Differential SPIHT for Image Sequence Coding

CIPR Technical Report TR-2010-1

Yang Hu and William A. Pearlman

IEEE International Conference on Acoustics, Speech, and Signal Processing
(ICASSP2010)
March 2010

Center for
Image Processing Research

Rensselaer Polytechnic Institute
Troy, New York 12180-3590
http://www.cipr.rpi.edu

DIFFERENTIAL-SPIHT FOR IMAGE SEQUENCE CODING

Yang Hu and William A. Pearlman

Department of Electrical, Computer, and Systems Engineering
Rensselaer Polytechnic Institute, Troy, NY 12180

ABSTRACT

Efficient image sequence coding exploits both spatial correlation and temporal correlation. SPIHT is a powerful algorithm in exploiting spatial correlations for still image coding. Based on SPIHT, we present a novel approach to exploit temporal correlations for image sequence coding, while maintaining single frame, random access decoding. The so-called Differential-SPIHT takes advantage of inter-frame correlations by reusing the significance testing data of a SPIHT coded frame. From simulation results of videos and volumetric medical images, Differential-SPIHT offers significant reductions in bitrate from conventional two-dimensional (2D) SPIHT for the same reconstruction PSNR, while retaining the desirables features of rate scalability and computational simplicity.

Index Terms— Compression, SPIHT, significance map, video, volumetric medical image.

1. INTRODUCTION

Image sequence and video coding achieves high coding efficiency by exploiting both spatial (intra-frame) correlation and temporal (inter-frame) correlation. To exploit temporal correlation, there are two main approaches: motion compensated predictive coding (MCPC) and three-dimensional (3D) wavelet transform. MCPC is the more computationally intensive and is widely used in the MPEGx and H.26x video coding standards. MCPC is applied to reduce the inter-frame dependence and the resultant frames are transformed by the Discrete Cosine Transform (DCT) and entropy coded[1]. In the 3D wavelet transform approaches, both 2D and 3D coding methods are used. Three-dimensional JPEG2000 coding can be realized with a 3D wavelet transform followed either by coding cubic subblocks [1] or by coding planar (x-y) subblocks in every z-axis transformed slice (or frame) [2]. SPIHT [3], SPECK [4], or related tree-based coding methods can also be performed on 2D or 3D hierarchical structures [5] [6] induced by the wavelet transform.

We gratefully acknowledge the support of the Office of Naval Research under Award no. N0014-05-10507.

[1]The first frame in a GOF (Group of Frames) acts as a reference and is intra-frame coded, either by entropy coding following a DCT and quantization or by 2D zerotree coding following a 2D wavelet transform.

In this paper, we present a new approach to exploit the inter-frame correlation. It is an extension of conventional 2D SPIHT to compress image sequences. The temporal correlation between frames is exploited by reusing the significance map of a previous SPIHT coded frame. Re-use of the significance map is an idea borrowed from the different scenario of distributed coding of hyperspectral images with SPIHT [7]. This method in our application gives excellent results in compressing volumetric medical images and video with low to moderate motion. One desirable feature of this method is that, unlike predictive and some other differential methods, random access to any frame or slice is assured, because each one is encoded and decoded individually. It also retains the SPIHT properties of a rate-scalable bit-stream and low computational complexity.

A brief description of SPIHT is given in section 2. In section 3, Differential-SPIHT is presented. Simulation results are given in section 4. Finally, section 5 presents our conclusions.

2. CONVENTIONAL 2D SPIHT

SPIHT [3] is a powerful algorithm for still image coding. It iteratively tests wavelet coefficients by comparing them with the threshold of each iteration. These tests are called significance tests. Let $C(i,j)$ be the wavelet coefficient of the encoded image at location (i,j). The thresholds are usually chosen as $T = 2^n$, starting from integer $n = n_{max} = \lfloor \log_2(\max_{(i,j)} |C(i,j)|) \rfloor$, the highest bitplane of the largest coefficient magnitude. Then n decreases by one at each of the following iterations.

