

# Motion Differential SPIHT for Image Sequence Coding

**CIPR Technical Report TR-2010-2**

Yang Hu and William A. Pearlman

September 2010



Center for  
Image Processing Research

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

# MOTION DIFFERENTIAL SPIHT FOR IMAGE SEQUENCE CODING

Yang Hu, William A. Pearlman

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

## ABSTRACT

Efficient image sequence coding exploits both intra- and inter-frame correlations. SPIHT is efficient in intra-frame decorrelation for still images. Based on SPIHT, differential-SPIHT removes inter-frame redundancy by reusing the significance map of a SPIHT coded frame. The motion differential SPIHT (MD-SPIHT) automatically decides the coding methods for each frame, according to the inter-frame correlations in terms of PSNR. From the simulation results of hyperspectral images, 3D/4D medical images, and video, MD-SPIHT is more efficient, compared with conventional 2D SPIHT and motion JPEG2000, while retaining the desirable features of rate scalability, random frame access, and computational simplicity.

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

## 1. INTRODUCTION

Image sequence coding achieves high efficiency by exploiting both spatial (intra-frame) and temporal (inter-frame) correlations. To exploit temporal correlation, there are two main approaches: inter-frame prediction and three-dimensional (3D) wavelet transform. Inter-frame prediction is the more computationally intensive and is widely used in the MPEGx and H.26x video coding standards. In the 3D wavelet transform approaches, both 2D and 3D zerotree coding methods can be used. For example, 3D 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 transformed frame [2].

In this paper, we present an image sequence coding system called *motion differential SPIHT* (MD-SPIHT), employing the inter-frame decorrelation approach presented in [3]. In this system a frame is coded either by conventional 2D SPIHT or by differential-SPIHT, where inter-frame correlation is exploited by reusing the significance map of a previously SPIHT coded frame. This system gives excellent results in compressing hyperspectral images, 3D and 4D medical images, and video with low to moderate motion. One desirable feature of this method is that, unlike 3D wavelet transform or

predictive and some other differential methods, random access to any frame 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, MD-SPIHT is presented. Simulation results are given in section 4, and section 5 is the conclusion.

## 2. CONVENTIONAL 2D SPIHT

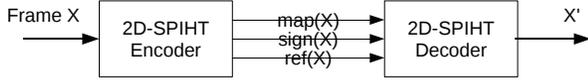
SPIHT [4] is powerful for still image coding. It iteratively tests wavelet coefficients by comparing them with the threshold of each iteration, called significance tests. Let  $C(i, j)$  be the wavelet coefficient at location  $(i, j)$ . The thresholds are usually chosen as  $T = 2^n$ , starting from integer  $n = n_{max} = \lfloor \log_2(\max_{(i,j)} |C(i, j)|) \rfloor$ , and decreases by one at each following iteration. There is a flexible number of iterations and each iteration includes a sorting pass and a refinement pass.

The 2D wavelet transformed structure is viewed as a collection of non-overlapping trees rooted in coefficients of the lowest frequency subband. When the magnitudes of all coefficients in a set (a tree or sub-tree) 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. 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 compared to the current threshold and signified as either significant or insignificant by sending the appropriate bit to the codestream. Initially, all coefficients and all sets rooted at the lowest subband are insignificant. When an individual coefficient tests as significant, its sign bit is outputted and its coordinates are moved to an ordered list, called the LSP (List of Significant Points). These tests proceed until all the insignificant sets and coefficients have been tested at the operative threshold. This procedure is called *sorting pass*.

After a sorting pass at threshold  $T = 2^n$  (bit plane  $n$ ), the bits in the  $n$ -th bitplane of LSP coefficients previously found significant at higher thresholds are outputted to the codestream. This procedure is called the *refinement pass*. Then the threshold is lowered by a factor of 2 for the next iteration.

The output of significance tests from the sorting passes is

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



**Fig. 1.** Conventional 2D SPIHT.

called the *significance map*, which 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 from the refinement passes. The SPIHT codec is shown in Figure 1, where  $map(*)$ ,  $sign(*)$ , and  $ref(*)$  indicate significance map, sign bits, and refinement bits, respectively. For medium bit rates, the significance map consumes about 75% to 85% of the output bitstream. As the bitrate increases, this percentage decreases, since the refinement bits belonging to a bit plane are coded for every coefficient that is significant for previous higher bitplanes.

