

HIGH PERFORMANCE VIDEO CODING TECHNIQUE BASED ON HYBRID TRANSFORMATION

Usama S. Mohammed

Department of Electrical & Electronics Engineering, Faculty of Engineering,
Assiut University, Assiut, Egypt Usama@aun.edu.eg

(Received November 30, 2010 Accepted January 10, 2011)

This paper describes a computationally efficient hybrid video coding scheme. The proposed hybrid coding scheme combines the advantages of simple hardware complexity of the 3-D discrete cosine transform coding and the high performance of the 3-D set partitioning in hierarchical tree (3-D SPIHT) coding. The modification of the subband video data in the wavelet domain is done based on the DCT transformation and the classification of the coefficient in the low-frequency video subband (LL). The modification process provides a new subband video data containing almost the same information of the original one but having smaller values of the wavelet coefficient. Simulation results of the proposed method demonstrate that, with a small addition in the computational complexity of the coding process, the PSNR performance of the proposed algorithm is much higher than that of the 3-D SPIHT test coder.

1. INTRODUCTION

Video compression is an area of research that has received considerable attention over the last few decades. Various compression techniques have been used to store, transmit, and manipulate video data efficiently. Video compression is commonly achieved by removing redundancies in the frequency, spatial and temporal domains. Some video coding techniques, such as predictive coding, transform coding and vector quantization, treat the video as random signals and exploit their stochastic properties to achieve compression [1]. In general, video compression algorithms are at the heart of a digital media explosion, playing an important role in internet and multimedia applications, digital appliances, and handheld devices. Multimedia appliances are ubiquitous in every day life, encompassed by standards such as DVD [2], HDTV [3], and satellite broadcasting [4]. With even budget cell phones capable of playing and recording videos, these appliances continue to increase in pervasiveness. Virtually all video playback devices rely on compression techniques to minimize storage, transmission bandwidth, and overall cost. Athar Ali Moinuddin et.al [5] proposed an efficient and embedded 3-D wavelet based video coding technique. The key idea is the use of a composite block-tree hierarchical structure to link blocks of wavelet coefficients in spatial, temporal and color planes in such a way that a large number of the insignificant sets are coded together. The proposed scheme uses $n \times n$ block of wavelet coefficients as a basic unit, in contrast to the existing 3-D set partitioning in hierarchical trees (3-D SPIHT) algorithm that uses a single coefficient as a basic unit. Zhang C. et.al [6] introduced Spatially Varying Transforms (SVT) for video coding, where the location of the transform block within the macroblock is not fixed but

varying. In this paper [6], they extend this concept and present a novel method, called Variable Block-size Spatially Varying Transforms (VBSVT). VBSVT utilizes Variable Block-size Transforms (VBT) in the SVT framework, and is shown to be more preferable for coding prediction error with different characteristics than fixed block-size SVT and also the standard methods that use fixed or adaptive block sizes at fixed spatial locations. G. Bernabé et.al [7] presented a lossy compression scheme based on the application of the 3D fast wavelet transform to code medical video. This type of video has special features, such as its representation in gray scale, its very few interframe variations, and the quality requirements of the reconstructed images. These characteristics as well as the social impact of the desired applications demand a design and implementation of coding schemes especially oriented to exploit them. They analyze different parameters of the codification process, such as the utilization of different wavelets functions, the number of steps the wavelet function is applied to, the way the thresholds are chosen, and the selected methods in the quantization and entropy encoder. In order to enhance their original encoder, they propose several improvements in the entropy encoder: 3D-conscious run-length, hexadecimal coding and the application of arithmetic coding instead of Huffman.

The remainder of this paper is organized as follows. The 3-D SPIHT is introduced in Section 2. Section 3 introduces the proposed Hybrid coding algorithm (HVC) algorithm. Simulation results are introduced in Section 4. Finally, concluding remarks are given in Section 5.

2. 3-D SET PARTITIONING IN HIERARCHICAL TREE CODING

3-D SPIHT algorithm is based on three basic concepts: (1) code/transmit important information first based on the bit-plane representation of pixels (2) ordered refinement bit-plane transmission, and (3) coding is performed along the predefined path/trees called spatio-temporal orientation trees, which efficiently exploit the properties of a 3-D wavelet transformed video.

