3D Dual-Tree Discrete Wavelet Transform Based Multiple Description Video Coding

Jing Chen[1, 2], Canhui Cai[1] and Li Li[1]
[1] *School of Information Science and Engineering, Huaqiao University, Xiamen, China*
[2] *School of Information Science and Engineering, Xiamen University, Xiamen, China*

Keywords: Multiple Description Coding, Video Coding, Dual-Tree Discrete Wavelet Transform, Layer Coding.

Abstract: A 3D dual-tree discrete wavelet transform (DT-DWT) based multiple description video coding algorithm is proposed to combat the transmitting error or packet loss due to Internet or wireless network channel failure. Each description of the proposed multiple description coding scheme consists of a base layer and an enhancement layer. First, the input image sequence is encoded by a standard H.264 encoder in low bit rate to form the base layer, which is then duplicated to each description. Second, the difference between the reconstructed base layer and the input image sequence is encoded by a 3D dual-tree wavelet encoder to produce four coefficient trees. After noise-shaping, these four trees are partitioned into two groups, individually forming enhancement layers of two descriptions. Since the 3D DT-DWT equips 28 directional subbands, the enhancement layer can be coded without motion estimation. The plenty of directional selectivity of DT-DWT solves the mismatch problem and improves the coding efficiency. If all descriptions are available in the receiver, a high quality video can be reconstructed by a central decoder. If only one description is received, a side decoder can be used to reconstruct the source with acceptable quality. Simulation results have shown that the quality of reconstructed video by the proposed algorithm is superior to that by the state-of-the-art multiple description video coding methods.

1 INTRODUCTION

Due to the effect of transmission error, packet loss and over-delay problems caused by unreliable networks, the quality of video transmitted by Internet or wireless network is degraded, which restrains the development of multimedia services. As an effective error resilient coding technique, Multiple Description Coding (MDC) has received a lot of attention from researchers and has been applied in variety of scenarios to insure a robust video transmission (Goyal, 2001).

The MDC includes three steps. First, the source data is coded to form multiple self-decodable bit streams, named descriptions. Then, descriptions are transmitted over diverse channels to the receiver. In the receiver, there are two sorts of decoder, central decoder and side decoder, with which, video can be reconstructed by either high quality or acceptable quality according to the number of correctly received descriptions. When a network transmission error occurs so that only one description is received, the received description is decoded by the side decoder. The quality is incrementally improved with more received descriptions. When all descriptions are received correctly, the highest fidelity of reconstruction can be achieved by the central decoder.

Since the first MDC image coding scheme, multiple description scalar quantizer (MDSQ), published by Vaishampayan in 1993 (Vaishampayan, 1993), many other MDC methods has been developed, including multiple description scalar quantizing coding, multiple description transform coding (MDTC), multiple description subband coding (MD-SPIHT), polyphase transform and selected quantization multiple description coding (PTSQ), Unequal protected MDC and layered based MDC, etc.

Researches on multiple description video coding (MDVC) emerged in 1999. The pioneer ones, such as MDSQ based video coding (Vaishampayan, 1999) and MDTC based multiple description video coding (Reibman, 2002), were direct extension of the multiple description image coding method. The problem with these methods is the reference

Chen J., Cai C. and Li L..
3D Dual-Tree Discrete Wavelet Transform Based Multiple Description Video Coding.
DOI: 10.5220/0005057701800185
In *Proceedings of the 11th International Conference on Signal Processing and Multimedia Applications* (SIGMAP-2014), pages 180-185
ISBN: 978-989-758-046-8

mismatch between the transmitting side and the receiving side, introduced by the errors of unreliable network transmission. How to solve the mismatch problem is the primary task for multiple description video coding. To eliminate the mismatch, motion estimation and compensation were individually performed in each description (Tillier, 2007), which lowered the coding efficiency and increased the computational complexity.

As the emergence of H.264 coding standard, several H.264 based multiple description video coding method were proposed to improve the error resilient ability of H.264. Bernadini and Durigon proposed a pholyphase spatial subsampling multiple description coding (PSS-MDC) (Bernardini, 2004), which separated the source video into four subsequences by spatial subsampling, and then coded each subsequence by an H.264 encoder to form four bitstreams, which were transmitted through diverse channels. Even three of them were lost, the method was able to insure an acceptable reconstructed quality of the input video. Based on PSS-MDC, Wei et. al. paired subsequences to form two descriptions and proposed a neighboring pixel prediction algorithm to reduce the redundancy of PSS-MDC (Wei, 2006). Campana et. al. innovated the MDSQ by mapping more zero coefficients to improve the coding efficiency (Campana, 2008). A slice optimal allocation for H.264 multiple description video coding was proposed by Tillo (Tillo, 2008). However, it was not suitable for low bitrate transmission.

