REGION-BASED SPIHT CODING AND MULTIRESOLUTION DECODING OF IMAGE SEQUENCES

Sungdae Cho and William A. Pearlman
Center for Next Generation Video
Department of Electrical, Computer, and Systems Engineering
Rensselaer Polytechnic Institute
110 8th St. Troy, NY 12180-3590
chos@rpi.edu, pearlw@rpi.edu

ABSTRACT

This paper presents a region-based encoding/decoding for image sequences using Spatio-Temporal Tree Preserving 3-D SPIHT (STTP-SPIHT) algorithm, which is based on the 3-D SPIHT concepts, then further describes multiresolution decoding of the sequences. We have already proved the efficiency and robustness of the STTP-SPIHT in both of noisy and noiseless channels. This coder has also added benefit of parallelization of the compression and decompression algorithm, thus can be used in real-time implementation in hardware and software. In addition, we demonstrate that STTP-SPIHT has the functionality of region-based video encoding/decoding and spatial and/or temporal scalability. These features are highly desirable for today's multimedia applications, such as picture in picture function, video or volumetric image database browsing, distance learning, and video transmission over channels with highly constrained bandwidth.

1. INTRODUCTION

Wavelet zerotree image coding techniques were developed by Shapiro (EZW) [8], and further developed by Said and Pearlman (SPIHT) [7], and have provided unprecedented high performance in image compression with low complexity. Later, Kim *et al*. extended the idea to the three dimensional SPIHT (3-D SPIHT) algorithm [3, 4], and compared the result with the video compression algorithms which are generally used nowadays, such as MPEG-2, and H.263 even without any motion estimation/compensation. They showed promise of a very effective and computationally simple video coding, and also obtained excellent results in numerically and visually. We modified the 3-D SPIHT algorithm to work independently in a number of so-called spatio-temporal (s-t) blocks, composed of packets that are interleaved to deliver a fidelity embedded output bit stream. This algorithm is called STTP-SPIHT (Spatio-Temporal Tree Preserving 3-D SPIHT) [1]. Therefore a bit error in the bit stream belonging to any one block does not affect any other block, so that higher error resilience against channel bit errors is achieved.

In addition to the error resilience, the functionalities of region-based video encoding/decoding and spatial and/or temporal scalability are highly desirable to meet today's multimedia applications. Many classes of video sequences contain areas which are more important than others. It is unnecessary and unwise to equally treat all the pixels in image sequences. To minimize the total number of bits, unimportant areas should be highly compressed, thereby reducing transmission time and cost. One could preserve the features with nearly no loss, while achieving high compression overall by allowing degradation in the unimportant regions termed regionally lossy coding or region-based lossy coding. Compression schemes, of which are capable delivering higher reconstruction quality for the significant portions, are attractive in many cases including volumetric medical image areas, where doctors are interested only in a specific portion that might contain a disease.

In this paper, we first show how the STTP-SPIHT can be implemented to realize region-based encoding/decoding without changing any coding structures of the algorithm. This means this coder retains its error resilience to channel bit errors. This method extends easily to multiple ROI's. Then, we demonstrate this coder has spatial and/or temporal scalability.

The organization of this paper is as follows: Section 2 shows how region-based SPIHT coding can be implemented in STTP-SPIHT. Section 3 shows the multiresolution decoding of image sequences. Section 4 provides simulation results. Section 5 concludes this paper.

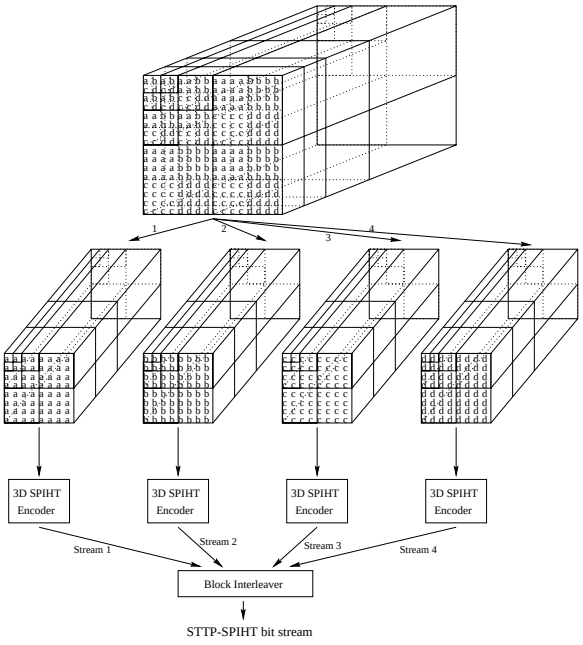

Figure 1. Structure of the Spatio-Temporal Tree Preserving 3-D SPIHT(STTP-SPIHT) compression algorithm

2. REGION-BASED SPIHT CODING

Most works in region-based coding have focused on goal-identifying regions with different gray scale characteristics which would benefit from the use of different encoding methods [2, 6, 5]. However, we can take advantage of the STTP-SPIHT coder of simple, robust, and very efficient embedded video coding method to get a region-based compression of image sequences. Using the coder, a specific region of interest (ROI) gets reproduced with higher quality than the rest of the image frame.