3-D SPIHT consists of two main stages as in 2-D SPIHT [8]: sorting and refinement. In the sorting stage, 3-D SPIHT sorts the pixels by magnitude with respect to a threshold, which is a power of two, called the level of significance. However, this sorting is a partial ordering, as there is no prescribed ordering among the coefficients with the same level of significance or highest significant bit. The sorting is based on the significance test of pixels along the spatio-temporal orientation trees rooted from the highest level of the pyramid in the 3-D wavelet transformed video.

Spatio-temporal orientation trees were introduced to test significance of groups of pixels for efficient compression by exploiting self-similarity and magnitude localization properties in a 3-D wavelet transformed video. In other words, the 3-D SPIHT exploits the circumstance that if a pixel in the higher level of pyramid is insignificant, it is very likely that its descendants are insignificant.

For practical implementation, 3-D SPIHT maintains three linked lists, the list of insignificant pixels (LIP), the list of significant pixels (LSP), and the list of insignificant sets (LIS). At the initialization stage, 3-D SPIHT initializes the LIP with all the pixels in the highest level of the pyramid, the LIS with all the pixels in the highest level of pyramid except the pixels which do not have descendants, and the LSP as an empty list. The basic function of actual sorting algorithm is to recursively

partition sets in the LIS to locate individually significant pixels, insignificant pixels, and smaller insignificant sets and move their co-ordinates to the appropriate lists, the LSP, LIP and LIS, respectively. After each sorting stage, 3-D SPIHT outputs refinement bits at the current level of bit significance of those pixels which had been moved to the LSP at higher thresholds. In this way, the magnitudes of significant pixels are refined with the bits that decrease the error the most. This process continues by decreasing the current threshold successively by factors of two until the desired bit-rate or video quality is reached. One can refer to [8] and [9] for more details.

3. PROPOSED HYBRID CODING ALGORITHM

In this work, combination of the 3-D DCT coding and the high performance of the 3-D SPIHT coding is used. The modification of the subband image data in the wavelet domain is done based on the 3-D DCT transformation and the classification of the wavelet coefficients in the LL subband. The modification process provides a new subband video data containing almost the same information of the original one but having smaller values of the wavelet coefficient. The proposed hybrid coding algorithm is described next.

First, one-level of the 3-D discrete wavelet transform is applied on the input video. This process will generate 4-subband data (x_{ll} , x_{lh} , x_{hl} , x_{hh}) in this layer. Next, the based-band video $x_{ll}(i,j,t)$ is compressed using the 3-D DCT coding method, and its reconstructed video is designed as $x'_{ll}(i,j,t)$. The difference between $x_{ll}(i,j,t)$ and $x'_{ll}(i,j,t)$ is the residual based-band video $x''_{ll}(i,j,t)$, i.e.

$$x''_{ll}(i,j,t) = x_{ll}(i,j,t) - x'_{ll}(i,j,t) \tag{1}$$

Weber's law indicates that for a wide range of image sequence intensity, the ratio of the just-noticeable difference to the image sequence's intensity is a constant. To make use of this phenomenon in the proposed coding algorithm, The subband video $x''_{ll}(i,j,t)$, $x_{lh}(i,j,t)$, $x_{hl}(i,j,t)$ and $x_{hh}(i,j,t)$ are normalized as indicated below:

$$\begin{aligned} x''_{ll}(i,j,t) &= \frac{x''_{ll}(i,j,t)}{x_{ll}(i,j,t) + k_1}, & x''_{lh}(i,j,t) &= \frac{k_2 x_{lh}(i,j,t)}{x_{ll}(i,j,t)} \\ x''_{hl}(i,j,t) &= \frac{k_3 x_{hl}(i,j,t)}{x_{ll}(i,j,t)}, & x''_{hh}(i,j,t) &= \frac{k_4 x_{hh}(i,j,t)}{x_{ll}(i,j,t)}. \end{aligned} \tag{2}$$