Stimulated by the thought of layered based MDC scheme, we proposed an H.264 and dual-tree discrete wavelet transform based multiple description video coding. The base layer is produced by feeding input into an H.264 encoder with low bitrate and copied into each description. The enhancement layer is formed by four trees of wavelet transform coefficients outputed by the dual-tree discrete wavelet transform. These coefficient trees are paired into two sets and sent them into separate description. Each description comprises a base layer and an enhancement layer, which is transmitted through diverse channel. The simulation results have shown the efficiency and the error resillient ability of the proposed method.

The rest of this paper is organized as follows. H.264 and dual-tree discrete wavelet transform is introduced in Section 2. Section 3 concentrates on how to generate descriptions of the proposed multiple description video coding scheme. Simulation results and analysis are illustrated in Section 4. Concluding remarks are given in Section 5.

2 DUAL-TREE DISCRETE WAVELET TRANSFORM

2.1 Dual-tree Discrete Wavelet Transform (DT-DWT)

To improve the directional selection and shift-invariant property of the traditional discrete wavelet transform (DWT), Nick Kingsbury proposed the dual-tree complex wavelet transform (DT-CWT) in 1998. Its directional subband decomposition make DT-CWT nearly shift-invariant and higher directional selectivity. However, the DT-CWT is an over-complete transform with plenty of redundancy (2^n:1 for n-dimensional signal). By analysis of the real part and imaginary part of wavelet coefficients, Selesnick found these two parts have the same directional selectivity and either one could serve as a wavelet transform to halve the number of coefficients. In this light, Selesnick (Selesnick, 2005) proposed the dual-tree discrete wavelet transform (DT-DWT). Figure 1 shows the six high frequency directional subbands of 2D DT-DWT.

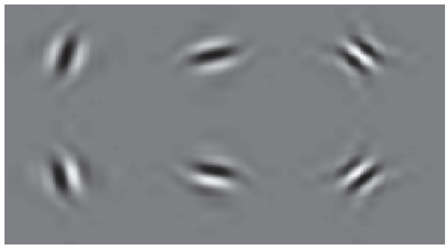


Figure 1: The directional subbands of 2D DT-DWT.

2.2 Implementation of 3D Dual-tree Discrete Wavelet Transform

3D dual-tree discrete wavelet transform can be separately performed by four 3D discrete wavelet transforms (Wang, 2007). Figure 2 shows one of the transform, where $*$ denotes the convolution operation, $\downarrow 2$ is down sampling by 2; x, y and z represent three directions of axis, horizontal, vertical and time, respectively; h_0 and h_1 are low-pass filter and high-pass filter, which form a Hilbert transformed pair, and insure the perfect reconstruction of the discrete wavelet transform. For convenient, $*x\, h_0$ is used to denote the convolution of video source with low-pass filter h_0; and $\downarrow_x 2$ to show the down sampling process in horizontal direction; Many symbols are likewise defined in Figure 2. As shown in Figure 2, there are one low frequency subband (LLL) and seven high frequency subbands (LLH，LHL，LHH，HLL，HLH，HHL，HHH) in one 3D DWT. Since there are four 3D DWT in one 3D DT-DWT structure, 28 high frequency subbands guarantee variety of directional selectivity of 3D DT-DWT, as shown in Figure 3.

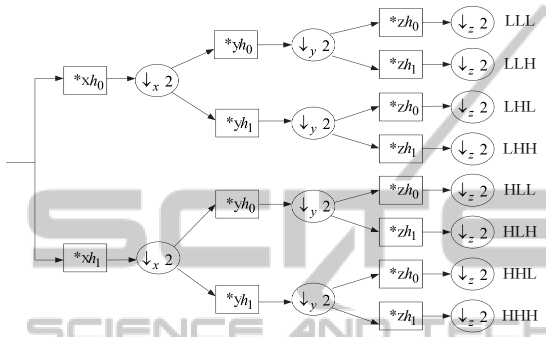


Figure 2: The filter band of 3D DT-DWT.

