MULTIPLE DESCRIPTION VIDEO CODING WITH 3D-SPIHT EMPLOYING A NEW TREE STRUCTURE

Moyuresh Biswas, Michael Frater, John Arnold

School of Information Technology & Electrical Engineering
University College, University of New South Wales
Australian Defence Force Academy, Australia

ABSTRACT

Multiple Description Coding (MDC) of video is a useful technique for robust video transmission over unreliable networks. MDC not only provides at least acceptable quality of video in error-prone network but it also efficiently utilizes the multipath nature of internet and wireless networks. In this paper, multiple description video coding with 3-D Set Partitioning in Hierarchical Tree (3D-SPIHT) codec has been proposed. In particular, we propose branch-pruning technique to generate multiple descriptions. We then propose a new tree structure for 3D-SPIHT which is particularly efficient for MDC. Experimental results prove the effectiveness of the proposed method.

***Index Terms*—** Multiple description coding, set partitioning, group of frames.

1. INTRODUCTION

Transmission and sharing of multimedia (i.e. voice, image, video) over wired and/or wireless networks is an attractive and useful application. However, as the compression performance of the state-of-the-art image/video coders increases, so does the likelihood that the resulting coded information becomes more sensitive to errors caused by unreliable transmission channels. Error-resilient coding is therefore often required.

Multiple Description Coding (MDC) [1] is a promising technique for error-resilient video transmission. The concept behind MDC is to generate two or more independent descriptions of image/video content so that in presence of network congestion or other sources of packet loss, an acceptable quality can be achieved if at least one description is received correctly; and the received video quality gradually increases as more descriptions are received successfully. One important aspect of MDC is that it can efficiently utilize the diversity communication framework of current wired/wireless networks (commonly known as ‘path diversity’ where multiple communication links between sender and receiver are available.

Existing works on MDC include Multiple Description Scalar Quantizer (MDSQ) [2] proposed by Vaishampayan. In MDSQ a scalar quantizer is represented by two indices each of which belongs to different descriptions. The idea of MDSQ was applied on predictive video coding in [3]. MD method with Pairwise Correlating Transform (PCT) was proposed in [4]. PCT introduces controlled correlation between a pair of DCT transformed coefficients; the correlation helps estimating the lost component. Another class of MD method is based on pre and/or post-processing approach e.g. zero-padding technique as proposed in [5].

In this paper, we propose a Multiple Description video coding technique with 3D- Set Partitioning in Hierarchical Tree (SPIHT) [6] codec. The SPIHT algorithm was selected because it exhibits state-of-the-art compression along with other suitable features such as progressive transmission, SNR scalability. Experimental results show that our method performs better than the existing methods.

The rest of the paper is organized as follows. In Section 2 SPIHT algorithm and its extension for video coding is briefly described. Section 3 explains MD video coding with 3D-SPIHT. Experimental results are presented in Section 4. Section 5 concludes the paper.

2. SPIHT AND ITS 3D EXTENSION

Set partitioning in hierarchical trees (SPIHT) [6], invented by Said and Pearlman, is a popular image coding technique which is a derivative of the Embedded Zerotree Wavelet (EZW) algorithm originally proposed by Shapiro [7]. The SPIHT algorithm proceeds with two passes: *sorting* pass and *refinement* pass. In the sorting pass, coefficients are tested for significance relative to a given threshold. Instead of having each coefficient tested, SPIHT introduces a set-partitioning sorting technique where a set (of coefficients) is tested for significance such that the result of this significance test is true (or false) for all coefficients in the set. This is done by introducing a spatial orientation tree which utilizes the spatial relationship among the pixels resulting from hierarchical subband decomposition. Figure 1 shows the spatial tree structure of SPIHT. As can be seen,

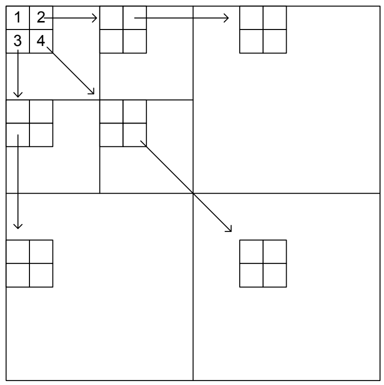

Fig. 1. Spatial Orientation Tree of SPIHT

coefficients in the lowest subband form a group of 2x2 pixels. SPIHT is a progressive bitplane coding technique. In the refinement pass, the coefficients are successively refined at each bitplane. For detailed algorithmic description of SPIHT, the reader is referred to [6].