Where k_i is a constant used to minimize the number of the dominant passes in the 3D- SPIHT coder, and will be determined as follows:

$$\begin{aligned} k_1 &= \left\lceil \frac{(\text{Max}(X''_{ll}) - 4 * \text{Max}(X'_{ll}))}{4} \right\rceil, \\ k_2 &= \left\lceil \frac{2 * \text{Max}(X'_{ll})}{(\text{Max}(X_{lh}))} \right\rceil, \\ k_3 &= \left\lceil \frac{2 * \text{Max}(X'_{ll})}{(\text{Max}(X_{hl}))} \right\rceil, \\ k_4 &= \left\lceil \frac{2 * \text{Max}(X'_{ll})}{(\text{Max}(X_{hh}))} \right\rceil \end{aligned} \tag{3}$$

As in predictive image coding [10], the absolute values of the normalized coefficients are modified as follows,

$$y_p^n(i, j, t) = y_p^n(i, j, t) \times \left[2 - y_p^n(i, j, t) \right] \quad p=1, \dots, q \quad (4)$$

Where $y_0^n(i, j, t)$ is the value of the normalized coefficients $x_{ll}^n(i, j, t)$, $x_{lh}^n(i, j, t)$, $x_{hl}^n(i, j, t)$ and $x_{hh}^n(i, j, t)$, and q is an integer constant which may change with different subbands. After normalization and mapping, four subimages are generated and designated as $y_{ll}^n(i, j, t)$, $y_{lh}^n(i, j, t)$, $y_{hl}^n(i, j, t)$ and $y_{hh}^n(i, j, t)$, which are then located in the corresponding positions of $x_{ll}(i, j, t)$, $x_{lh}(i, j, t)$, $x_{hl}(i, j, t)$ and $x_{hh}(i, j, t)$ in the wavelet decomposition. Then, the data in $y_{ll}^n(i, j, t)$ is rearranged into four subimages,

$$\begin{aligned} y_{ll1}^n(i, j, t) &= y_{ll}^n(2i, 2j, t) \\ y_{ll2}^n(i, j, t) &= y_{ll}^n(2i+1, 2j, t) \\ y_{ll3}^n(i, j, t) &= y_{ll}^n(2i, 2j+1, t) \\ y_{ll4}^n(i, j, t) &= y_{ll}^n(2i+1, 2j+1, t) \end{aligned} \quad (5)$$

This rearrangement is further applied on $y_{ll1}^n(i, j, t)$ and so on to obtain hierarchical representation of $y_{ll}^n(i, j, t)$, $y_{lh}^n(i, j, t)$, $y_{hl}^n(i, j, t)$ and $y_{hh}^n(i, j, t)$ to be coded by the 3-D SPIHT coder. 3-D SPIHT coding (without adaptive arithmetic code) is finally applied on the resulting hierarchical representation to generate symbol streams.

4. SIMULATION RESULTS

The performance of the proposed technique is introduced in this section. The proposed algorithm has been applied to the luminance component of Akiyo, Mthr_dotr, Grandma, salesman, MISS_AM and Claire image sequences with a QCIF format. 104 frames of these sequences are tested and 8 GOF is used. All results are achieved without motion compensation in both coders. In the simulation results the exact number of bits used is selected as follows:

For MISS_AM image sequences, 620000 bits in 3-D DCT coding, and 4000 bits in 3-D SPHIT coding is used so that the total bit rate per pixel (bpp) is 0.23674.

To achieve total bit rate per pixel (bpp) is 0.23674 for MTHR_DOTR, 576109 bits in 3-D DCT coding and 47891 bits in 3-D SPHIT coding is used.

The total bit rate per pixel of 0.23674 bpp for GRANDMA is achieved using 505347 bits in 3-D DCT coding, and 118653 bits in 3-D SPHIT coding.

Figures 1 and 2 show PSNR results for 3-D SPIHT and proposed method for 104 frames of the luminance of the sequences, at a bit rate of 0.23674. Table 1 shows the resulting mean PSNR values for 0.23674 bit rates and the reconstructed image sequences for the proposed method and the 3-D SPHIT coding are shown in Fig.3 and Fig.4, respectively. The results clearly show that the proposed coding performance is better than the 3-D SPIHT video coding performance (about 1.07- 5.88 dB) with the current videos used.