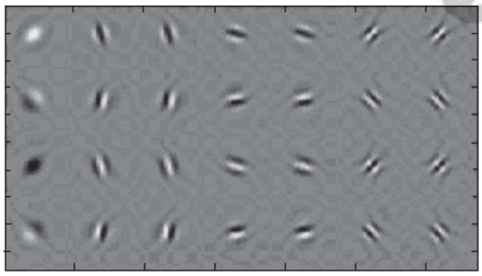


Figure 3: The 28 directional high frequency subbands of 3D DDWT.

2.3 The Sparse Representation of Coefficients of 3D DT-DWT

As mentioned above, the number of 3D DT-DWT coefficients is four times as much as that of the traditional 3D DWT. Based on the observation that most 3D DT-DWT coefficients with small values have little contribution on reconstructed images, Reeves and Kingsbury (Reeves, 2002) proposed a noise shaping method to thin out the coefficients. The so-called noise shaping in fact is an iterative projection between the image domain and the wavelet transform domain, during which only significant coefficients are picked up. And then they are modified to compensate the error incurred due to the loss of insignificant coefficients. The rationality and effectiveness of the sparse representation resulted from noise shaping can be found in some video coding schemes (Wang, 2007; Li, 2009).

3 3D DT-DWT BASED LAYERD MULTIPLE DESCRIPTION VIDEO CODING

3.1 The Proposed Layered Multiple Description Video Coding Scheme

It is well known that the layered structure of video coding can provide a convenient way to change its bit rate for different bandwidth requisition of heterogeneous clients and for dynamic network congestion reduction. Consequently, the layered MDC method can take the advantage of the layer structure by providing common information and private information for each description to form the base layer and the enhancement layer respectively. To avoid the mismatch, the inter-frame prediction in the enhancement layer is undesired.

Figure 4 shows the proposed layered MDC scheme, where the base layer of each description is encoded by an H.264 encoder, the residual between source video and the one reconstructed from base layer is encoded by a 3D DT-DWT encoder to form the enhancement layer. Since the 3D DT-DWT equips 28 directional subbands, the enhancement layer coding can be done without motion estimation. In this way, the mismatch is eliminated and the computational complexity of the enhancement layer coding is reduced. Moreover, the proposed multiple description video coding scheme can keep higher coding efficiency.

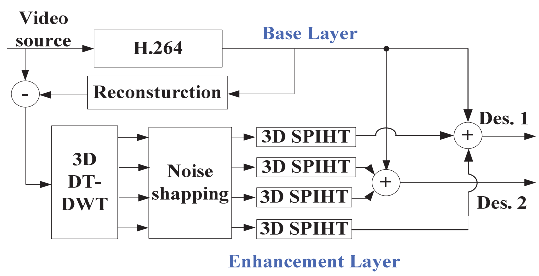


Figure 4: The 3D DT-DWT layered MDC.

3.2 The Base Layer Coding

Since the base layer contains the common information of all descriptions, to reduce the redundancy of the MDC scheme, H.264 video coding is adopted to provide the high coding efficiency. The coded bitstream is then copied to both descriptions, to ensure the primary information of video. The features of H.264, such as multi-reference frame mode and 1/4 pixel precision of motion estimation, guarantee the accuracy of its inter-frame prediction. The precise intra-frame prediction coding further improves its coding efficiency.

3.3 The Enhancement Layer Coding

The study of 3D DT-DWT coefficient trees (Figure 3) shows the high directional correlation between tree 1 and tree 3, tree 2 and tree 4, respectively. Combining tree 1 and tree 4 to form the enhancement layer of description 1, tree 2 and tree 3 to form the enhancement layer of description 2, even one of descriptions is lost, the other one can effectively provide the directional information.

Based on the above observation, the residual of souce video and reconstructed one from the base layer is encoded by a 3D DT-DWT encoder to form four wavelet coefficient trees. The noise shaping then applied to reduce the redundancy of 3D DT-DWT. The resulted four 3D wavelet trees are coded by 3D SPIHT (Kim, 2000) to form the enhancement layers. These four tree bistreams are grouped to form enhancement layers of two descriptons: tree 1 and tree 4 are allotted to description 1, tree 2 and tree 4 are allotted to description 2. Since the directional correlation of two descriptions conveys some information of each other, even one description losts, the other one can provide an acceptable quality reconstruction.

