

# Error-Resilient and Error Concealment 3-D SPIHT Video Coding with Added Redundancy\*

Jie Zhu, R.M. Dansereau, and Aysegül Cuhadar

Department of Systems and Computer Engineering,  
Carleton University, Ottawa, Ontario, Canada  
{jzhu, rdanse}@sce.carleton.ca

**Abstract.** In this paper, we propose a multiple description (MD) video coding algorithm based on error-resilient and error concealment set partitioning in hierarchical trees (ERC-SPIHT). Compared to ERC-SPIHT, the novelty of the proposed approach is the injection of additional redundancy into the substreams so that the coefficients in the spatial root subband are protected highly from transmission errors. Experimental results on different video sequences show that the proposed method maintains error-resilience with high coding efficiency. In particular, our results demonstrate that the proposed algorithm achieves a significant improvement on video quality by up to 2.24 dB in the presence of a substream loss compared to ERC-SPIHT.

**Keywords:** multiple description (MD) coding, set partitioning in hierarchical trees (SPIHT), error-resilience, error concealment.

## 1 Introduction

The set partitioning in hierarchical trees (SPIHT) algorithm has made great strides in image and video coding [1]. It provides outstanding rate-distortion (R-D) performance with low computational complexity. However, SPIHT-coded bitstreams are extremely sensitive to transmission errors. To remedy this problem, it is important to incorporate error-resilience with the SPIHT encoding algorithm for robust video transmission over error-prone networks.

There have been various error control mechanisms proposed to provide error-resilience for video coders [2]. One common approach is to request retransmission of any corrupted information from the receiver. However, retransmission may not be acceptable for real-time video applications due to long round-trip delays. Another popular error-resilience method is forward error correction (FEC), which is designed to add additional channel redundancy to the bitstream. However, FEC is ineffective in packet-based networks when a large number of packets may be lost in clusters and such loss exceeds the recovery capability of FEC codes [3].

Multiple description (MD) video coding has recently attracted attention as a promising approach for error-resilient video transmission [4]. Among existing MD

---

\* This research was partially funded by Nortel and the Natural Sciences and Engineering Research Council (NSERC) of Canada.

video coding schemes, several domain-partitioning based algorithms using three-dimensional (3-D) SPIHT have shown to be effective in achieving error resilience, such as spatial and temporal tree preserving 3-D SPIHT (STTP-SPIHT) [5] and error resilient and error concealment 3-D SPIHT (ERC-SPIHT) [6]. In these algorithms, the wavelet coefficients are divided into multiple groups and each group is independently encoded with the 3-D SPIHT algorithm and then separately transmitted over networks. Error-resilience of a domain-partitioning based MD video coder can be further ameliorated when the coder is supplemented with an error concealment technique at the decoder. It has been demonstrated that even simple error concealment, such as bilinear interpolation (missing coefficients are estimated by the average value of its four adjacent neighbors), can improve the R-D performance when combined with STTP-SPIHT and ERC-SPIHT [7]. More specifically, ERC-SPIHT provides better error concealment performance over STTP-SPIHT because ERC-SPIHT uses dispersive grouping such that the spatially adjacent coefficients are spread onto different substreams [7].

Although error concealment can be used in conjunction with STTP-SPIHT and ERC-SPIHT, it may still cause reduced presentation quality. Small relative errors in the estimation of missing coefficients in the spatial root subband (i.e., the approximation subband on the highest spatial decomposition level) can have a significant impact in the overall distortion. An elegant solution for a video coder to reduce such impact is to inject additional redundancy into the substreams so that the missing coefficients in the spatial root subband can be recovered with a high level of accuracy. By additional redundancy, we mean redundancy that is added intentionally to the signal instead of the inherent redundancy present in the signal. The key challenge in developing such a video coder is how to balance the tradeoff between the source coding efficiency at the encoder and the error concealment performance at the decoder.

In this paper, we propose an error-resilient MD video coding framework for ERC-SPIHT-coded video sent across error-prone networks. The framework is an improvement of our earlier related scheme [8]. Although the work in [8] allows an important quality improvement, it is restricted to a specific scenario (i.e., only four substreams are created and only one of the substreams allowed lost due to noisy channels). In this paper, we generalize the scheme for the generation of an arbitrary number  $S$  of substreams and extend the recovery of missing coefficients.

The rest of this paper is organized as follows. In Section 2, we present our error-resilient video coding framework. We discuss the simulation results in Section 3 and draw conclusions in Section 4.

