# A 3-D Virtual SPIHT for Scalable Very Low Bit-Rate Embedded Video Compression

Habibollah Danyali and Alfred Mertins

University of Wollongong
School of Electrical, Computer and Telecommunications Engineering
Telecommunication and Information Technology Research Institute (TITR)
Wollongong, NSW 2522, Australia
Email: hd04@uow.edu.au, mertins@uow.edu.au

## Abstract

In this paper we propose a modification of the 3-D Set Partitioning in Hierarchical Trees (3-D SPIHT) algorithm for very low bit-rate wavelet-based video coding. The modified algorithm, called 3-D Virtual SPIHT (3-D VSPIHT), virtually decomposes the coarsest level of the wavelet coefficients to reduce the number of three-dimensional (spatio-temporal) sets in the 3-D SPIHT algorithm. Our simulation results show that the proposed codec has better performance than the original 3-D SPIHT algorithm, especially for very low bit-rate video coding. The low complexity of the codec and the embeddedness property of the output bitstream make it a convenient coding technique for applications such as Internet video streaming via low bit rate channel. Moreover it has a good potential to carry spatial and temporal scalability, which are especially important for the new multimedia applications.

## 1 Introduction

Image and video coding plays an important role in many multimedia applications. At very low bit-rate, traditional video coding based on DCT suffers from blocking artifacts. There is a high demand for an efficient video coding technique which can provide an acceptable quality for very low bit-rate and support new functionalities such as PSNR, frame rate and spatial scalability.

Recently many researchers have focussed on wavelet-based image and video compression. There are two main groups of wavelet-based video coding: The first group called hybrid coding uses a motion estimation and compensation algorithm for reducing temporal redundancy of video frames and wavelet transform for spatial domain [1–4]. These schemes are mainly similar to the H.263 standard [5], but with the DCT replaced by the wavelet transform. In the second group, which can be with or without motion compensation, a wavelet transform is applied to both temporal and spatial domains [6–11]. Due to the multiresolution property of the wavelet transform these methods are highly scalable in terms of spatio-temporal resolution and PSNR, have low complexity (in the case of no motion compensation), and are suitable for progressive video transmission, especially for video streaming through heterogeneous networks with a large variation in bandwidth and user-device capabilities.

An efficient algorithm for coding of the decomposed wavelet coefficients is a necessary part of a wavelet-based image and video coding system. The Embedded Zerotree Wavelet (EZW) algorithm by Shapiro [12] opened a new and important window toward zero tree coding of wavelet coefficients. The further development of EZW by Said and Pearlman [13], which is known as Set Partitioning in Hierarchical Trees (SPIHT), provides one of the best performing wavelet-based image compression algorithms. The excellent rate-distortion performance and scalable nature of SPIHT for still images make it an attractive coding strategy also for video coding.

A 3-D extension of SPIHT for video coding has been proposed by Kim and Pearlman [6]. They apply a 3-D dyadic wavelet to a group of video frames (GOF) and code the wavelet coefficients by 3-D SPIHT. Even with no motion estimation and compensation this method performs measurably and visually better than MPEG-2 which employs complicated means of motion estimation and compensation. Recently Kim et al. [7] have used a 3-D wavelet packet and 3-D SPIHT with and without motion compensation for low bit rate scalable video coding.

