



3D Wavelet Transform of Motion-Compensated Highly Scalable Video Compression

P.Sravan Kumar

Assistant Professor, Department of MECH
Sri Sai Institute of Technology and Science, Rayachoti
Email: nasravan350@gmail.com

Abstract: - *The following paper introduces a low bit rate video coding method on the basis of 3D motion compensated wavelet transform. Motion compensated transform techniques, as distinct from motion compensated predictive coding, represent a key tool in the development of highly scalable video compression algorithms. Motion vectors are compressed, too, therefore the time and spatial redundancy of coding is exploited here as well. A scalable video coder cannot be equally efficient over a wide range of bit-rates unless both the video data and the motion information are scalable. The proposed method involves the construction of quality layers for the coded sample data and a separate set of quality layers for the coded motion parameters. When the motion layers are truncated, the decoder receives a quantized version of the motion parameters used to code the sample data. Experimental results indicate that the cost of scalability is small. In addition, considerable performance improvements are observed at low bit-rates, relative to lossless coding of the motion information.*

Keywords: 3D Wavelet transform, Video coding, Motion compensation, MPEG-4, Motion vector compression.

I. Introduction

The objective of highly scalable video coding is to produce a dense family of embedded bit-streams, each an efficient compressed representation of the video, at successively higher bit-rates. Scalability has important applications in many areas including simulcast, videoconferencing and remote video browsing. In addition to bit-rate scalability, other important forms of scalability for video coding include spatial resolution and temporal (frame-rate) scalability. We propose a wavelet-based, highly scalable video compression scheme with rate- scalable motion coding. The proposed method involves the construction of quality layers for the coded sample data and a separate set of quality layers for the coded motion parameters. When the motion layers are truncated, the decoder receives a quantized version of the motion parameters used to code the sample data. The effect of motion parameter quantization on the reconstructed video distortion is described by a linear model. The optimal tradeoff between the motion and sub band bit-rates is determined after compression.

Video is a real-time medium whose transport places exceptional demands on the capacity of existing network infrastructure, so that compression is essential for virtually any digital video communications. For effective utilization of limited channel capacity, it is desirable for the compressed

representation to contain embedded subsets corresponding to successively higher bit-rates and reconstructed video qualities. Such a representation is said to be "rate scalable"; terms such as "SNR scalability" and "distortion scalability" are also

Interestingly, under these conditions the motion compensated lifting scheme proposed here is equivalent to a frame warping scheme. Most importantly, however, the use of separate forward and backward motion models within the lifting framework allows it to effectively exploit motion redundancy in the case of more general (and more realistic) motion, without sacrificing perfect reconstruction.

For scalable video compression, a successful motion compensated DWT should exhibit the following two properties:

- 1) High-pass temporal sub band samples should be close to zero wherever the corresponding original frame samples are predictable through motion modeling. This is important since the energy of the high-pass sub band samples has a direct impact on the transform's coding gain.
- 2) Low-pass temporal sub bands should be as free as possible from ghosting artifacts. That is, they should be believable images of the original scene, taken at the appropriate time instants. This is important both for temporal scalability and to minimize the presence of unwanted edges, which have a direct impact on the effectiveness of subsequent spatial compression techniques.

The proposed scheme exhibits both of these properties. To see why this is so, observe firstly that high-pass sub band frames are none other than the motion compensated residuals, so that their energy depends only upon the success of the motion model. The second property follows from the fact that each lifting step involves the addition (or subtraction) of two frames, one being mapped to the coordinate system of the other, so that spatial features from different parts of the scene will not be super-imposed. In effect, the low-pass sub band frames correspond to temporal filtering of the original frames along the motion trajectory. Moreover, to the extent that the high-pass sub band samples have amplitudes close to 0, it is clear that

$l[m,n]$ will be similar to $x_1[m,n]$.

Of course, both of these properties break down in regions where the motion model is unsuccessful. Nevertheless, the inevitability of the transform is not sacrificed. Moreover, the structure of the transform does not in any way restrict the motion models which may be supported. As mentioned previously, pre- warping methods are limited to exploiting only very simple motion. Therefore such methods cannot possess the above properties while preserving perfect reconstruction.

II. The Operation of the Algorithm

4.1 Motion Compensated 3D Wavelet Transform

The motion compensation method complementing the original 3 D wavelet transform [1, 4, 6] must be inserted into the time domain steps of the transform. Fig. 1 depicts the block diagram of the algorithm, with a GOF (Group of Frames) size of 8 frames. Regarding the time domain transform, Haar base [8] has been chosen. Prior to the time domain transform, an attempt is made to increase the similarity between the input frames using motion compensation. The number of frames used in a filtering phase is equivalent to the length of the filters of the given base, i.e. this is the frame count that must be made similar to each other by motion compensation (the required computational performance and the number of motion vectors to be transferred is proportional to this). As a result, the application of a filter as short as possible is recommended. This induced our decision for Haar base which features an impulse response of a length of 2.