## 2 Proposed Multiple Description Video Coding Algorithm With 3-D SPIHT

Motivated by the approach in [8], now we propose a new error-resilient MD video coding scheme. The overall framework of the proposed coding algorithm is shown in Fig. 1, where two new functions are added to the ERC-SPIHT video coding system [7], namely additional redundancy generation at the encoder and

root subband recovery at the decoder. In our implementation, we use the same order of spatial and temporal transformation as in [7]. In addition, we adopt the 3-D SPIHT with asymmetric tree structure (AT-SPIHT) in our work since it exhibits better R-D performance than its counterpart, the 3-D SPIHT with symmetric tree structure [7].

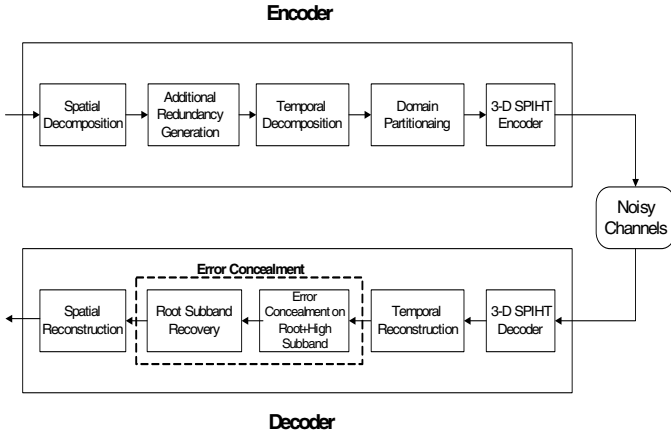


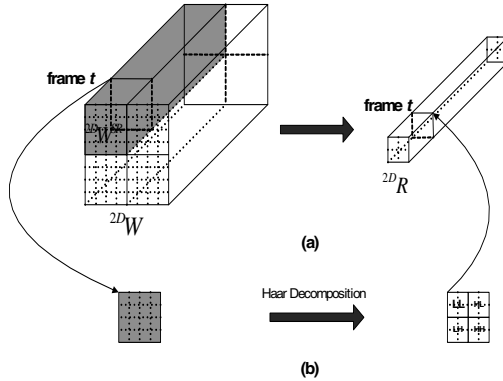
Fig. 1. Framework of the proposed video coding algorithm

### 2.1 Additional Redundancy Generation and Encoding

After a group of frames (GOF) is spatially transformed into the 2-D wavelet transform domain, the next step is to generate the additional redundancy. In essence, the additional redundancy generation operates as illustrated in Fig. 2(a). To simplify the illustration, we assume there are  $N$  frames in the GOF and the frames are spatially transformed in one level. Symbol  ${}^{2D}W$  denotes the wavelet coefficients resulting from the spatial transform and symbol  ${}^{2D}W^{SR}$ , shown by the shaded area, represents the spatial root subband. The additional redundancy is obtained by spatially decomposing each frame of the spatial root subband  ${}^{2D}W^{SR}$  by one more level. Then, the resulting coefficients in the LL (low-low) subband (i.e., approximation coefficients in the lowest frequency subband) are grouped together as the additional redundancy, denoted by  ${}^{2D}R$ . Note that the wavelet transform chosen at this step can be different from the one used for decomposing the GOF into the 3-D wavelet domain. In this work, we employ the Haar wavelet transform. One reason to choose the Haar wavelet is that the computation is much simpler. Another advantage of the Haar wavelet is that it is a 2-tap filter. As a result, each element in  ${}^{2D}R$  is only correlated to a block of  $2 \times 2$  adjacent wavelet coefficients in  ${}^{2D}W^{SR}$ . Recall that the coefficients of the Haar low-pass filter are  $\left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right]$  and the coefficients of the high-pass filter are  $\left[\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right]$ . Therefore,

let  ${}^{2D}R(x, y, t)$  and  ${}^{2D}W(x, y, t)$  denote the coefficient at the spatial location  $(x, y)$  of frame  $t$  in  ${}^{2D}R$  and  ${}^{2D}W$ , respectively. We find that

$${}^{2D}R(x, y, t) = \{ {}^{2D}W(2x - 1, 2y - 1, t) + {}^{2D}W(2x - 1, 2y, t) + {}^{2D}W(2x, 2y - 1, t) + {}^{2D}W(2x, 2y, t) \} / 2. \tag{1}$$