### 3. MOTION DIFFERENTIAL SPIHT

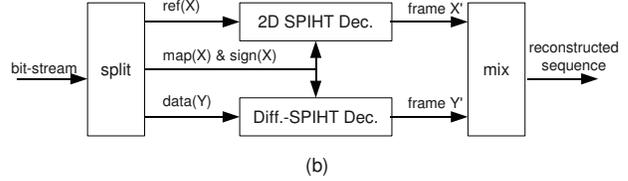
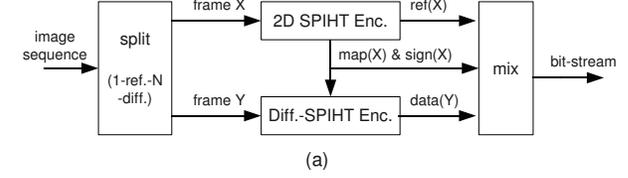
The MD-SPIHT has a similar framework with other so-called *motion-coding* methods, such as motion JPEG and motion JPEG2000, in that there are no inter-frame interactions. In MD-SPIHT, the frames of an image sequence are compressed either as *reference-frames* or as *differential-frames*. The former are coded with the conventional 2D SPIHT. The latter are coded with differential-SPIHT, which utilizes the significance map and the sign data of a previously coded reference-frame to predict the locations and signs of the significant coefficients of the current frame.

After each reference-frame, there could be  $N$  (0, 1, 2, ...) differential-frames. In this "1-reference- $N$ -differential" frame structure,  $N$  can be fixed to some value for the whole image sequence, or determined adaptively during the coding procedure. To be concrete, a correlation metric is defined as the PSNR between the reference-frame and the current frame. If this PSNR is greater than a threshold  $T_{PSNR}$ , the current frame will be coded as a differential-frame; otherwise, the current frame will be coded with 2D SPIHT and serves as a new reference-frame, until a newer reference-frame appears. The system is illustrated in Figure 2, where  $X$  and  $Y$  represent the reference- and differential-frames, respectively.

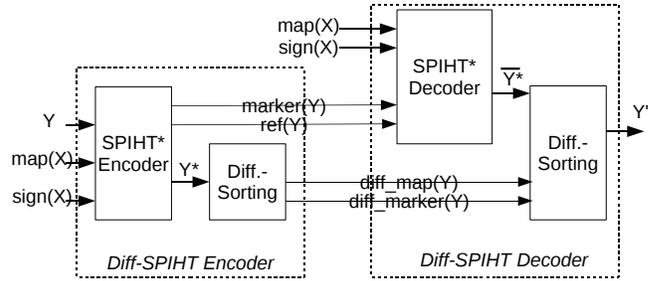
There are four relationships between the wavelet coefficients  $C_X(i, j)$  and  $C_Y(i, j)$  at location  $(i, j)$  of frames  $X$  and  $Y$  in any iteration.

- I** Both  $C_X(i, j)$  and  $C_Y(i, j)$  are significant.
- II** Both  $C_X(i, j)$  and  $C_Y(i, j)$  are insignificant.
- III**  $C_X(i, j)$  is significant and  $C_Y(i, j)$  is insignificant.
- IV**  $C_X(i, j)$  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. Let the threshold in the last iteration be  $T_{last}$ . To ensure no loss of reconstruction



**Fig. 2.** Motion Differential SPIHT (a) encoder (b) decoder.



**Fig. 3.** Differential-SPIHT

fidelity compared to conventional SPIHT coding, all significant coefficients with respect to  $T_{last}$  in frame  $Y$  should be coded. Such coefficients are covered by Relations I and IV. In differential-SPIHT, Figure 3, the *SPIHT\** and *differential-sorting* stages deal with these two cases, respectively.

#### 3.1. SPIHT\*

The *SPIHT\** is the first step in differential-SPIHT. It codes the significant coefficients at the locations indicated by the significance map  $map(X)$ . The *SPIHT\** reads in  $map(X)$  to iteratively identify the possibly significant positions for  $Y$ . At each iteration, the  $Y$  coefficients in the selected positions are significant with a good probability, due to the correlation between  $X$  and  $Y$ . However, there are other possible cases, because of the difference of  $X$  and  $Y$ .

The coefficient  $C_Y(i, j)$  involves the sign and the magnitude, where  $(i, j)$  is the selected location by  $map(X)$  at the iteration. For the sign of  $C_Y(i, j)$ , it may be the same as or different from the sign of  $C_X(i, j)$ , which is available for *SPIHT\** from  $sign(X)$ . For the magnitude of  $C_Y(i, j)$ , it may be insignificant ( $|C_Y(i, j)| < T$ ), just significant ( $T \leq |C_Y(i, j)| < 2T$ ), or super significant ( $|C_Y(i, j)| \geq 2T$ ). Instead of a sign bit output in the con-

ventional 2D SPIHT, a marker symbol is sent to indicate the specific case to which  $C_Y(i, j)$  belongs. The marker symbols are arithmetic coded [5]. These marker data are denoted by  $marker(Y)$ . 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  $m = \lfloor |C_Y(i, j)|/T \rfloor$ .