In application to video coding, three-dimensional SPIHT (3D-SPIHT) algorithm was proposed by Kim *et al* [8] by extending the idea of set partitioning to three dimensions. 3D-SPIHT operates on a group of frames (GOF) which are first wavelet transformed both spatially and temporally. In [8], a *symmetric* spatio-temporal orientation tree structure is proposed where coefficients in the lowest spatio-temporal subband form a group of $2 \times 2 \times 2$ adjacent pixels. Later, Cho and Pearlman proposed an *asymmetric* tree structure for the 3D-SPIHT (AT-SPIHT) [9] which always gives a longer tree than its *symmetric* counterpart. This is motivated by the fact that tree-based coders like SPIHT give better performance when the tree depth is longer. As AT-SPIHT tree is of particular relevance to our work, it is illustrated graphically in Figure 2 for a group of 8 frames with 2 levels of spatial decomposition and 2 levels of temporal decomposition. The spatial tree structure in each frame is exactly the same as the 2-D case. However, the top-left coefficient of each 2×2 group in the lowest spatial band of each frame (except those belonging to highest temporal subband) also forms a tree branch, unlike the 2-D SPIHT case where such coefficients have no descendant. The top-left coefficient of each 2×2 group in the lowest spatio-temporal subband (LL) has one 2×2 offspring group in next higher temporal level (LH). Again, the top-left coefficient of each 2×2 group in temporal LH band has two 2×2 offspring groups in temporal H band. Such coefficients in temporal H band have no descendant. Another related work is the Spatio-Temporal Tree Preserving SPIHT (STTP-SPIHT) [10] where wavelet coefficients are divided into different blocks according to their spatial and temporal relationships and each block is coded independently.

3. MDC WITH SPIHT

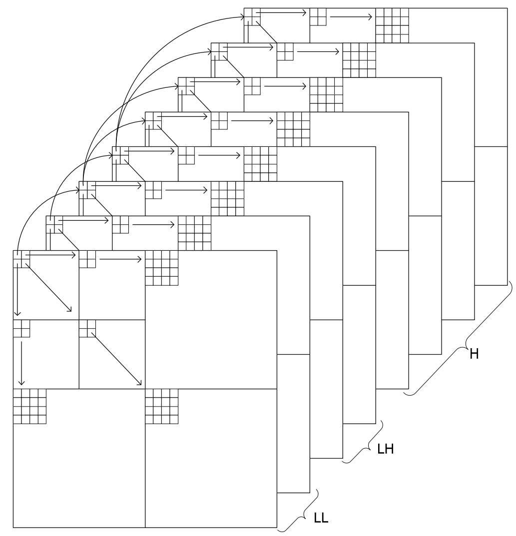

Fig. 2. Asymmetric tree structure of AT-SPIHT

The excellent performance of 3D-SPIHT algorithm in single description case suggests its possible extension to MDC. There has only been preliminary work in this regard. Kim *et al* [11] proposed an unbalanced MDC with STTP-SPIHT [10] as the base coder. The idea is to distribute the spatio-temporal tree blocks generated by STTP-SPIHT among the descriptions in such a way as to satisfy the rate constraints of different paths.

In this paper, we propose a new technique for MD extension of 3D-SPIHT. Unlike other techniques, our approach utilizes the inherent tree structure of the 3D-SPIHT coder to generate descriptions. We focus on the practical case of two descriptions scenario. In particular, the contributions of this paper are:

a) *We propose a new approach to MD video coding using branch-pruning technique to separate the descriptions.*
b) *We propose a modified tree structure for 3D-SPIHT which is particularly efficient for MDC.*

3.1. Overview of Proposed Method

In 3D-SPIHT, all trees are rooted at the lowest spatio-temporal subband, i.e. each coefficient at the lowest spatio-temporal subband is the root of a different tree. As a tree grows along the spatial and temporal directions, new branches originating from coefficients of different subbands are added to the tree. The benefit of this tree structure is that it gives the flexibility to selectively prune the branches of a tree. For example, we can prune the branches which add higher spatial and/or temporal frequency coefficients to a tree and still get a reduced quality picture. This opens a new avenue for multiple description coding where a description can be generated by pruning some branches of a tree which are included in the other description pruning here those included in the first one. As a simple example, branches originating only at even-indexed coefficient in a frame row can be retained in a description and the other description