TABLE I. PSNR_{av} COMPARISON OF THE COMPRESSION METHODS

Video Sequence	3-D Spiht	Proposed method
MISS_AM	35.2253 (624000 bits)	38.2139 620000 bits in 3D -DCT coding, and 4000 bits in 3D- SPIHT is used
MTHR_DOTR	29.4302 (624000 bits)	30.9752 576109 bits in 3D -DCT coding, and 47891 bits in 3D- SPIHT is used
GRANDMA	26.8265 (624000 bits)	32.7134 505347 bits in 3D -DCT coding, and 118653 bits in 3D- SPIHT is used
AKIYO	29.5201 (624000 bits)	31.009 503549 bits in 3D -DCT coding, and 120451 bits in 3D- SPIHT is used
CLAIRE	32.2698 (624000 bits)	33.3451 620000 bits in 3D -DCT coding, and 4000 bits in 3D- SPIHT is used
SALESMAN	27.4049 (624000 bits)	30.0953 618154 bits in 3D -DCT coding, and 5846 bits in 3D- SPIHT is used

5. CONCLUSIONS

In this paper, a hybrid video coding algorithm is introduced. The proposed algorithm works much more efficiency than the 3-D SPIHT coding method. The new distribution of the wavelet coefficients provides a reduction in computational complexity because only two sorting pass is taken. The simulation results indicate that the PSNR performance of the proposed algorithm is much higher than that of 3-D SPIHT test coder. The coder is proposed without motion compensation, which makes it less complex than another motion compensated coders. As shown in [11], 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 proposed method. Moreover, using an arithmetic coding in the output stage can improve the PSNR. Like the 3-D SPIHT, the coder 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.

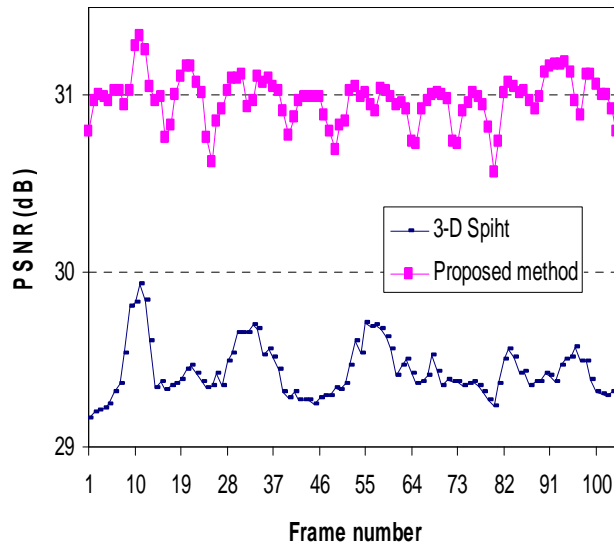


Figure 1. MTHR_DOTR sequence luminance, PSNR comparison of 3-D SPIHT and proposed method.

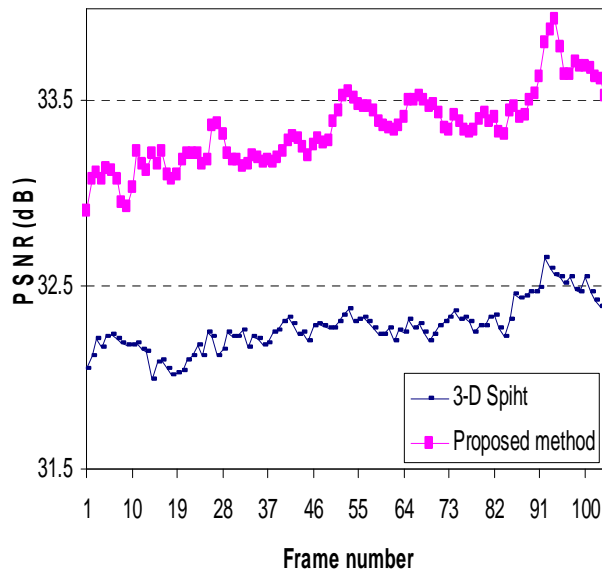


Figure 2. Claire sequence luminance, PSNR comparison of 3-D SPIHT and proposed method.