The refinement pass of SPIHT\* is similar to that of conventional SPIHT. Given  $T = 2^n$ , the bits in bit-plane  $n$  of the coefficients that are previously found significant are outputted as  $ref(Y)$ . Then, update these coefficients by setting the bits in bit-plane  $n$  to be 0's. Restricted by the input significance map, the SPIHT\* has the same number of iterations as the SPIHT procedure, which produces the map. The  $Y^*$  in Figure 3 denotes the frame with updated coefficients after all SPIHT\* iterations. So far, all the significant coefficients in Relation I are coded.

### 3.2. Differential-Sorting

Differential-sorting is adopted to code the  $Y$  coefficients in Relation IV. In differential-sorting, the threshold is set to be  $T_{last}$ , and the sorting is enacted on the initial trees. The significance tests are performed as in conventional SPIHT, and produce its significance map  $diff\_map(Y)$ , which locates the significant coefficients of  $Y^*$ .

For each significant coefficient, a marker  $diff\_marker(Y)$  is sent to indicate the sign and the magnitude. Unlike  $marker(Y)$  in SPIHT\*, these markers indicate the signs (+, -) directly, and there are no insignificant magnitudes. Similarly, the markers are arithmetic coded.

In Differential-SPIHT *decoding*, differential-sorting follows SPIHT\* decoding to make sure all the significant coefficients are being reconstructed.

### 3.3. Performance Analysis

Given a  $T_{last}$ , MD-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 reversible integer transforms, MD-SPIHT is able to compress losslessly, which is required by some medical image applications.

Although MD-SPIHT exploits inter-frame correlation, it has essential differences from the 3D DWT or predictive or other differential methods. First, without inter-frame interactions, multiple differential-frames can be coded in parallel, random frame access is supported, and there is no latency. Secondly, it maintains low computational complexity, because there is no special temporal decorrelation process, unlike the computation-consuming inter-frame prediction and 3D DWT. Furthermore, unlike predictive and other differential methods, the encoder need not simulate decoding to create reference frames. Finally, it reduces the memory requirements by buffering only the reference frame and its significance map and sign data, instead of groups of frames.

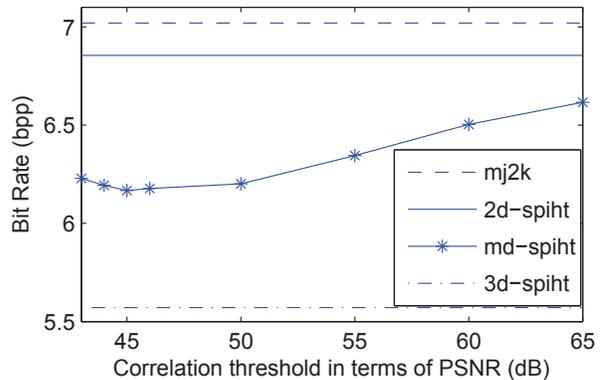


Fig. 4. Lossless bit rates at different correlation thresholds

Sequence	SPIHT			MJ2K
	MD	2D	3D	
cuprite sc04	6.27	6.93	5.63	7.10
moffet field sc01	6.16	6.85	5.71	7.57
moffet field sc03	6.27	6.93	5.08	6.29
jasper ridge sc01	7.17	7.70	5.94	7.89
jasper ridge sc03	7.15	7.68	5.92	7.86

Table 1. Hyperspectral image lossless bit rates (bpp)

## 4. SIMULATION RESULTS

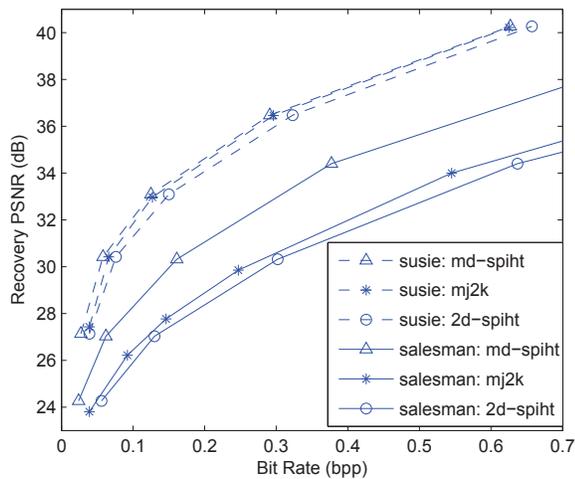
MD-SPIHT is simulated to compare with other coding schemes having the frame-by-frame coding structure. One is the conventional 2D SPIHT, which is applied consecutively to every frame of the image sequence. The other is motion JPEG2000 (MJ2K), which codes each frame as a JPEG2000 image. The MD-SPIHT uses the adaptive "1-reference-N-differential" frame structure. In simulation, all schemes adopt 3-level 2D DWT and arithmetic coding. As in the JPEG2000, the 5/3 [6] and 9/7 [7] filters are adopted for lossless and lossy coding, respectively. The comparisons are in terms of the compression efficiency in bpp (bits per pixel).