In this paper we employ a three-dimensional wavelet packet of [7] for decomposing a group of frames (GOF) to wavelet coefficients and propose a virtual definition of the wavelet coefficients sets for 3-D SPIHT. This algorithm is referred to as 3-D VSPIHT. We will show that it outperforms 3-D SPIHT particularly for very low bit-rate video coding and sup-
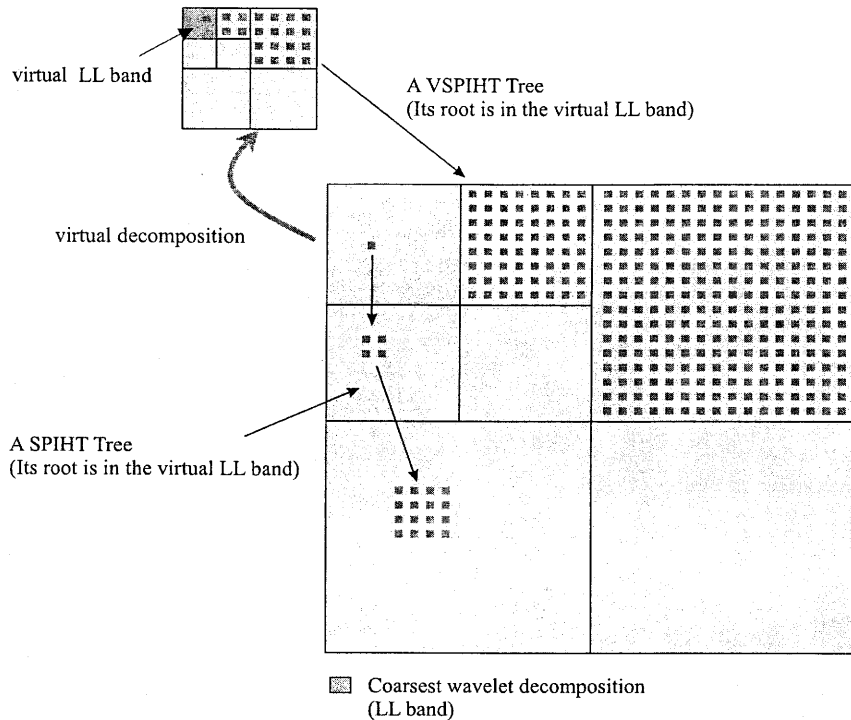
Figure 1: Virtual SPIHT concept.

ports attractive features of 3-D SPIHT such as speed, simplicity, and providing a fully embedded bit stream.

This paper is organized as follow. The next section, Section 2, describes the proposed 3-D VSPIHT algorithm. Section 3 gives a brief overview of the proposed video coding system. The implementation details and simulation results are given in Section 4. Finally, some conclusions are presented in Section 5.

## 2   Three Dimensional Virtual SPIHT (3-D VSPIHT)

In this section we first give a brief description of the SPIHT algorithm, then explain the VSPIHT. The SPIHT consists of three stages: initialization, sorting and refinement. It sorts the information of wavelet coefficients in three ordered lists: list of insignificant sets (LIS), list of insignificant pixels (LIP) and list of significant pixels (LSP). At the initialization stage the SPIHT first defines a start threshold due to the maximum value in the wavelet coefficients pyramid, then sets the LSP as an empty list and puts the coordinates of all coefficients in the coarsest level of the wavelet pyramid (LL band) in the LIP and those which have descendants to the LIS. In the sorting pass, the algorithm first starts to sort the elements in the LIP then in the LIS. For each pixel in the LIP it performs a sig-

nificance test against the current threshold and outputs the test result (0 or 1) to the output bitstream. If a coefficient is significant, its sign is coded and then its coordinate is moved to the LSP. During the sorting pass of LIS, the SPIHT does the significance test for each set in the LIS and outputs the significance information (0 or 1). If a set is significant, it is partitioned into its offspring and leaves. After the sorting pass for all elements in the LIP and LIS, SPIHT does a refinement pass with the current threshold for all entries in the LSP, except those which have been moved to the LSP during the last sorting pass. Then the current threshold is divided by 2 and the sorting and refinement stages are continued until a predefined bit-budget is exhausted.

The number of sets in the LIS in the initializing stage can affect the PSNR performance of the SPIHT bitstream particularly at very low bit rates. For the case of a high number of sets in the LIS, at early sorting passes many bits in the output bitstream are wasted for coding the significance of these sets, because most of the sets will be insignificant during several stages until the threshold has been sufficiently reduced. These sorting bits could be better spent for sorting of insignificant coefficients in the LIP and the refinement of the previously known significant coefficients.

A 2-D VSPIHT has been introduced in [14] for coding of wavelet coefficients of motion compensated
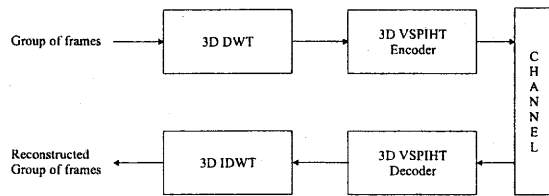
Figure 2: 3-D VSPIHT coding system block diagram.