The blocks (LP, HP) below the motion compensation (MCij) in the block diagram correspond to the low pass and high pass filters of the time domain Haar wavelet. In case of exact motion compensation the high frequency components resulted by the high pass filtering will feature small amplitude, increasing the efficiency of the entire compression. An arbitrary base independent of the time domain can be chosen for the spatial transform. The procedure is continued after this (motion compensation, time and spatial domain

transform) for the low frequency components. Having completed every step of the transform, a bit stream is generated from the obtained components by SPIHT algorithm [1, 5], which is also used for the quantization of the bit stream. The beginning of the stream comprises the most significant bits of the individual coefficients, followed by the lower significant bits in decreasing order. If this stream is interrupted somewhere, than the lower significant bits are rejected, i.e. the coefficients are quantized. Depending on the place of interruption, either constant bit rate or constant quality (varying bit rate) coding can be set. Finally, the quantized bit stream is lossless compressed by an entropy encoder.

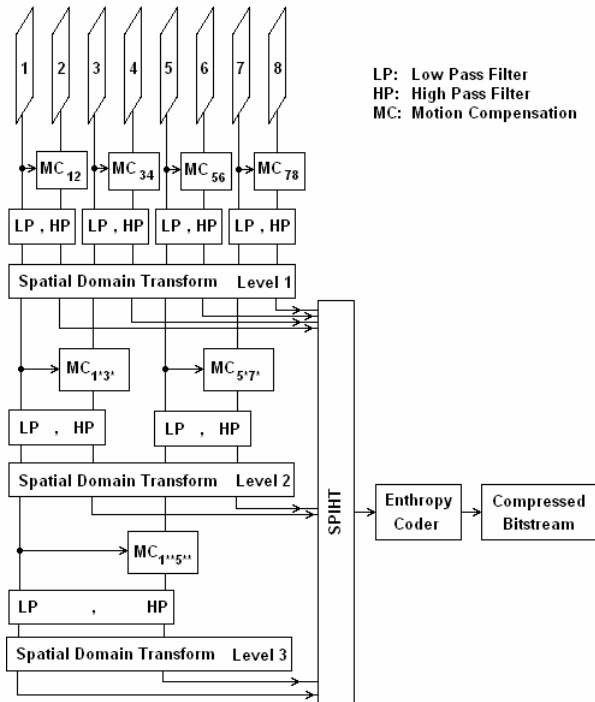


Fig.1. Motion Compensated 3D Wavelet Encoder

4.2 Motion Compensation

For the sake of simplicity, block based motion compensation method has been chosen from the several possibilities available (e.g. block based, mesh based, etc.) [2, 3]. This method works well for linear movements, but is unsuitable for handling more complex (e.g. rotating) movements. A motion estimation algorithm

required for these movements is complicated, giving a further reason for deciding for the block based solution. The selected motion compensation method permits the selection of the block size as well as the dimensions of the search window. Also the less consistent movements can accurately be described if the block sizes are set to smaller, but the increased number of blocks resulted by this approach leads to an increased number of motion vectors, i.e. increases the amount of side information to be transferred.

The amount of side information also depends on the extent of the search domain. Bigger search domains may result longer motion vectors, enabling the handling of faster movements, though leading to more bits in the description of the vectors (in case of an image in CIF format, assuming a block and search domain size of 8x8 and a frame rate of 30 fps, the information content of the vectors exceeds 285kbps). By doubling the size of the search domain, the amount of side information increases by approximately 100 kbps (in the case of the above example), while the computation requirement of motion estimation quadruples. This example reveals that also the motion vectors must be compressed in case of low bit rate coding, because the decisive amount of the bandwidth is occupied by the side information, while at very low data rates the transmission of even these becomes impossible. If longer bases were used for time domain transform, more frames would have to be made similar to each other, i.e. the number of motion vectors would multiply.

Fig. 2. Foreman and Coastguard series.

4.3 Compression of Motion Vectors

The motion vectors are generated in a structure similar to that of the individual frames; therefore the application of 3 D wavelet transform for the compression of these proves to be sensible. If the motion field is consistent, the motion vectors corresponding to the blocks that are close to each other have similar values (just like the pixels of the frames). The number of the motion vectors is much smaller than the pixel count; therefore the compression rate for the motion vectors is much worse. This ratio is further degraded by the requirement that the vectors must be transferred without losses. Haar base was selected for the wavelet transform of the motion vectors, and one vector was stored by less than 2 bits (depending on the contents of the image). The operation of the SPIHT algorithm is stopped when the last important bit has been coded, too. The bit stream is cut upon achieving a constant quality, not a constant transmission speed. The more movements occur in a video and the more dynamic it is, the less consistent the motion field will be, therefore more bits will have to be used for coding the motion vectors.