**Fig. 2.** Relationship between original video sequence and additional redundancy

Next,  ${}^{2D}W$  and  ${}^{2D}R$  are both temporally transformed to the 3-D wavelet domain and the resulting coefficients are denoted by  ${}^{3D}W$  and  ${}^{3D}R$ , respectively. The above relationship is still valid for  ${}^{3D}W$  and  ${}^{3D}R$  such that

$${}^{3D}R(x, y, t) = \{ {}^{3D}W(2x - 1, 2y - 1, t) + {}^{3D}W(2x - 1, 2y, t) + {}^{3D}W(2x, 2y - 1, t) + {}^{3D}W(2x, 2y, t) \} / 2, \tag{2}$$

where  ${}^{3D}R(x, y, t)$  and  ${}^{3D}W(x, y, t)$  represent the coefficient at the spatial location  $(x, y)$  of frame  $t$  in  ${}^{3D}R$  and  ${}^{3D}W$ , respectively. This relation holds true since the Haar wavelet transform is a separable transform.

The issue is now how to encode  ${}^{3D}R$  efficiently and resiliently. An effective solution is to distribute the additional redundancy  ${}^{3D}R$  over different substreams. We assume that the coefficients in  ${}^{3D}W$  are divided into  $S$  groups by dispersive grouping. Similarly, we use dispersive grouping to organize the coefficients in  ${}^{3D}R$  into  $S$  groups. Groups  ${}^{3D}W_i$  and  ${}^{3D}R_i$  are then independently encoded with AT-SPIHT to produce substream- $i$ , where  $i = 1, \dots, S$ .

Recall that in our previous work [8], the additional redundancy  ${}^{3D}R$  is comprised of the average value of four coefficients with the same location in the spatial root subband of groups  ${}^{3D}W_1$ ,  ${}^{3D}W_2$ ,  ${}^{3D}W_3$  and  ${}^{3D}W_4$ . When dispersive grouping is applied, we had

$${}^{3D}R(x, y, t) = \{ {}^{3D}W(2x - 1, 2y - 1, t) + {}^{3D}W(2x - 1, 2y, t) + {}^{3D}W(2x, 2y - 1, t) + {}^{3D}W(2x, 2y, t) \} / 4. \tag{3}$$

Compared to the proposed video coder, a smaller amount of additional redundancy is inserted among substreams in the previous work [8] since each element in  ${}^3D R$  is only half the magnitude of the one in (2).

### 2.2 Reconstruction of Lost Wavelets Coefficients

We now describe how the proposed root subband recovery algorithm can be used with other 2-D error concealment schemes in each frame. We define  ${}^3D W'$  and  ${}^3D R'$  as the coefficients decoded from the received substreams. We assume that when the error occurred in any substream induced by noisy channels, all coefficients in this substream are set to zero. Then, the reverse temporal transform is applied on  ${}^3D W'$  and  ${}^3D R'$  separately and the resulting frames are denoted by  ${}^2D W'$  and  ${}^2D R'$ .

The issue now is how to reconstruct the missing coefficients in  ${}^2D W'$ . In this work, we focus on the reconstruction of the missing coefficients in the spatial root subband. When the additional redundancy is generated, each element in  ${}^2D R$  is only correlated to a block of  $2 \times 2$  adjacent wavelet coefficients in  ${}^2D W$ . As a result, root subband recovery is implemented in each block of size  $2 \times 2$  adjacent wavelet coefficients. For notation convenience, let us denote the coefficients in the same block  ${}^2D W'(2x - 1, 2y - 1, t)$ ,  ${}^2D W'(2x - 1, 2y, t)$ ,  ${}^2D W'(2x, 2y - 1, t)$  and  ${}^2D W'(2x, 2y, t)$  as  $W'_1$ ,  $W'_2$ ,  $W'_3$  and  $W'_4$ . In this paper, we consider two cases as below.

- Case 1: only one coefficient is missing

It is straightforward to reconstruct the lost coefficient by (2). For example, if  $W'_1$  is missing, it can be reconstructed by

$$W_1^* = 2 \cdot {}^2D R'(x, y, t) - W'_2 - W'_3 - W'_4. \tag{4}$$

- Case 2: two coefficients are missing

Note that (4) is not suitable when there is more than one coefficient lost. A solution for this problem is to recalculate the missing coefficients by exploiting the Haar wavelet decomposition and reconstruction. Let's assume that two coefficients,  $W'_2$  and  $W'_4$ , in the same block are affected by the noisy channels. Initially, the lost coefficients  $W'_2$  and  $W'_4$  are estimated with any existing 2-D error concealment technique. For example,  $W'_2$  and  $W'_4$  are reconstructed as  $\hat{W}_2$  and  $\hat{W}_4$ . This block is then spatially decomposed by one level by applying the Haar wavelet transform. We have