The two-dimensional wavelet transform is viewed as a collection of non-overlapping trees rooted in coefficients of the lowest frequency subband. These trees are called *spatial orientation trees*, due to their branching to coefficients of like spatial orientation. Sets of coefficients are nodes in a subtree. When the magnitudes of all coefficients in a set are less than the current threshold, the set is deemed *insignificant* and a '0' is sent to the codestream; otherwise, the set is deemed *significant* and a '1' is sent to the codestream. When a set tests as significant, it is partitioned into four individual coefficients (tree offspring) and the set of all descendants of these offspring. Individual coefficients are similarly compared to

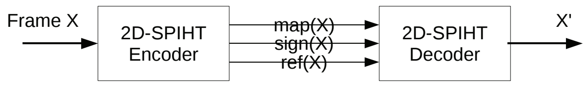

Fig. 1. Conventional 2D SPIHT.

the current threshold and signified as either significant or insignificant by sending the appropriate bit to the codestream. Initially, all coefficients at the roots and all sets are insignificant. When an individual coefficient tests as significant, its sign bit is also outputted and its coordinates are moved to an orderd list, called the LSP (List of Significant Points).

These tests proceed until all the sets and coefficients not significant at the operative threshold have been tested. This procedure is called *sorting pass*. The threshold is lowered by a factor of 2 for the next sorting pass through the insignificant sets and coefficients remaining from the previous pass.

The bit stream consisting of the outputs of significance tests from the sorting passes is called the *significance map*. This sequence of bits conveys the execution path of the encoder to the decoder. The sign bits constitute another class of bits from the sorting passes. A third class contains the refinement bits. When a sorting pass at threshold $T = 2^n$ (bit plane n) finishes, the bits in the n-th bitplane from LSP coefficients previously found significant at higher thresholds are outputted to the codestream. This procedure is called *refinement pass*, because it refines the values of the significant coefficients.

SPIHT coding contains a flexible number of iterations and each iteration includes a sorting pass and a refinement pass. The SPIHT codec is shown in Figure 1, where *map*(*), *sign*(*), and *ref*(*) indicate significance test bits (significance map), sign bits, and refinement bits, respectively. For medium bitrates, the significance map consumes about 75% to 85% of the output bitstream. As the bitrate increases, this percentage decreases, since the refinement and sign bits become more numerous.

3. DIFFERENTIAL-SPIHT

Differential-SPIHT is an extension of the well-known SPIHT algorithm to compress image sequences. The key idea of Differential-SPIHT is the reuse of the significance map.

The block diagram of the presented scheme is given in Figure 2. Let Frame X be a *reference-frame* and coded with conventional 2D SPIHT (Figure 1). Frame Y is called a *map-frame*, since it is coded using the significance map *map*(X) and sign data *sign*(X) of Frame X, as shown in Figure 3.

3.1. Correlation Analysis

As is known, the wavelet coefficients at the same position, (i, j), in neighboring frames, X and Y, are correlated. Four

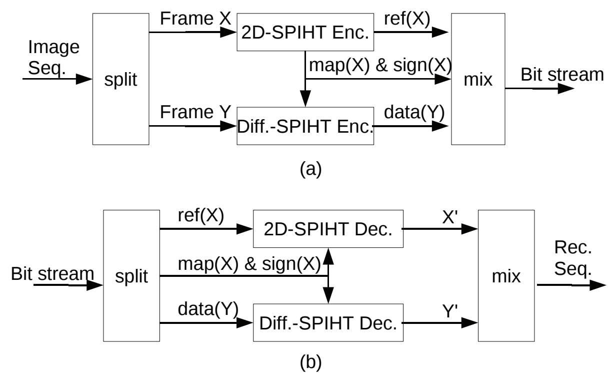

Fig. 2. New scheme (a) encoder (b) decoder.