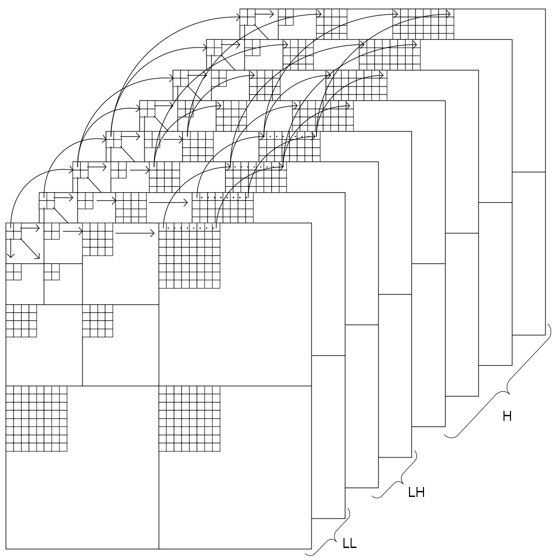

Fig. 3. Proposed tree structure for MDC

may contain only those originating at odd-indexed coefficients. The decision as to whether to prune a branch can be taken in rate-distortion sense. But in our approach we have used the even-odd subsampling to make the pruning decision which, although somewhat ad-hoc, is simple and still found to give good results. Redundancy can be simply added by retaining certain important coefficients (important in the sense that these add coefficients in lower spatio-temporal subband to the tree) common in both descriptions.

We investigate the implication of applying this 'branch pruning' technique to AT-SPIHT as the asymmetric tree structure of AT-SPIHT gives better performance for single description coding. We notice from Fig 2 that in AT-SPIHT tree structure, when a tree branch is pruned e.g. one at a higher temporal level, we tend to lose information about large number of pixels clustered together at a picture area. So we modified the AT-SPIHT tree structure and propose a new tree structure which is more suitable for multiple description coding. The new tree structure is shown in Figure 3 for GOF size 8 with 3 levels of spatial decomposition and 2 levels of temporal decomposition.

In this structure, parent-offspring relationship in the lower spatio-temporal subbands is the same as AT-SPIHT. We modify the tree as we move towards higher frequencies in both spatial and temporal directions. As can been seen, coefficients at higher frequency subbands have either one or two descendants in the temporal direction. The benefit is that it reduces the chance of clustered information loss when such a branch is pruned. This new tree structure has inevitable advantage in application to MDC. For example, if we prune a temporal tree branch we will loose only a few coefficients in the same spatial position at higher temporal level which has less drastic an effect than in the case of AT-SPIHT.

4. EXPERIMENTAL RESULTS

We examined the performance of MD video coding employing the proposed tree structure with the asymmetric tree structure of AT-SPIHT; this is termed AT-SPIHT-MDC. In both cases, we use branch pruning with simple even-odd subsampling for description generation. We also implemented the MDC technique proposed by Kim *et al* [11] but, instead of unbalanced descriptions, we generated two balanced descriptions. This is termed STTP-SPIHT-MDC. We conducted experiments with wide variety of test sequences having diverse motion. Experiment parameters are as follows. For the QCIF sequences, we performed 2 levels of spatial wavelet decomposition; for the CIF sequences, we did 3 levels of decomposition. In all cases, we used 8 frames in a GOF and 3 levels of temporal decomposition. All comparisons are made assuming that only one description is received entirely. A selection of rate distortion curves for various picture sizes (QCIF and CIF), test sequences and MDC redundancies is presented below. Each test was performed using 296 frames from the sequence. Results for QCIF sequences are shown in Figure 4, while results for CIF sequences are shown in Figure 5.

The results in Figures 4 and 5, which are consistent with results we have obtained across a broad range of test sequences with low-to-medium motion, demonstrate that our proposed method outperforms previously proposed techniques by 0.2~1.0 dB in PSNR.