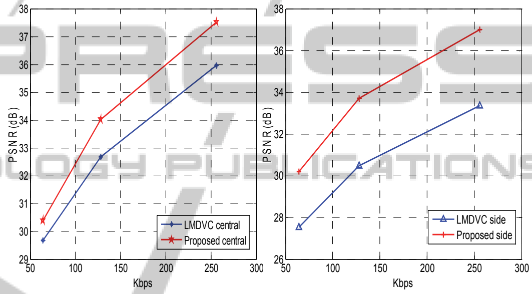
3.4 Decoding Algorithm

If all descriptions are correctly received, the source video can be perfectly decoded from the base layer and enhancement layers of both descriptions by using the central decoder. If only one description is received, the side decoder is used instead. Firstly, the base layer is reconstructed, and then two wavelet trees in the correctly received description are used to reconstruct two received wavelet coefficient trees and two missed ones by exploiting the directional correlation of four wavelet coefficient trees. The base layer ensures the elementary quality of reconstructed video, and the received enhancement layer can effectively improve the quality of reconstructed image based on the received residual and directional information.

4 EXPERIMENTAL RESULTS AND ANALYSIS

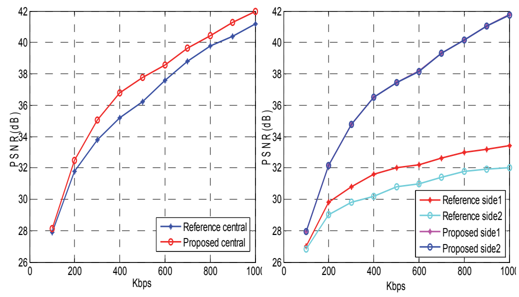
To verify the proposed method, lots of experiments are conducted using different motion types of video sequences (slow, moderate, fast) with different formats (cif, qcif) and different frame rates.

The first set of experiments is to compare the proposed method with other wavelet transform based multiple description video codecs. As shown in Figure 5 and Figure 6, the proposed method outperforms the LMDVC (Chen, 2008) and the one by Tiller (Tillier, 2007) in both central decoder and side decoder with different bitrates, especially in the high bitrate.



(a) central decoder (b) side decoder

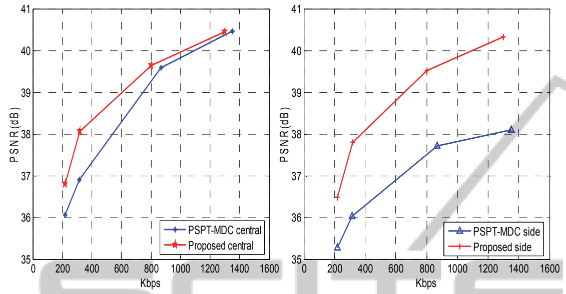
Figure 5: The comparison of the proposed method with LMDVC (Chen, 2008) on Foreman (qcif, 15fps).



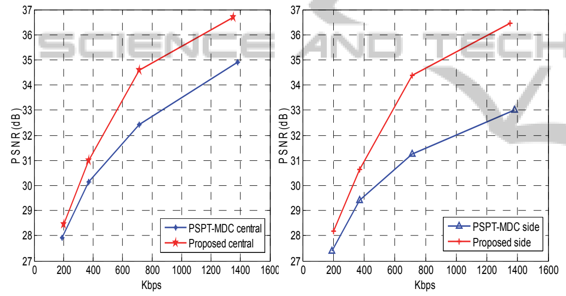
(a)central decoder (b) side decoder

Figure 6: The comparison of the proposed method with Tillier's (Tillier, 2007) on Foreman (qcif, 30fps).

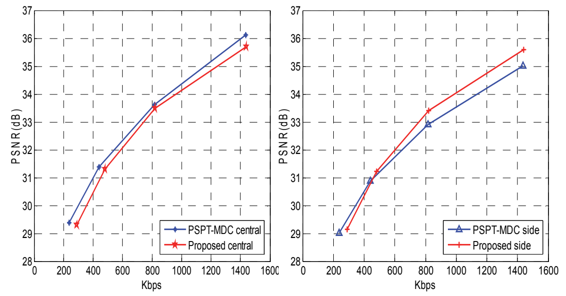
The second set of experiments is to compare the proposed method with other H.264 based multiple description video codecs. Figure 7 shows experimental results on different motion activity types of video sequences. On slow (Miss-america) and moderate motion (Foreman) video sequences, shown in Figure 7(a) and Figure 7(b), the proposed method outperforms the PSPT-MDC (Wei, 2006), especially when only one description is available. While tested on high speed motion (Football) video sequences, shown in Figure 7(c), 28 directions of 3D DT-DWT subbands are still not good enough to represent the scene, which lowered the performance of central decoder compared with the H.264 based MDC. However, when the bitrate is higher than 450K/s, the side reconstruction quality is still better than PSPT-MDC, which proves the error resilient ability of the proposed method.