For the AVIRIS hyperspectral images ( $512 \times 512 \times 224$ , 16-bit depth), Figure 4 presents the lossless compressed bit rates at different  $T_{PSNR}$  for *cuprite* scene 01. We can see that the MD-SPIHT is most efficient at  $T_{PSNR} = 45$ dB, which is used to compress the other hyperspectral images and the results are given in Table 1, where the 3D SPIHT [8] results are also provided as a reference.

Medical images are coded losslessly. In 3D volume images, each 2D slice is seen as a frame, and MD-SPIHT exploits the dependency between slices. While 4D images are time series of 3D volumes, MD-SPIHT exploits the inter-volume (temporal) dependency. Simulation results are provided in Table 2, where MD-SPIHT has specific  $T_{PSNR}$  for each subject. All the tested medical images are of 8-bit depth.

Sequence	SPIHT			MJ2K
	MD ( $T_{PSNR}$ )	2D	3D	
3D medical images:				
CT_aperts	1.16 ( 33 dB )	1.28	0.92	1.22
CT_carotid	1.88 ( 33 dB )	2.04	1.56	1.96
CT_skull	3.01 ( 27 dB )	3.37	2.34	3.32
CT_wrist	1.97 ( 35 dB )	2.21	1.48	2.07
MR_liver_t1	3.05 ( 33 dB )	3.32	2.34	3.25
MR_ped_chest	2.63 ( 40 dB )	3.05	1.93	3.00
MR_sag_head	2.65 ( 31 dB )	2.88	2.19	2.91
4D medical images:				
volunteer4d	3.28 (39 dB)	3.39		3.35
ct4d	3.17 (41 dB)	3.33		3.24
siem	4.81 (37 dB)	4.81		5.13

**Table 2.** Medical image lossless bit rates (bpp)



**Fig. 5.** Video (susie, salesman) lossy compression

The tested 3D CT volumes are slices 1-64, and 3D MR volumes are slices 1-48, all of size  $512 \times 512$ . The tested 4D images are volunteer4d ( $128 \times 120$ , slices 1-16, volumes 1-16), ct4d ( $256 \times 256$ , slices 1-192, volumes 1-18), and siem ( $64 \times 64$ , slices 1-16, volumes 1-112).

Lossy compression is performed for video. The testing results for *susie* ( $352 \times 240$ , frame 1-64) and *salesman* ( $360 \times 288$ , frame 1-64) are given in Figure 5 with  $T_{PSNR} = 27$ dB, which is achieved from the simulation of *foreman* ( $352 \times 288$ , 300 frames). MD-SPIHT has much better performance for *salesman*, because *salesman* has higher inter-frame correlation, due to lower scene motion and lack of camera motion.

From the simulation results, MD-SPIHT outperforms the other two 2-D schemes significantly, especially for sequences with high inter-frame correlation, e.g., hyperspectral images, 3D medical volumes, and video with low to moderate motion.

For *salesman* at 30 dB recovering PSNR, MD-SPIHT reduces the overall bit rate by about 40% of those of the other two schemes.

## 5. CONCLUSION

MD-SPIHT is an effective system for image sequence compression. 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 hyperspectral images, 3D/4D medical images, and video, MD-SPIHT outperforms 2D SPIHT and MJ2K significantly, especially for sequences with high inter-frame correlation. At the same time, MD-SPIHT maintains such features as low computational complexity, low memory requirement, support of parallel coding, support of lossless compression, no latency, and no fidelity loss. We consider MD-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 individually 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] Y. Hu and W.A. Pearlman, "Differential-SPIHT for image sequence coding," in *Proc. ICASSP*, 2010.
- [4] 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.
- [5] I.H. Witten, R.M. Neal, and J.G. Cleary, "Arithmetic coding for data compression," *Communications of the ACM*, vol. 30, no. 6, pp. 540, 1987.
- [6] AR Calderbank, I. Daubechies, W. Sweldens, and B.L. Yeo, "Lossless image compression using integer to integer wavelet transforms," in *International Conference on Image Processing*, 1997, pp. 596–599.
- [7] 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.
- [8] S. Cho and W.A. Pearlman, "Error-resilient video coding with improved 3D SPIHT and error concealment," in *Proceedings of SPIE*, 2003, vol. 5022, pp. 125–136.