III. Experimental Results

Random access has also been taken into consideration during the implementation, which sets up requirements related to the GOF size. The frames of a GOF can simultaneously be coded or decoded, therefore, in case of random access the beginning of a GOF must definitely be waited for. If a GOF includes a huge number of frames, the access time increases proportionally with the frame count. The GOF size has therefore been chosen as 16, which corresponds to about 0.5 sec in case of 30 fps, yielding an expected random access time of 0.25 sec.

Haar base for the time domain transform. Symmetrical extension was applied at the edges. The size of the search domain of the motion compensation was 8x8 pixels, while the block size was varied between sensible limits. At the introduction of the results the quality is investigated in the function of the bit rate for different block sizes. Quality is characterized by the PSNR (Peak Signal to Noise Ratio) value, here given in dB. The horizontal axis shows the bit rate in kbps. Experiments were performed on various block sizes. The figures indicate that the coding resulted bad quality in case of 32 pixel blocks. The reason of this is that the chosen block size is too large for handling the fine movements; therefore motion compensation does not operate correctly. The individual movements can already be well described by blocks of 8 pixels, but considerably huge amount of motion vectors are generated in this case, requiring a relatively large transmission bandwidth. In case low bitrates this bandwidth is not available, therefore the motion vectors cannot be properly transferred, leading to the abrupt degradation of the picture quality. The application of blocks of 16 pixels provided optimal results with both series, because the movements can already be described with sufficient accuracy in this case, while the number of the generated motion vectors still remains acceptable. In the Coastguard sequence, most of the image is constituted by the background, while the two ships occur in a smaller area. The motion vectors of the background are mostly identical as a result of the movement of the camera, yielding a consistent motion field that can be stored by a relatively small amount of bits. There is little time and spatial redundancy on the waving surface of the water, therefore high bit rate is required to achieve good quality. Considering the Foreman sequence, also the man's head is moving in every direction besides the camera.

The results shown (Fig. 3. and Fig. 4.) were obtained by the Coastguard and Foreman test sequences (Fig. 4.) inCIF format (352x288 pixels/frame, 30fps). A Daubechies 9/3 [7] base was used for the spatial wavelet transform, while

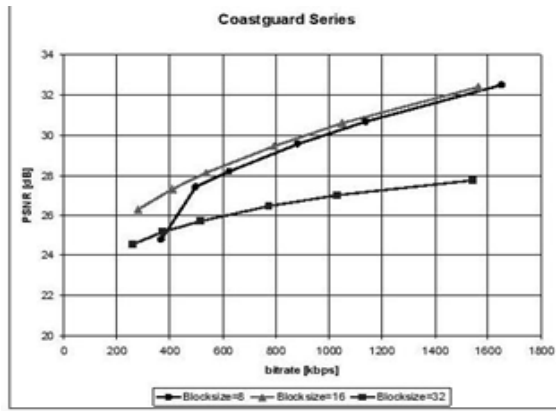


Fig.3 Quality vs. Bit Rate

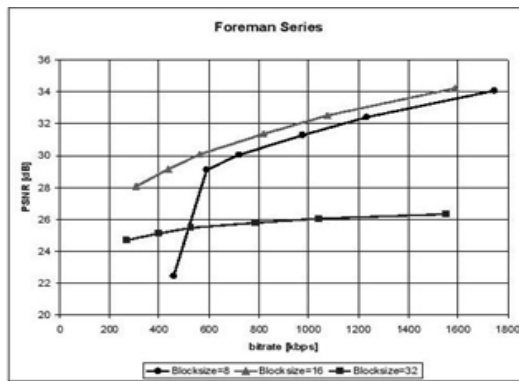


Fig.4 Quality vs. Bit Rate

IV. Conclusion

A new frame work for motion compensated highly scalable DWT is proposed. This new transform is amenable to any motion model and yields excellent preliminary results even with simple block-based motion. It also produces high quality low temporal resolution frames for temporal scalability.

Future work will be directed at implementing superior motion models, and optimizing rate allocation for the motion information.

V. References

- [1] B. J. Kim and W. Pearlman, "An embedded wavelet video coder using three-dimensional set partitioning in hierarchical trees (SPIHT)," *Proc. DCC'97, IEEE Data Compression Conference*, pp. 251-260, Mar. 1997.

- [2] J. Tham, S. Ranganath and A. Kassim, "Highly scalable wavelet-based video codec for very low bit-rate environment," *IEEE Journal on Selected Areas in Comm.*, vol. 16, pp. 12-27, Jan. 1998.
- [3] R. Calderbank, I. Daubechies, W. Sweldens and B. Yeo, "Wavelet transforms that map integers to integers," *Applied and Comp. Harm. Analysis*, vol. 5, pp. 332-369, July 1998.
- [4] Y. Nakaya and H. Harashima, "Motion compensation based on spatial transformations," *IEEE Trans. Circuits and Systems for Video Tech.*, vol 4, pp. 339-367, June 1994.
- [5] BOTTREAU, V., BENETICRE, M., FELTS, B., PESQUETPOPESCU, B. A fully scalable 3D subband video codec. *Image Processing*, vol.2, 2001, pp. 1017-1020.