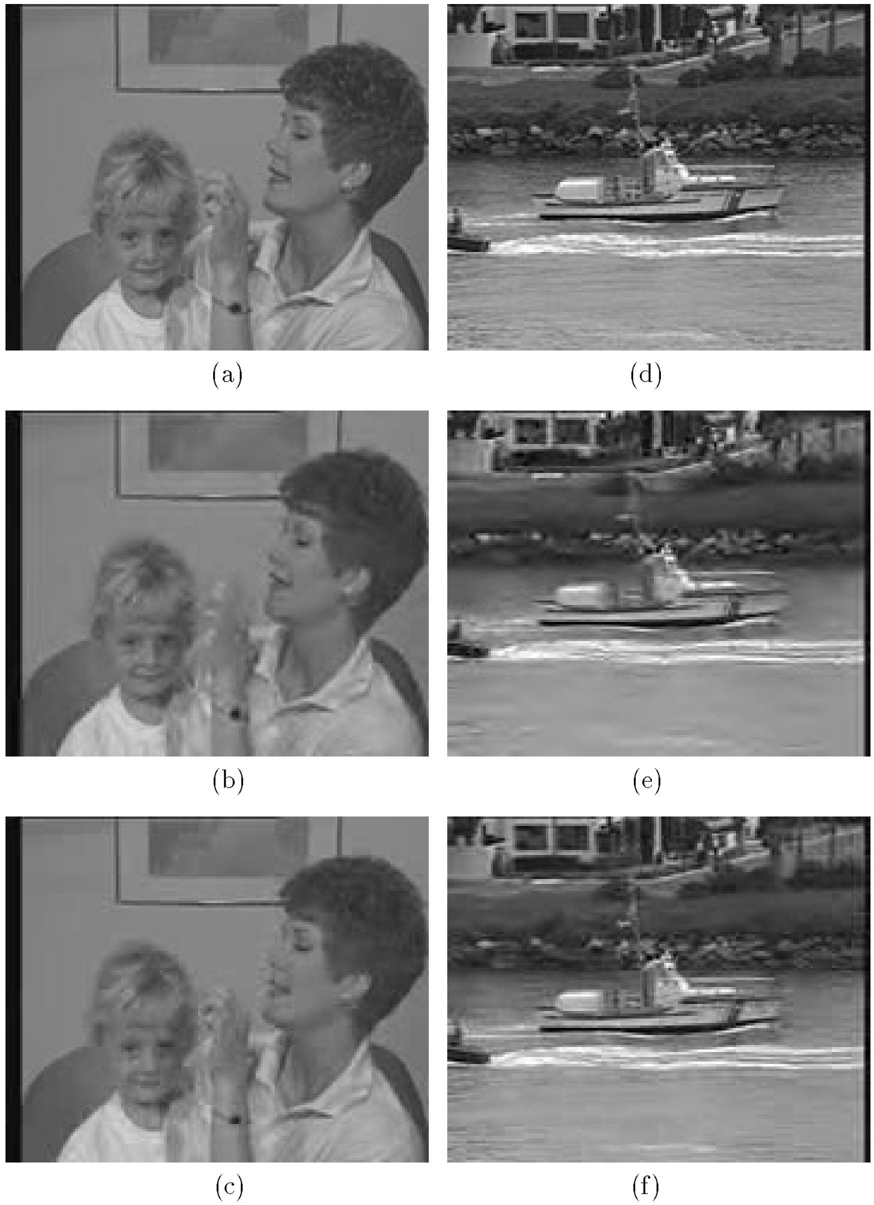

Figure 3: Interframe coding results for the 87th frame of the "Mother & Daughter" sequence (a)–(c) at 10 fps, 24 kbps and the 111th frame of the "Coast Guard" sequence (d)–(f) at 10 fps, 48 kbps. Original frames (a) and (d), and the reconstructed frames by 3DSLCCA (b) and (e), and by H.263 (c) and (f).

Function	Encoder	Decoder
Wavelet Decomposition/Reconstruction	3.42	3.31
Cluster Elimination	4.38	-
SLCCA Encoding/Decoding	3.19	3.89
Overall	10.99	7.20

Table 1: Timing results (seconds) for 3DSLCCA on an SGI O_2 workstation for 256 frames in QCIF resolution.

References

[1] ITU-T Draft Recommendation H.263, "Video coding for low bitrate communications," Dec. 1995.

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

[3] 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.

[4] D. Taubman and A. Zakhor, "Multirate 3-D subband coding of video," *IEEE Transactions on Image Processing*, vol. 3, no. 4, pp. 572–590, Sept. 1994.

[5] J.-R. Ohm, "Three-dimensional subband coding with motion compensation," *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 559–571, Sept. 1994.

[6] B.-J. Kim and W. 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.

[7] 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.

[8] B.-B. Chai, J. Vass, and X. Zhuang, "Significance-linked connected component analysis for wavelet image coding," *IEEE Transactions on Image Processing, to appear*, 1998.

[9] J.M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3445–3462, Dec. 1993.

[10] A. Said and W.A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 243–250, June 1996.

[11] S. Servetto, K. Ramchandran, and M.T. Orchard, "Wavelet based image coding via morphological prediction of significance," in *Proceedings of IEEE International Conference on Image Processing*, Oct. 1995, pp. 530–533.

[12] 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.

[13] 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.

[14] 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*, May 1998, pp. 2829–2832.

[15] Z. Xiong, B.-J. Kim, and W.A. Pearlman, "Multiresolution encoding and decoding in embedded image and video coders," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 1998, pp. 3709–3712.

[16] G. Côté, B. Erol, M. Gallant, and F. Kossentini, "H.263+: Video coding at low bit rates," *Submitted to IEEE Transactions on Circuits and Systems for Video Technology*, 1998.