Figure 1 shows the structure and the basic idea of the STTP-SPIHT compression algorithm. STTP-SPIHT algorithm divides the 3-D wavelet coefficients into some number n of different groups according to their spatial and temporal relationships, and then to encode each group independently using the 3-D SPIHT algorithm, so that n independent embedded 3-D SPIHT substreams are created. These bitstreams are then interleaved in blocks. Therefore, the final STTP-SPIHT bitstream will be embedded or progressive in fidelity, but to a coarser degree than the normal SPIHT bitstream. In this figure, we show an example of separating the 3-D wavelet transform coefficients into four independent groups, denoted by a, b, c, d, each one of which retains the spatio-temporal tree structure of normal 3-D SPIHT[3, 4], and these trees correspond to specific regions of the image sequences. The s-t block, which is denoted by a, matches

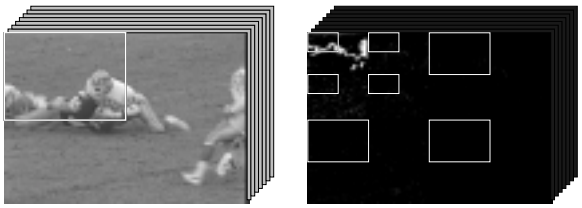

Figure 2. (a)Left : 352×240 "Football" sequence (frame 0) (b)right : Lowest subband frame with 2 levels of dyadic spatial decomposition

the top-left portion along with the image sequences, and the other s-t blocks correspond to the top-right, bottom-left, bottom-right fractions of the image sequences, and those s-t blocks are denoted by b, c, d, respectively. The normal 3-D SPIHT algorithm is just a case of n = 1, and we can flexibly choose n. When we choose n, the number of coefficients of x and y axis of sub-dimension should be divisible by 16 for 3 level decomposition, and divisible by 8 for 2 level decomposition. If the axis is not divisible by those numbers, we should extend the original image to be divisible. For 352×240 Football sequences, we can choose n from 1 up to 330.

As we can see, the STTP-SPIHT has a nice property that each substream retains its spatio-temporal relationship of wavelet coefficients. Furthermore each spatio-temporal related tree corresponds to a certain region of the image sequences. Therefore, we can assign more bits to the substream, which has the information of the region of interest, and assign the remaining bits of the bit budget to the other substreams to get region-based coding. Therefore, the sequence belonging to the background and the ROI's are coded independently at the specified bitrates. Figure 2 illustrates how region of interest in the image sequences is mapped into wavelet coefficient domain. In this figure, two level spatio-temporal transform using 9/7 wavelet filters is applied to the 352×240 "Football" image sequences. For any regions of interest, we can easily find out corresponding wavelet coefficients, since all the coefficients are composed of spatio-temporal orientation trees.

When we encode video sequences for region-based coding, we enter more information to the encoder, such as the axis of top-left and bottom right positions for the cross-section of the rectangular region of interest of the image, desired bit rate for the region of interest, and total bit rate or background bit rate, so that we can decide which bitstream would correspond to the region. There are two ways to decide the background bit rate. One, is to specify the total bit rate and desired bit rate of the ROI, and the other is to specify the ROI bit rate and background's. In the first case, we can always meet the target bit rate because we can assign