Figure 3: A 3-D wavelet packet decomposition with 2 levels temporal and 3 levels spatial decomposition (21 bands).

error frames in a video coding system similar to H.263. The philosophy behind the VSPIHT is to reduce the number of sets in the LIS by assuming longer trees than naturally provided by the wavelet transform at the initialization stage. It virtually assumes further decomposition for the coarsest level (LL band) of the wavelet pyramid and considers the roots of trees that should be put into the LIS at initialization stage in the coarsest LL virtual band. See Figure 1 for an illustration of the concept. Each of the defined sets consists of two parts: a real part which is located in the actually decomposed section of the wavelet tree and a virtual part which is in the virtually decomposed section of the pyramid. During sorting pass of LIS, for test of significance, VSPIHT looks only to the real part of each set. Moreover when a set is partitioned to its subsets, if its offspring are in the virtual part of the set, there is no need to send information about significance of the offspring. The decoder follows the same procedure as the encoder.

In this paper, we extend the VSPIHT method from 2-D to 3-D and study its performance in conjunction with 3-D wavelet transforms. The 3-D VSPIHT extends the concept of the VSPIHT for both spatial and temporal dimensions. Because the size of the lowest band that is to be virtually decomposed may not allow a straightforward extension of the wavelet trees we first virtually extend this band to a size $N_v \times N_h \times N_\tau$ where $N_v$, $N_h$ and $N_\tau$ are the smallest powers of two that are equal or larger than the size of this band. Then the virtually extended band is virtually decomposed until the size of the coarsest virtual LL band becomes $2 \times 2 \times 2$. The number of sets in the LIS is then reduced to 7, which is the minimum number of the sets for the 3-D case.

## 3   The Video Coding System

Figure 2 shows the block diagram of the proposed video codec. A group of frames (GOF) is transformed with a three-dimensional discrete wavelet transform (3-D DWT) which provides a spatio-temporal wavelet decomposition. Then the encoder executes the 3-D VSPIHT encoding algorithm and the output bitstream is sent to the channel. At the decoder side, the received
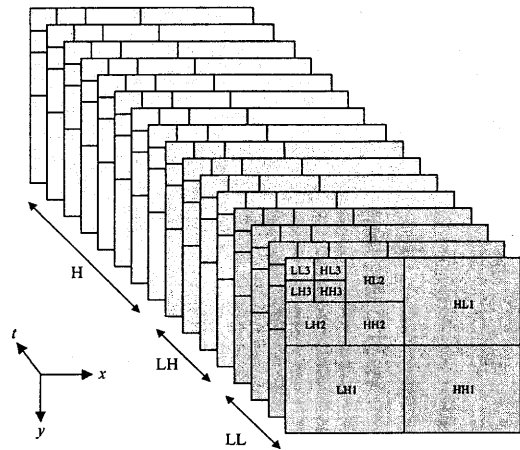
bitstream is first decoded by a 3-D VSPIHT decoder and then an inverse 3-D DWT provides a reconstructed version of the GOF. There are two main types of 3-D DWTs: dyadic DWT and wavelet packets [7]. In the dyadic case, a temporal decomposition is followed by a spatial decomposition and this procedure is repeated for the lowest spatio-temporal band for further decomposition levels, therefore the number of decomposition levels in the temporal and spatial domain is the same. The number of subbands in this case is $7N + 1$ where $N$ is the number of the spatio-temporal decomposition levels. For the wavelet packets transform, the number of temporal and spatial decompositions can be different. In this case, a wavelet transform is applied in temporal direction to produce any desired temporal decomposition levels. Then all frames in the GOF are individually decomposed in the horizontal and vertical spatial directions. The number of subbands is $(N_t + 1)(3N_s + 1)$ where $N_t$ and $N_s$ are the temporal and spatial decomposition levels respectively. 3-D wavelet packets are more efficient and flexible than a pure dyadic decomposition especially for cases when the number of frames in a GOF is small and the spatial size of image is big and more levels of wavelet decomposition for spatial than temporal direction are needed.

## 4   Implementation Details and Results

In this section we first describe some features of our implementation such as the 3-D wavelet transform, filter choice, size and extension of video sequences, and number of frames in a GOF. Then we present some results of our simulation.