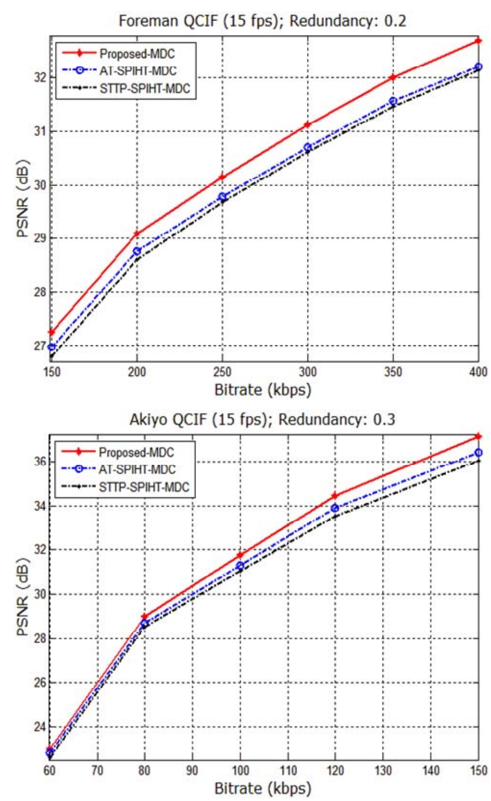

Fig. 4. PSNR results for QCIF sequences

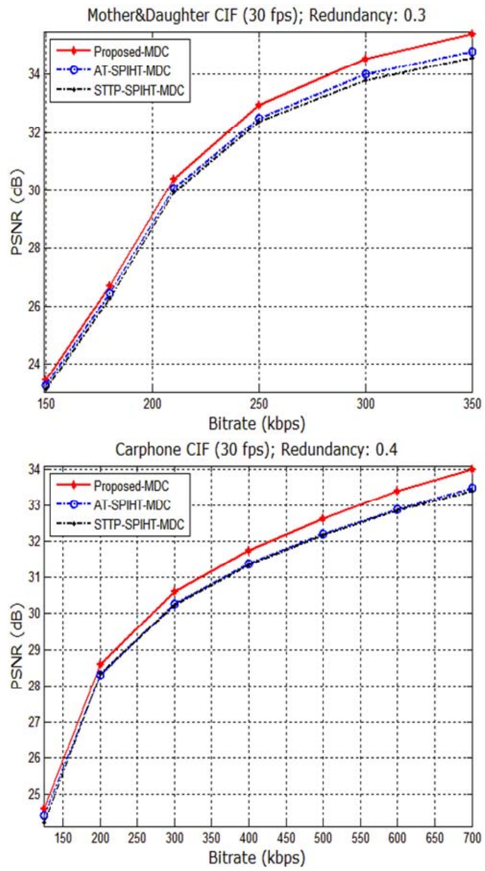

Fig. 5. PSNR results for CIF sequences

An extract from a sample from of the Carphone sequence is shown in Figure 6, which demonstrates a quite clear subjective-quality improvement using our proposed method, in spite of the small difference in PSNR (0.5 dB) for the two pictures. For high motion sequences, including Football and Stefan, out proposed technique still provides a higher level of performance, with improvements of approximately 0.15~0.35 dB noticed over the same bit rate range as shown in this paper.

5. CONLUSION AND FUTURE WORK

In this paper, a new technique for multiple description video coding with SPIHT algorithm has been presented. A modified tree structure for 3D-SPIHT has been proposed which suites better the case of MDC. It is evident from the experimental results that, the proposed technique gives better rate-distortion performance than the existing techniques [11]. The possible applications of the proposed method include video transmission over wireless networks and packet video networks. We are currently investigating these applications.

Fig. 6. Extract from Carphone QCIF Frame#88; bps=100k, 40% redundancy. (a) Proposed Method (PSNR 30.4 dB) (b) STTP-SPIHT-MDC (PSNR 29.9 dB)

6. REFERENCES

[1] V.K.Goyal, "Multiple Description Coding: Compression Meets the network," *IEEE Signal Proc. Magazine,* vol. 18, pp. 74-93, Sept. 2001.

[2] V. A. Vaishampayan, "Design of Multiple Description Scalar Quantizers," *IEEE Trans. Inform. Theory,* vol. 39, pp. 821-834, May 1993.

[3] V. Vaishampayan and S. John, "Balanced interframe multiple description video compression," in *Proc. ICIP*, 1999.

[4] Y. Wang, M. T. Orchard, V. A. Vaishampayan, and A. R. Reibman, "Multiple description coding using pairwise correlating transforms," *IEEE Trans. Image Proc.,* vol. 10, pp. 351-366, March 2001.

[5] D. Wang, N. Canagarajah, D. Redmill, and D. Bull, "Multiple Description video coding based on zero padding," in *Proc. ISCAS*, 2004.

[6] A. Said and W. A. Pearlman, "A New Fast and Efficient Image Codec based on Set Partitioning in Hierarchical Trees," *IEEE Trans. on Cir. & Sys. on video Tech.,* vol. 6, pp. 243-250, June 1996.

[7] J. M. Shapiro, "Embedded Image Coding using zerotrees of wavelet coefficients," in *IEEE Trans. on Signal Proc.* vol. 41, Dec 1993, pp. 3445-3462.

[8] B.-J. Kim, Z. Xiong, and W. A. Pearlman, "Low bit-rate scalable video coding with 3-D set partitioning in hierarchical trees (3D-SPIHT)," *IEEE Trans. on Cir. & Sys. on video Tech.,* vol. 10, pp. 1374-1387, Dec. 2000.

[9] S. Cho and W. A. Pearlman, "Error Resilient Video Coding with improved 3-D SPIHT and Error Concealment," in *Proc. of SPIE/ID&T Int. Conf. on Electronic Imaging*, Santa Clara, CA, 2003.

[10] S. Cho and W. A. Pearlman, "A full-featured, error resilient, scalable wavelet video codec based on the Set Partitioning in Hierarchical Trees (SPIHT) algorithm," *IEEE Trans. on Cir. & Sys. on video Tech.,* vol. 12, pp. 157-171, March 2002.

[11] J.Kim, R. M. Mersereau, and Y. Altunbasak, "Distributed Video Streaming using Multiple Description Coding and Unequal Error Protection," *IEEE Trans. Image Proc.,* vol. 14, pp. 849-861, July 2005.