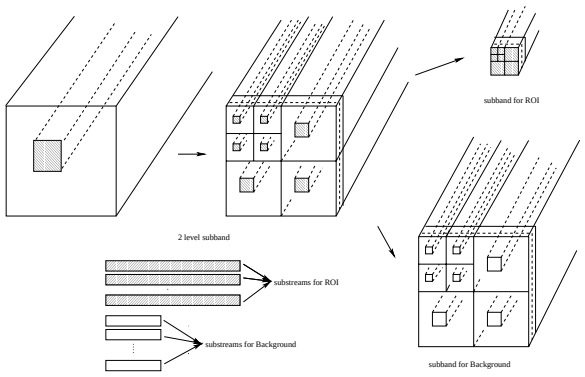

Figure 3. 3D Encoder System Configuration

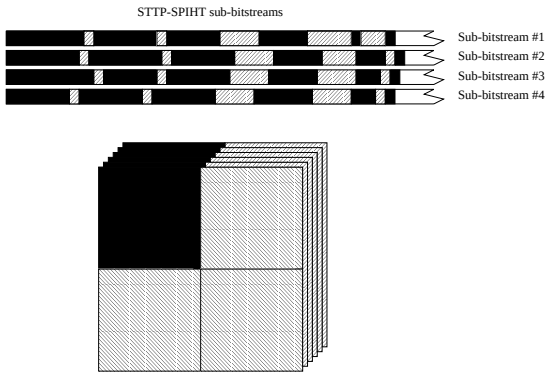

Figure 4. STTP-SPIHT sub-bitstreams and their corresponding spatial locations

the remaining bit budget of total bit rate after assigning to the ROI, and the second case, we can handle the quality of background image while maintaining the quality of the ROI.

To the decoder side, there are two kinds of substreams, one for the ROI and the other for the background. The STTP-SPIHT decodes the sub-bitstreams independently. The only difference from the original STTP-SPIHT is that there are two kinds of bit rates specified by the encoder. After their independent decodings, the decoded wavelet coeffieicnts are reordered according to their spatial and temporal relationships, and then the inverse wavelet transform is applied.

Figure 3 illustrates this idea. The ROI is shown as the shaded area, and white for the background. We can easily figure out which sub-block of coefficients corresponds to the ROI, because of the spatio-temporal relationships of each block. Then we assign more bits to the block which has the information of the ROI, and the remaining bits to the other blocks. The figure illustrates that the final substreams for the ROI are longer than the other substreams.

3. MULTIRESOLUTION DECODING OF IMAGE SEQUENCES

STTP-SPIHT is scalable in rate, using block interleaving/de-interleaving of the sub-bitstreams. In addition to that, it is highly needed for STTP-SPIHT to have temporal and/or spatial scalability for working with today's multimedia applications, such as video or volumetric image database browsing.

The STTP-SPIHT coder is based on a multiresolution wavelet decomposition, so it should be simple to add the function of multiresolution decoding to the STTP-SPIHT algorithm. Each STTP-SPIHT sub-bitstreams are composed of portions, and each of which contributes to the spatio/temporal locations. Therefore, we can just partition the embedded STTP-SPIHT sub-bitstreams into segments according to their subbands, and only decode the segment that corresponds to the desired resolution.

Figure 4 shows how the STTP-SPIHT sub-bitstreams (n = 4) are composed of portions according to their corresponding spatial/temporal locations. In this figure, dark areas represent the low resolution of video sequence, and the other part is used for high resolution. As we can see, lower resolution information is usually located at the beginning part of the sub-bitstreams. This is an example of 2 scales only, and we can easily extend to higher levels. Temporal scalability means frame rate scalability, and the STTP-SPIHT supports any combination of spatial and temporal scalability. After coding some point of the image sequences, most of the remaining bit budget is used for coding the higher frequency bands which contain the detail of the sequences, and the higher frequency areas are not usually visible at reduced spatial/temporal resolution.

The most valuable benefit of resolution scalable decoding is saving of decoding time, because the wavelet transformation consumes most of the decoding time of the process. For example, in a low resolution video of two spatial scales only, one-quarter of the number of wavelet coefficients is transformed, while the spatio-temporal resolution of two spatial and temporal scales involves transformation of one-eighth of the number of wavelet coefficients in the decoder. Therefore the lower resolutions of the sequences need much less time to transform due to the considerably fewer number of wavelet coefficients. This feature can be used in multimedia or volumetric medical image database browsing for faster search.

4. RESULT

Figure 5 shows an example of region-based coded image sequences. We specified the region of interest in command line as [198, 148, 245, 188], the bitrate for ROI as 1.0, and the bit rate for background as 0.05 for 'Football' sequence

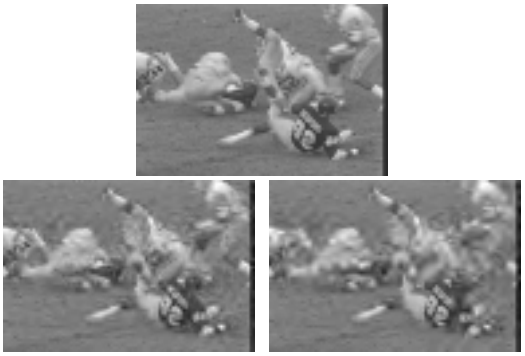

Figure 5. 352×240 "Football" sequence (frame15) (a)Top : Original sequence(b)Bottom-left : 3-D SPIHT compressed at 0.0845 bpp (c)Bottom-right : region-based STTP-SPIHT result, requiring an overall rate of 0.0845 bpp