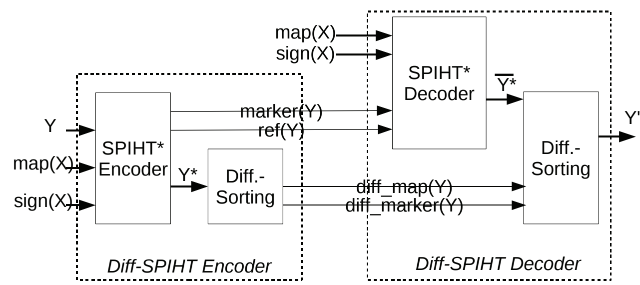

Fig. 3. Differential-SPIHT

relationships between the wavelet coefficients $C_X(i, j)$ and $C_Y(i, j)$ at location (i, j) in any iteration are important to us.

1. Both $C_X(i, j)$ and $C_Y(i, j)$ are significant.
2. Both $C_X(i, j)$ and $C_Y(i, j)$ are insignificant.
3. $C_X(i, j)$ is significant and $C_Y(i, j)$ is insignificant.
4. $C_X(i, j)$ is insignificant and $C_Y(i, j)$ is significant.

If X and Y are highly correlated, the first two relationships are much more likely to happen. The significance map, *map*(X), conveys the information that identifies the positions of the significant coefficients at each iteration. The coefficients of Y at these positions are likely to have the same state of significance in each corresponding iteration.

3.2. SPIHT*

In Differential-SPIHT (Figure 3) encoding, Frame Y is first performed with SPIHT* encoder. Different from conventional SPIHT, there is no significance test to generate a new significance map. Instead, SPIHT* reads in the significance map of X, *map*(X), to iteratively identify the significant positions for Y. At each iteration, the selected positions of Y are exactly the same as those of X. Most Y coefficients in these positions are also significant due to the high correlation between X and Y. However, there are other possible cases, because of the difference of X and Y.

To be concrete, assume (i, j) is a selected position at the iteration with threshold T. The coefficient $C_Y(i, j)$ involves sign and magnitude. For the sign, it may be the same or different from the sign of $C_X(i, j)$, which is available for SPIHT* from *sign*(X). For the magnitude, it may be insignificant ($|C_Y(i, j)| < T$), it may be *just* significant ($T \leq |C_Y(i, j)| < 2T$), or it may be *super* significant ($|C_Y(i, j)| \geq 2T$).

Instead of a sign bit output in the conventional 2D SPIHT, a marker code is outputted to indicate the specific case to which $C_Y(i, j)$ belongs. The markers are Huffman coded based on statistics gathered from simulations on sequences missa and salesman. These marker codewords are denoted by *marker*(Y) in Figure 3. Then the coefficient is updated to be $|C_Y(i, j)|$ or $(|C_Y(i, j)| - T \times m)$ for the insignificant and (just/super) significant cases, respectively, where $m = \lfloor |C_Y(i, j)|/T \rfloor$.

In the refinement pass of SPIHT*, given $T = 2^n$, the bits in the n-th bitplane of the coefficients previously found significant are outputted as *ref*(Y). Then, update these coefficients by setting the bits in the n-th bitplane to be 0's. Restricted by the input significance map, SPIHT* has the same number of iterations as SPIHT, which produces the map. The Y^* in Figure 3 denotes the frame with updated coefficients after SPIHT*.

3.3. Differential-Sorting

SPIHT* does not guarantee all the significant coefficients being coded. Assume the threshold in the last SPIHT* iteration is T_{last}. At some position (i, j), if $|C_Y(i, j)| \geq T_{last}$ but $|C_X(i, j)| < T_{last}$, the position would not be indicated by the significance map *map*(X), and, consequently, $C_Y(i, j)$ would not be coded in SPIHT*. With such coefficients uncoded, there may be a large PSNR loss. Thus Differential-Sorting (Figure 3) is adopted to make sure all the signfiicant coefficients in Y being coded.