$$LH = \{W'_1 + \hat{W}_2 - W'_3 - \hat{W}_4\} / 2, \tag{5}$$

$$HH = \{W'_1 - \hat{W}_2 - W'_3 + \hat{W}_4\} / 2 \tag{6}$$

where  $LH$  and  $HH$  are the resulting coefficients in the low-high and high-high subbands, respectively. Meanwhile,  $LL$  and  $HL$  can be recovered from the correctly received substreams. That is

$$LL = \{W'_1 + W'_2 + W'_3 + W'_4\} / 2 = {}^2D R'(x, y, t), \tag{7}$$

$$HL = \left\{ W'_1 - W'_2 + W'_3 - W'_4 \right\} / 2 = W'_1 + W'_3 - {}^{2D}R'(x, y, t) \quad (8)$$

where  $LL$  and  $HL$  are the resulting coefficients in the low-low and high-low subbands, respectively. Then, the missing coefficients can be recalculated by applying the inverse Haar wavelet transform such that

$$W_2^* = \{LL + LH - HL - HH\} / 2 = \left\{ W'_2 + W'_4 + \hat{W}_2 - \hat{W}_4 \right\} / 2 \quad (9)$$

$$W_4^* = \{LL + HH - LH - HL\} / 2 = \left\{ W'_2 + W'_4 + \hat{W}_4 - \hat{W}_2 \right\} / 2. \quad (10)$$

The sum of squared difference (SD) between the original and reconstructed coefficients is

$$\begin{aligned} SD &= \left\{ W_2^* - W'_2 \right\}^2 + \left\{ W_4^* - W'_4 \right\}^2 \\ &= \left\{ (\hat{W}_2 - W'_2)^2 + (\hat{W}_4 - W'_4)^2 - 2 * (\hat{W}_2 - W'_2)(\hat{W}_4 - W'_4) \right\} / 2. \end{aligned} \quad (11)$$

Compared to the sum of SD obtained using only an existing error concealment technique, the sum of SD can be decreased by

$$\begin{aligned} \Delta &= \left\{ \hat{W}_2 - W'_2 \right\}^2 + \left\{ \hat{W}_4 - W'_4 \right\}^2 - SD \\ &= \left\{ (\hat{W}_2 + \hat{W}_4) - (W'_2 + W'_4) \right\}^2 / 2. \end{aligned} \quad (12)$$

Note that not all of the missing coefficients in the spatial root subband can be estimated by root subband recovery when the coefficient  ${}^{2D}R'(x, y, t)$  has to be set to zero (i.e.,  ${}^{2D}R'(x, y, t) = 0$ ). In this case, the missing coefficients are only able to be estimated by using other error concealment techniques, such as bilinear interpolation.

### 3 Simulation Results

The experiments are conducted using the  $352 \times 240 \times 48$  monochrome “Football” and “Susie” video sequences with frame rate of 30 frames/s. We use 16 frames for the GOF. A three-level decomposition using the 9/7 biorthogonal wavelet filter is applied in both the spatial and temporal domains. The wavelet coefficients are divided into multiple groups ( $S = 4$  or  $S = 16$ ) and each group is independently encoded at the coding rate  $c = 1.0$  bit/pixel (bpp). The peak signal-to-noise ratio (PSNR)

$$PSNR = 10 \log_{10} \left( \frac{255^2}{MSE} \right) dB \quad (13)$$

is used as the distortion metric where MSE is the mean-squared error between the original and the reconstructed video sequence.

**Table 1.** Comparison of PSNR (dB) for “Football” and “Susie” video sequences in noiseless channels

(a) *Football*

Total number of substreams	Proposed algorithm (dB)	Previous work in [8] (dB)	ERC-SPIHT (dB)
$S = 4$	33.16	33.20	33.35
$S = 16$	32.96	N/A	33.16

(b) *Susie*

Total number of substreams	Proposed algorithm (dB)	Previous work in [8] (dB)	ERC-SPIHT (dB)
$S = 4$	43.72	43.75	43.87
$S = 16$	43.51	N/A	43.66

**Table 2.** Comparison of PSNR (dB) for “Football” and “Susie” video sequences in a noisy channel

(a) *Football*

Total number of substreams	Substream(s) lost	Proposed algorithm (dB)	Previous work in [8] (dB)	ERC-SPIHT (dB)
$S = 4$	2	26.00	25.93	24.93
	4	25.91	25.84	24.81
	2 & 4	21.56	N/A	21.51
$S = 16$	2	30.56	N/A	29.71
	4	30.55	N/A	29.60
	2 & 4	28.48	N/A	27.64