central decoder side decoder
(a) Miss-america (slow motion, cif, 30fps)



central decoder side decoder
(b) Foreman (moderate motion, cif, 30fps)



central decoder side decoder
(c) Football (fast motion, cif, 30fps)

Figure 7: The comparison of proposed method with PSPT-MDC (Wei, 2006) on different sequences.

5 CONCLUSIONS

The proposed method takes the advantages of both multiple description coding and the layer coding. To form multiple description, the primary information of video is encoded with H.264 and copied into both descriptions to ensure the fundamental quality of the reconstruction. The directional correlation between descriptions is provided by 3D DT-DWT after noise shaping and form the enhancement layer of two descriptions. The high directional selectivity of 3D DT-DWT saves the bitrate of motion estimation, and reduces the computational complexity. The experimental results have shown the effectiveness of the proposed method.

ACKNOWLEDGEMENTS

This work is partially supported by the National Natural Science Foundation of China (No. 61372107) and the Xiamen Key Science and Technology Project Foundation under the Grant 3502Z20133024.

REFERENCES

V.K. Goyal, 2001. Multiple description coding: Compression meets the network. *IEEE Signal Processing Magazine*.

V.A. Vaishampayan, 1993. Design of multiple description scalar quantizers. *IEEE Transactions on Information Theory*.

V.A. Vaishampayan, S. John, 1999. Balanced interframe multiple description video compression. In *Proceeding of IEEE International Conference on Image Processing (ICIP)*.

R.A. Reibman, H. Jafarkhani, Y. Wang, M.T. Orchard, and R. Puri, 2002. Multiple description video coding using motion compensated temporal prediction. *IEEE Transactions on Circuits and System for Video Technology*.

C. Tillier, T. Petrisor, B. Pesquet-Popescu, 2007. A motion-compensated overcomplete temporal decomposition for multiple description scalable video coding. *EURASIP Journal on Image and Video Processing*.

R. Bernardini, M. Durigon, R. Rinaldo, L. Celetto, and A. Vitali, 2004. Polyphase spatial subsampling multiple description coding of video streams with h.264. In *Proceeding of IEEE International Conference on Image Processing (ICIP)*.

Z. Wei, C. Cai, K. Ma, 2006. H.264-based multiple description video coder and its DSP implementation. In *Proceeding of IEEE International Conference on Image Processing (ICIP)*.

O. Campana, R. Contiero, and G. A. Mian, 2008. An H.264/AVC video coder based on a multiple description scalar quantizer. *IEEE Transactions on Circuits and Systems for Video Technology*.

T. Tillo, M. Grangetto, and G. Olmo, 2008. Redundant slice optimal allocation for H.264 multiple description

coding, *IEEE Transactions on Circuits and Systems for Video Technology*.
I.W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury, 2005. The dual-tree complex wavelet transform. *IEEE Signal Processing Magazine*.
B. Wang, Y. Wang, I. Selesnick and A. Vetro, 2007. Video coding using 3D dual-tree wavelet transform. *EURASIP Journal on Image and Video Processing*.
T.H. Reeves, N.G. Kingsbury, 2002. Overcomplete image coding using iterative projection-based noise shaping. In *Proceeding of IEEE International. Conference on Image Proceeding (ICIP)*.
L. Li, C. Cai, 2009. Multiple description image coding using dual-tree discrete wavelet transform. In *proceeding of International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*.
J. Chen, C. Cai, 2008 . Layered multiple description video coding, In *Proceeding of IEEE International Conference on Signal Processing (ICSP)*.
B.J. Kim, Z. Xiong, and W.A. Pearlman, 2000. Low bit-rate scalable video coding with 3-D set partitioning in hierarchical trees (3-D SPIHT). *IEEE Transactions on Circuits and Systems for Video Technology*.