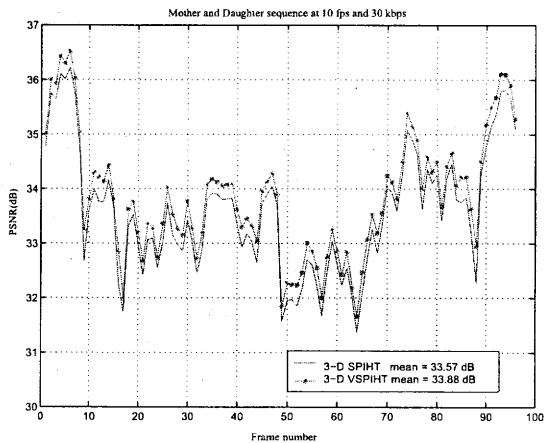For testing of the 3-D VSPIHT we have selected

125

Figure 4: Mother and daughter sequence luminance, PSNR comparison of 3-D SPIHT and 3-D VSPIHT.



Figure 5: Hall Monitor sequence luminance, PSNR comparison of 3-D SPIHT and 3-D VSPIHT.

QCIF sequences (image size of 144 x 176) and a frame rate of 10 fps, which is suitable for very low bit rate video coding. We have used a GOF size of 16 and the 9/7-tap [15] filters for both temporal and spatial decomposition. A 3-D wavelet packet firstly applies two levels of decomposition in the temporal direction then three levels in the spatial domain for all frames in a GOF. See Figure 3 for an illustration of the transformed GOF.

The codec was tested for two different video sequences: "Mother and Daughter" and "Hall Monitor". "Mother and Daughter" is a typical head and shoulder sequence with small object motion. "Hall Monitor" is an example of a monitoring application. It has a fixed background and occasionally some persons appear and disappear. Figures 4 and 5 show PSNR results for 3-D SPIHT and 3-D VSPIHT for 96 frames of the luminance of the sequences, at a bit rate of 30 kbps. For individual frames the PSNR improvements due to 3-D VSPIHT are between 0.15 to 0.67 dB for Mother and Daughter and between 0.38 to 0.58 for Hall Monitor sequences. The mean PSNR value for all frames for 3-D VSPIHT outperforms the mean PSNR value of the 3-D SPIHT by 0.31 dB for the Mother and Daughter and 0.51 dB for the Hall Monitor sequence. To give some visual examples, Figures 6 and 7 show original frames and their reconstructed versions at bit rates of 30 kbps. In these examples, 3-D SPIHT and 3-D VSPIHT are very close to each other in terms of visual performance, but especially on some edges one can see a slightly better quality provided by 3-D VSPIHT.

## 5   Conclusions

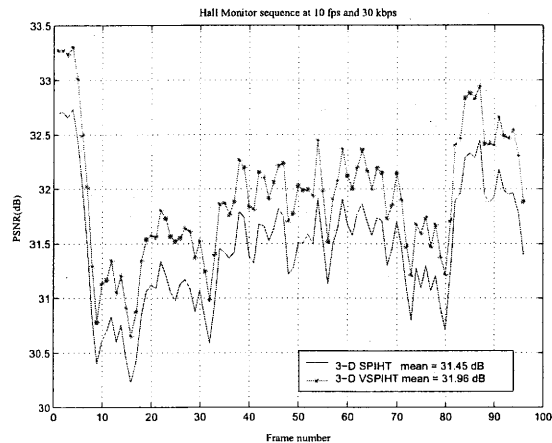We proposed a 3-D VSPIHT for video coding which improves the 3-D SPIHT PSNR quality especially for very low bit-rates. The codec is without motion compensation, which makes it less complex than other motion compensated codecs. As shown in [7], a motion compensated 3-D SPIHT can improve the PSNR quality of 3-D SPIHT in some cases, especially for video sequences in which there is a considerable camera pan and zoom. It is clear that the same situation is correct for 3-D VSPIHT. Also using an arithmetic coding on output bit stream can improve the PSNR. Like the 3-D SPIHT the codec provides a fully embedded bitstream and has a great potential for scalability in both temporal and spatial directions, which is very important for many multimedia applications.

## Acknowledgment

## References

[1] Y.-Q. Zhang and S. Zafar, "Motion-compensated wavelet transform coding for video compression," *IEEE Trans. Circ. and Syst. for Video Technology*, vol. 2, pp. 285–296, Sept. 1992.