with n = 55. The specified region of interest corresponds to the 38th, and 39th substreams. When we assign two bitrates to the substreams, the overall bitrate becomes 0.0845 bpp. In this figure, (a) is the 352×240 original Football sequence (frame 15) and (b) is the image decoded by original 3-D SPIHT with an overall bit rate of 0.0845 bpp, and the PSNR for the specified region of the image (frame 15) is 20.37 dB, and (c) is region-base coded image with the same overall bit rate of 0.0845 bpp, and the PSNR for the region of the image (frame 15) is 28.88 dB. In addition to the numerical result, in (b), the player's back number and name are hard to discern, but in (c), they are much clearer, but the background is not good as original.

Figure 6 illustrates the idea of multiresolution decoding for two levels of spatially scaled 352×240 "Football"sequence (frame 5). A low resolution video can be decoded from the first lower scale only, and the image size is 176×120. If we use three scales, the lowest scale's size would be 88×60 . Figure 7 is a typical example of multiresolution decoding of the "Football" and "Susie" sequences. In this figure, (a) and (b) are spatial half resolution of the sequence, and (c) and (d) are full resolution of the sequence.

5. Conclusion

We have implemented a region-based coding for image sequences and multiresolution decoding using STTP-SPIHT algorithm without changing any coding structure. Therefore, we can use all the features of the STTP-SPIHT algorithm. Furthermore, just little or no side information is needed by the encoder and decoder to implement the region-based coding and multiresolution decoding.

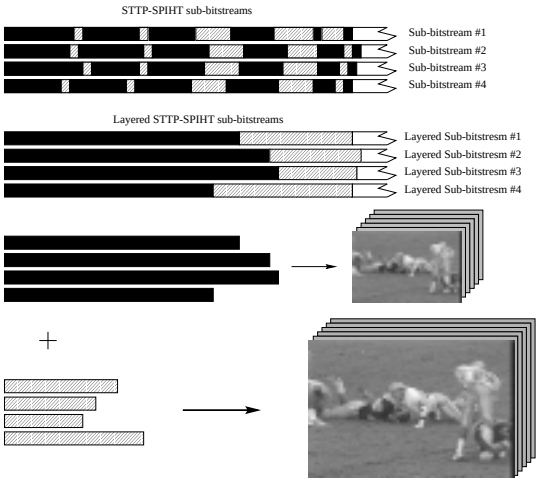

Figure 6. Multiresolution decoder uses the higher resolution layer to increase the spatial/temporal resolution of the video.

References

[1] S. Cho and W. A.Pearlman. Error resilient compression and transmission of scalable video. *Applications of Digital Image Processing XXIII, Proceedings SPIE,* 4115, 2000.

[2] A. Ikonomopoulos, M.Kunt, and M. Kocher. Second generation image coding techniques. *Proc. IEEE*, 73:549–574, 1985.

[3] B.-J. Kim and W. A. Pearlman. An embedded wavelet video coder using three-dimensional set partitioning in hierarchical trees. *Proc. of Data Compression Conference*, pages 251–260, 1997.

[4] B.-J. Kim, Z. Xiong, and W. A. Pearlman. Low bit-rate scalable video coding with 3d set partitioning in hierarchical trees (3d spiht). *IEEE Trans. Circuits and Systems for Video Technology*, 10:1374–1387,, December 2000.

[5] D.-K. Kim, Y.-D. C. Man-Bae Kim, and N.-K. Ha. On the compression of medical images with region of interest. *Proc. SPIE*, 2501:733–744, 1995.

[6] T. W. Ryan, L. D. Sanders, and H. D. Fisher. Wavelet-domain texture modeling for image compression,. *Proceedings of Visual Comm. Image Processing*, pages 380–383, 1994.

[7] A. Said and W. A. Pearlman. A new, fast and efficient image codec based on set partitioning in hierarchical trees. *IEEE Trans. on Circuits and Systems for Video Technology*, 6:243–250, June 1996.

[8] J. M. Shapiro. Embedded image coding using zerotrees of wavelet coefficient. *IEEE Transactions on Signal Processing*, 41:3445–3562, December 1993.

Figure 7. Multiresolution decoded sequence with STTP-SPIHT video coder (a)Top-left : spatial half resolution of “Football” sequence (frame 5) (b) Top-right : spatial half resolution of “Susie” sequence (frame 21) (c) Bottom Left : full resolution of “Football” sequence (frame 5) (d) Bottom right : full resolution of “Susie” sequence (frame 21).