In Differential-Sorting, the threshold is set to be T_{last}, and the sorting is enacted on the initial spatial-orientation tree. The significance tests are performed as in conventional SPIHT. The significance map bits convey the positions of significant coefficients of Y^* and are expressed as *diff_map*(Y). For each significant coefficient, instead of the sign bit output for conventional SPIHT, a marker code is outputted to indicate the sign and the magnitude. Different from *marker*(Y) in SPIHT*, these markers indicate the signs ($+$ or $-$) directly, and there is no such case that the magnitude is insignificant. Another Huffman code is designed for the markers according to statistics gathered from simulations. These codewords are indicated by *diff_marker*(Y) in Figure 3

In Differential-SPIHT decoding, Differential-Sorting follows SPIHT* decoding to ensure that all the significant coefficients will be reconstructed.

3.4. Performance Analysis

Given a specific T_{last}, Differential-SPIHT guarantees no PSNR loss, compared to 2D SPIHT, because it encodes and reconstructs all the coefficients with magnitude larger than or equal to T_{last}, just as 2D-SPIHT does. Combined with the reversible integer transform, Differential-SPIHT is able to compress losslessly, which is required by medical images.

Although Differential-SPIHT exploits inter-frame correlation, it has essential differences from the methods with MCPC, 3D-DWT, or DPCM (differential pulse code modulation). First, without interactions between frames, multiple map frames $(Y1, Y2, etc.)$ referring to the same reference frame (X) can be coded in parallel. These frames can be decoded independently, and there is no latency or error propagation. Secondly, it maintains low computational complexity; on the one hand, there is no special temporal-decorrelation process, unlike the computation-consuming MCPC and 3D-DWT; on the other hand, unlike MCPC and DPCM, the encoder need not execute decoding steps to reconstruct reference frames. Finally, it reduces the memory requirements by buffering only the significance map (of a much smaller size than a frame), instead of groups of frames.

4. SIMULATION RESULTS

As proof of concept, we limit our scope to a single map frame and provide comparison between the proposed scheme (Figure 2) and the conventional 2D SPIHT for videos and volumetric medical images. We use three levels of 2D wavelet transform and no entropy coding in these simulations.

The *2D-SPIHT* is simulated by coding every frame of a sequence with conventional 2D-SPIHT algorithm. The *New Scheme* is simulated by setting the image sequence to be "$XYXY$...", where X is a reference frame and Y is a map frame. That is, every odd-numbered frame in the sequence is coded as a reference frame by 2D SPIHT and the produced significance map is used by its following map frame.

In video simulations, the 9/7 biorthogonal wavelet filters [8] are used. Testing results for *salesman* (frame 1 to 58, 360×288) and *susie* (frame 1 to 58, 352×240) are given in Table 1 and Table 2, respectively, where $\overline{PSNR}(X, Y)$ is the average PSNR between map-frame and reference-frame. The larger the average PSNR, the higher the inter-frame correlation.

For volumetric medical images, the I(2+2,2) filters [9] are used to enable lossless compression. Table 3 provides the lossless coding results. Table 4 is lossy coding results for CT_skull (frame 1 to 60). All the volumetric medical images are of size 256×256 and 8-bit depth.

T_{last}	$PSNR_{rec}$ (dB)	2D-SPIHT (bpp)	New Scheme (bpp)
8	38.59	1.3116	1.1023
16	34.40	0.7419	0.5749
32	30.32	0.3769	0.2761
64	27.03	0.1870	0.1335
128	24.27	0.1036	0.0750

Table 1. Test results for salesman,$\overline{PSNR}(X,Y) = 33.11$ dB

T_{last}	$PSNR_{rec}$ (dB)	2D-SPIHT (bpp)	New Scheme (bpp)
8	40.17	0.8005	0.7528
16	36.32	0.4289	0.3723
32	32.95	0.2306	0.1818
64	30.31	0.1432	0.1054
128	27.07	0.0953	0.0691