Figure 3. the reconstructed 1st, 40th and 60th frames of Akiyo, Mthr_dotr, and Grandma video sequence using the proposed algorithm.



Figure 4. the reconstructed 1st, 40th and 60th frames of Akiyo, Mthr_dotr, and Grandma video sequence using the 3-D SPIHT.

REFERENCES

- [1] K. Aizawa and T. Huang, "Model Based Image Coding: Advanced Video Coding Techniques for Very Low Bit-Rate Applications," In the proceedings of the IEEE, 83(2), pp.259-271, 1995.
- [2] J. Taylor, "Standards: DVD-video: multimedia for the masses," IEEE MultiMedia, 6(3), July–September 1999.

- [3] Implementation Guidelines for the use of MPEG-2 Systems, Video and Audio in Satellite, Cable and Terrestrial Broadcasting Applications, ETSI ETR 154, Revision 2, 2000.
- [4] L. William Butterworth, "Architecture of the first US direct broadcast satellite system," In the proceedings of the IEEE National Telesystems Conference, 1994.
- [5] Athar Ali Moinuddin, Ekram Khan and Mohammed Ghanbari, "Fast and Efficient 3-D Wavelet Based Video Coding Technique", Proceedings of the World Congress on Engineering 2008 London, U.K (WCE 2008), vol I, pp.978-988, 2008.
- [6] C. Zhang, K. Ugur, J. Lainema, M. Gabbouj, "Video Coding Using Variable Block-Size Spatially Varying Transforms", in the proceeding of ICASSP, pp.905-908, 2009.
- [7] G. Bernabé, J. M. García, J. González, "A lossy 3D wavelet transform for high-quality compression of medical video", The Journal of Systems and Software, vol.82, pp. 526–534, 2009.
- [8] A. Said and W. A. Pearlman, "A New Fast and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees", IEEE Transactions on Circuits and Systems for Video Technology, vol.6, pp. 243-250, June 1996.
- [9] B.J Kim and W. A. Pearlman, "An Embedded Wavelet Video Coder Using Three Dimensional Set Partitioning in Hierarchical Trees", in the proc. Of Data Compression Conf., pp.251-260, Mar. 1997.
- [10] T. Yu and S. K. Mitra, "A novel DPCM algorithm using a nonlinear operator," in the Proc. IEEE international conference on image processing'94, Austin, Texas, USA, pp.871-875, November 1994.
- [11] 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.

"تقنية عالية الأداء لتشفير أشارات الفيديو باستخدام طريقة دمج تقنيات التحويل"

الهدف الرئيسي من هذه المقالة هو تقديم طرق فعالة لتشفير أشارات الفيديو وذلك عن طريق دمج تقنية التحويل المويجي (DWT) مع تقنية التحويل الجيب تامى (DCT). ففي هذا المقال تم تقديم طريقه جديده لتشفير الفيديو باستخدام طريقة الدمج للحصول على نسبة ضغط عالية، حيث انها تجمع بين مزايا تقنية التحويل المويجي ثلاثى الأبعاد 3-D DWT مع مزايا تقنية التحويل الجيب تامى ثلاثى الأبعاد 3-D DCT، كما يمكننا القول أنه في حالة التغاضي عن الدقة الحسابية العالية في النتائج، فإنه بالإمكان ضغط الفيديو إلى نسبة ليست بالقليلة، وذات جدوى عظيمة من جهة طاقة التخزين، وكفاءة المعالجة. وقد تم عمل تعديل فى المعاملات الناتجة فى التحويل المويجي باستخدام تقنية التحويل الجيب تامى وأيضا تصنيف هذه المعاملات الى معاملات مهمة ومعاملات غير مهمة. وأثبتت النتائج أن هذه الطريقة تتفوق على مثيلاتها من الطرق التى نشرت فى هذا المجال.