(b) *Susie*

Total number of substreams	Substream(s) lost	Proposed algorithm (dB)	Previous work in [8] (dB)	ERC-SPIHT (dB)
$S = 4$	2	30.74	30.73	29.71
	4	30.58	30.57	29.41
	2 & 4	26.52	N/A	25.97
$S = 16$	2	39.85	N/A	37.78
	4	39.51	N/A	37.27
	2 & 4	36.45	N/A	34.62

### • Source Coding Efficiency

We assume that all substreams are thus received correctly at the decoder. Table 1 shows the average PSNR values for the proposed algorithm, the previous work in [8] and ERC-SPIHT when the total number of substreams is  $S = 4$  or  $S = 16$ . When  $S = 4$ , the PSNRs obtained with the proposed algorithm and the previous work [8] are lower than ERC-SPIHT (0.15 – 0.19 dB). We also observe that the PSNR values of the proposed algorithm are slightly lower than those of [8]. It means that more bits are assigned to the additional redundancy in the final bitstreams in the proposed video coder. When  $S = 16$ , the previous work [8] is not applicable. It can be seen that the difference between the proposed algorithm and ERC-SPIHT is 0.15 – 0.20 dB in terms of PSNR.

### • Error Concealment Performance

We assume there is one or two substreams corrupted due to transmission errors. Table 2 presents the average PSNR values for the proposed algorithm, the previous work [8] and ERC-SPIHT. When only one substream is corrupted by noisy channels, the average PSNRs of the proposed algorithm outperforms ERC-SPIHT by up to 2.24 dB. In addition, the average PSNR value obtained with the proposed algorithm are slightly higher than [8] ( $S = 4$ ) since a larger amount of additional redundancy is inserted among substreams in the proposed algorithm, which will increase the accuracy in the estimation of missing coefficients in root subband recovery. When there are two substreams lost and these two missing substreams are in the same  $2 \times 2$  coefficient block, the missing coefficients can be recalculated by using (9)–(10). Compared to those with ERC-SPIHT, the average PSNRs with the proposed algorithm can be further improved by up to 1.83 dB.

## 4 Conclusions

We presented a domain-partitioning based MD video coding algorithm with 3-D SPIHT for robust video transmission. Although the proposed algorithm provides lower source coding efficiency than ERC-SPIHT in lossless transmission, the proposed approach has been shown to be more resilient in an error-prone transmission environment.

## References

- [1] Said, A., Pearlman, W.A.: A New, Fast, and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees. *IEEE Trans. Circuits Syst. Video Technol.* 6(3), 243–250 (1996)
- [2] Guo, Y., Chen, Y., Wang, Y.-K., Li, H., Hannuksela, M.M., Gabbouj, M.: Error Resilient Coding and Error Concealment in Scalable Video Coding. *IEEE Trans. Circuits Syst. Video Technol.* 19, 781–795 (2009)
- [3] Wu, D., Hou, Y.T., Zhang, Y.-Q.: Transporting Real-Time Video Over the Internet: Challenges and Approaches. *Proc. IEEE* 88(12), 1855–1877 (2000)
- [4] Goyal, V.K.: Multiple Description Coding: Compression Meets the Network. *IEEE Signal Processing Mag.* 18, 74–93 (2001)
- [5] Cho, S., Pearlman, W.A.: A Full-Featured, Error-Resilient, Scalable Wavelet Video Codec Based on the Set Partitioning in Hierarchical Trees (SPIHT) Algorithm. *IEEE Trans. Circuits Syst. Video Technol.* 12, 157–171 (2002)
- [6] Cho, S., Pearlman, W.A.: Error Resilience and Recovery in Streaming of Embedded Video. *Journal of Signal Process* 82, 1545–1558 (2002)
- [7] Cho, S., Pearlman, W.A.: Error Resilient Video Coding With Improved 3-D SPIHT and Error Concealment. In: *Proc. of SPIE/ID&T Int. Conf. on Electron. Imaging*, vol. 5022, pp. 125–136 (2003)
- [8] Zhu, J., Cuhadar, A.: Error-Resilient Multiple-Description Video Coding With 3-D SPIHT. In: *Proc. IEEE PacRim Conf. Communications, Computers and Signal Process.*, pp. 606–611 (2009)