Table 2. Test results for susie, $\overline{PSNR}(X,Y) = 29.86$ dB

Sequence	Frames	2D-SPIHT (bpp)	New Scheme (bpp)
CT_Aperts	1-60	1.4074	1.3436
CT_carotid	1-60	2.1648	2.0984
CT_skull	1-60	3.5348	3.3237
CT_wrist	1-60	2.3158	2.2255
MR_liver_t1	1-58	3.4460	3.3460
MR_ped_chest	1-58	3.2454	3.0600
MR_sag_head	1-58	3.2845	3.1427

Table 3. Lossless compression results.

T_{last}	$PSNR_{rec}$ (dB)	2D-SPIHT (bpp)	New Scheme (bpp)
2	46.19	2.4917	2.2568
4	42.34	1.5234	1.3151
8	37.27	0.8223	0.6730
16	31.52	0.4653	0.3619
32	24.70	0.2471	0.1842

Table 4. Lossy compression of CT_skull

Simulation results show that Differential-SPIHT outperforms conventional 2D SPIHT significantly, especially at low to moderate bitrates. The higher the inter-frame correlation, the better the performance of Differential-SPIHT. For lossless compression of volumetric medical images, the new scheme saves about 5% in bitrate. For video coding with moderate reconstructing PSNR (about 30 dB), the new scheme reduces the overall bit rates about 26%, and the average bitrate of Y frames is less than 50% of X frames.

5. CONCLUSION

Differential-SPIHT is a novel approach to exploit inter-frame correlation. It applies the significance map of a SPIHT coded frame to the following one or more frames to remove the inter-frame correlations. From the simulation results of videos and volumetric medical images, Differential-SPIHT outperforms conventional 2D SPIHT significantly, especially at low to moderate bitrates, and the performance increases with higher inter-frame correlation. At the same time, Differential-SPIHT maintains such features as low computational complexity, low memory requirement, supporting parallel coding, supporting lossless compression, no latency, and no PSNR loss. We consider Differential-SPIHT to be an attractive alternative for so-called *motion coding*, e.g., motion JPEG, motion JPEG2000, where frames of an image sequence are coded independently in order to preserve the desirable properties mentioned above.

6. REFERENCES

[1] J. Xu, Z. Xiong, S. Li, and Y.-Q. Zhang, "3-D embedded subband coding with optimal truncation (3-D ESCOT)," *J. Applied and Computational Harmonic Analysis*, vol. 10, no. 5, pp. 290–315, 2001.

[2] ISO/IEC Int. Standard 15444-2, Geneva, Switzerland, *Information Technology—JPEG2000 Extensions, Part 2: Core Coding System*, 2001.

[3] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning inhierarchical trees," *IEEE Trans. on Circuits & Syst. for Video Tech.*, vol. 6, no. 3, pp. 243–250, 1996.

[4] W.A. Pearlman, A. Islam, N. Nagaraj, and A. Said, "Efficient, low-complexity image coding with a set partitioning embedded block coder," in *IEEE Trans. Circuits & Syst. for Video Tech.*, 2004, vol. 14, pp. 1219–1235.

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

[6] Y. Liu and W.A. Pearlman, "Region of interest access with three-dimensional SBHP algorithm," in *Proc. SPIE*, 2006, vol. 6077, pp. 393–401.

[7] C. Tang, N.M. Cheung, A. Ortega, and CS Raghavendra, "Efficient inter-band prediction and wavelet based compression for hyperspectral imagery: a distributed source coding approach," in *Proc. IEEE Data Compression Conference*, 2005, pp. 437–446.

[8] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. on Image Processing*, vol. 1, no. 2, pp. 205–220, 1992.

[9] A. R. Calderbank, I. Daubechies, W. Sweldens, and B.-L. Yeo, "Wavelet transforms that map integers to integers," *J. Applied and Computational Harmonic Analysis*, vol. 5, pp. 332–369, 1998.