[2] P. Cheng, J. Li, and C.-J. Kuo, "Multiscale video compression using wavelet transform and motion compensation," in *Proc. IEEE Int. Conf. Image Processing*, 1995, pp. 606–609.

[3] S. A. Martucci, I. Sodagar, T. Chiang, and Y.-Q. Zhang, "A zerotree wavelet video coder," *IEEE Trans. Circ. and Syst. for Video Technology*, vol. 7, no. 1, pp. 109–118, Feb. 1997.

Figure 6: Mother and daughter frame 92, at 30kbit/sec and 10 fps. (a) Original (b) 3-D SPIHT (PSNR=34.62dB) (c) 3-D VSPIHT (PSNR=35.13dB)
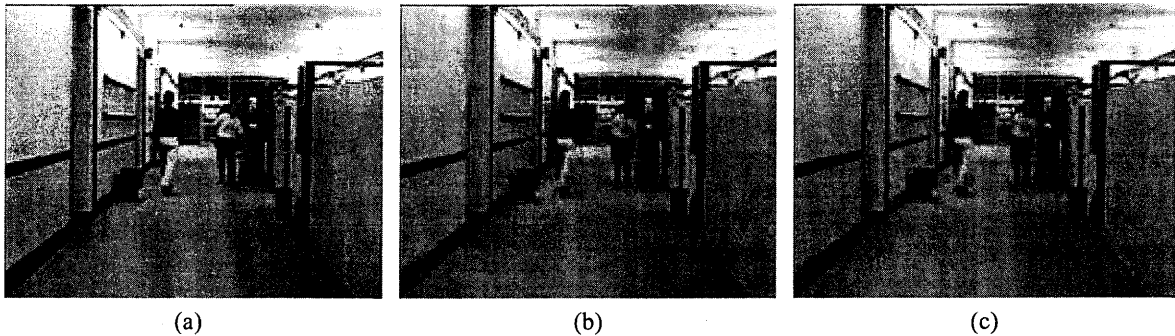


Figure 7: Hall monitor frame 55, at 30kbit/sec and 10 fps. (a) Original (b) 3-D SPIHT (PSNR=31.94dB) (c) 3-D VSPIHT (PSNR=32.41dB)

[4] J. Karlenkar and U. B. Desai, "SPIHT video coder," in *IEEE Region 10 International Conference on Glaobal Connectivity in Energy, Computer, Communication and Control, TEN-CON'98*, 1998, vol. 1, pp. 45–48.

[5] ITU-T, "Recommendation H.263 - Video coding for low bitrate communication," May 1996.

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

[7] B.-J. Kim, Z. Xiong, and W. A. Pearlman, "Low bit-rate scalable video coding with 3-d set partitioning in hierarchical trees (3-D SPIHT)," *IEEE Trans. Circ. and Syst. for Video Technology*, vol. 10, no. 8, pp. 1374–1387, Dec. 2000.

[8] C. Podichuk, N. Jayant, and N. Farvardin, "Three-dimensional subband coding of video," *IEEE Trans. Image Processing*, vol. 4, no. 2, pp. 125–139, Feb. 1995.

[9] S.-J. Choi and J. W. Wood, "Motion compensated 3-d subband coding of video," *IEEE Trans. Image Processing*, vol. 8, no. 2, pp. 155–167, Feb. 1999.

[10] J. R. Ohm, "Three-dimensional subband coding with motion compensation," *IEEE Trans. Im-age Processing*, vol. 3, no. 5, pp. 559–571, Sept. 1994.

[11] J. Y. Tham, S. Ranganath, and A. A. Kassim, "Highly scalable wavelet-based video codec for very low bit-rate environment," *IEEE J. Select. Areas Commun.*, vol. 16, no. 1, pp. 12–27, Jan. 1998.

[12] J. M. Shapiro, "Embedded image coding using zerotree of wavelet coefficients," *IEEE Trans. Signal Processing*, vol. 41, pp. 3445–3462, Dec. 1993.

[13] A. Said and W. A. Pearlman, "A new, fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circ. and Syst. for Video Technology*, vol. 6, pp. 243–250, June 1996.

[14] E. Khan and M. Ghanbari, "Very low bit rate video coding using virtual SPIHT," *IEE Electronics Letters*, vol. 37, no. 1, pp. 40–42, Jan. 2001.

[15] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Processing*, vol. 1, pp. 205–220, Apr. 1992.