Temporal and Spatial Scalability with 3D Wavelet Transforms

Calvin Lew <cdlew@stanford.edu>
Koohyar Minoo <kminoo@apple.com>
Daniel Simms <dsimms@dsimms.com>

April 22, 2002

Scalable video encoding allows specific subsets of the video bitstream with different transport and presentation properties, e.g. different bitrates and different temporal/spatial characteristics. These subsets are layered such that the delivery and decoding of the stream can change properties without too much complexity. Scalability presents a significant opportunity in today's diverse applications of streaming media. From HDTV migration scenarios to error resilience solutions, when multiple channels are available for broadcast of a video stream, scalability plays an important role.

Existing hybrid coding schemes like MPEG-1 and MPEG-2 which use block-based DCT are not optimal for temporal or spatial scalable coding[5]. Previous work has been done with subband coding, specifically wavelet transforms[6], to achieve spatial scalability. Furthermore, wavelet transforms have been extended into the temporal dimension, and, coupled with motion compensation prediction, have resulted in new techniques with scalable temporal resolution[2].

Previous papers have mainly concentrated on one or two wavelet bases, such as the Haar wavelet[1]. We intend to look at different wavelet bases for the temporal and spatial dimensions. We will explore different 3-D wavelet transforms for scalable frame rates and scalable frame sizes. We divide our project into two main phases. In the first phase, we will implement an encoder with the necessary functional blocks, such as motion-compensation prediction module, wavelet transform modules, and quantizer module, to provide a working framework but also one which we will be able to readily modify for the experiments in the second phase. We will use a block-based motion compensated predictor that utilizes the MCP information from lower spatial layers. Our predictor will help in the energy compaction of our signal in the temporal domain. Initially, we assume an ideal entropy-rate coder for ease of implementation. We also will use a quantizer that will minimize MSE and give good performance. With 3-D wavelets, we have the added freedom of scalable frame sizes, which allows us to add a new ingredient in exploring the tradeoffs of bitrate, distortion, and spatial resolution. Using the wavelets that we find to be most viable for our encoding scheme, we will collect rate-distortion data of different video streams. A major aim of the data will be measuring the performance of temporal scalability with fixed spatial resolution and, vice versa, spatial scalability with fixed temporal resolution.

References:

[1] J.R. Ohm, "Three-Dimensional Subband Coding with Motion Compensation," IEEE Trans. Imaging Processing, vol. 3, no. 5, pp. 559-571, September 1994.

[2] A. Seeker, D. Taubman, "Motion-compensated highly scalable video compression using and adaptive 3D wavelet transform based on lifting," Proc. IEEE Int. Conf. Image Processing, Thessaloniki, Greece, pp. 1029-1032, October 2001.

[3] D. Taubman, A. Zakhor, "Rate- and Resolution- Scalable Video and Image Compression with Subband Coding," 27th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, California, vol. 2, pp. 1489-1493, November 1993.

[4] P.C. Cosman, R.M. Gray, M. Vetterli, "Vector Quantization of Image Subbands: A Survey," IEEE Trans. Imaging Processing, vol. 5, no. 2, pp. 202-225, February 1996.

[5] J.W. Woods, G. Lilienfield, "A resolution and frame-rate scalable subband/wavelet video coder," IEEE Trans. Circuits and Systems for Video Technology, vol. 11, no. 9, pp. 1035-1044, September 2001.

[6] K. Ramchandran, M. Vetterli, "Best wavelet packet bases in a rate-distortion sense," IEEE Trans. Image Processing, vol. 2, pp. 160-175, April 1993.