WAVELET BASED DISTRIBUTED VIDEO CODING WITH SPATIAL SCALABILITY

Bo Wu[1], Xiangyang Ji[1]
[1]Institute of Computing Technology
Chinese Academy of Sciences
Beijing 100080, China

Debin Zhao[2], Wen Gao[3]
[2]School of Computer Sciences
Harbin Institute of Technology
Harbin, 150001, China
[3]Peking University
100872, China

Abstract—**In this paper, an improved spatial scalable wavelet domain distributed video coding (DVC) scheme is proposed. In this scheme, we adaptively apply differential pulse code modulation (DPCM) Wyner-Ziv (WZ) coding and discrete cosine transform (DCT) based intra block coding to reduced-resolution layer. Due to the low energy nature of high-pass subbands, WZ coding can be directly used. At decoder, in full-resolution layer WZ frame decoding, an efficient side information (SI) generation technique for low-pass subband is proposed, in which motion compensated interpolation (MCI) is performed on full-resolution signal. On the other hand, a refined technique is also employed to generate the SI for high-pass subbands by taking advantage of inter-band correlation. Experimental results show that the proposed scheme provides an efficient wavelet domain DVC coding paradigm.**

I. INTRODUCTION

Recently, the wireless communications of video signals over the wireless sensor network is emerging. This application scenario requires a light encoder in which the computation and memory resources on sensors are scarce. Wyner-Ziv Coding [1] allows implementing low-complexity encoders. The new video compression paradigm DVC based on Wyner-Ziv theory has been proposed in [3, 4]. The discrete wavelet transform (DWT) has been widely used in DVC paradigm due to its advantages of multi-resolution analysis and spatial decorrelation. In [7], the author proposed a linear prediction model to exploit the inter-band correlation of DWT coefficients of the hyperspectral imagery. In [8], the authors use the RWDT reference frames for finding matching blocks which satisfy CRCs to overcome the shift-variant of DWT domain. In [9], the author proposed a context correlation model between the source and its SI in the wavelet transform domain. However, the DWT is not a shift invariant operator. Compared to RDWT, spatial domain motion compensation is usually able to yield the better prediction efficiency for side information [2].

In this paper, we propose an efficient wavelet domain DVC scheme. In the scheme, the original video is decomposed into the reduced-resolution layer and the high-pass subbands by DWT. The reduced-resolution layer is coded by an adaptive DPCM WZ coding and intra block coding strategy. Considering the high efficiency and low complexity of DPCM, the Wyner-Ziv video coding scheme based on DPCM is proposed in [5]. As the temporal distance is enlarged or the motion is nonlinear, the temporal correlation between adjacent frames becomes weak. As a result, the correlation between the original and SI also becomes weak and thus, the WZ coding may be less efficient than DCT based intra block coding. Therefore, the intra mode is introduced for low resolution video coding for those irregular regions. For high-pass subband encoding, the PCM WZ coding are applied. The zero tree based coefficients organization are used as in [6]. At decoder, if the high-pass subbands decoding are not available, the reduced-resolution layer WZ decoding is performed. The SI for reduced-resolution layer is interpolated by reconstructed reduced-resolution frames. Otherwise, if the high-pass subband decoding is preferable, the full-resolution frames can be obtained by full-resolution WZ decoding procedure. At first, the low-pass subband is decoded. The MCI is performed on full-resolution key frames and the low-pass band of generated full-resolution prediction frame is used as the SI of low-pass subband. Then, with the help of reduced-resolution layer, the SI for high-pass subbands is motion compensated by using the motion information of decoded reduced-resolution layer. Both the temporal correlation among the already decode frames in different resolution layers and the in-band correlation among DWT subbands level-by-level are exploited.

The remainder of this paper is organized as follows: Section II describes the proposed scheme in detail. Section III gives experiment results and analysis. And conclusions are drawn in Section IV.

II. WAVELET DOMAIN DVC CODING SCHEME

2.1 Wavelet domain DVC encoding

As shown in Fig.1 to support one-level spatial scalability, every image of the input video is firstly decomposed into one low-low pass (LL) subband for reduced-resolution layer and three high-pass subbands. The low-pass subband and the high-pass subbands of key frames are all encoded with DCT based intra block coding. Here, we use H.263+ intra mode coding. The WZ frame's coding procedure is described as follows:

2.1.1 Low-pass subband WZ coding

The low-pass subband of WZ frame is encoded by an adaptive DPCM WZ coding and intra block coding. The coding scheme is shown as Fig.1 (a). In the scheme, the coefficients of low-pass subband are partitioned into blocks. Then, each block is adaptively classified into intra mode or DPCM mode. The DPCM residue signal is the difference between the blocks in the low-pass subband of current frame and the co-located blocks in the low-pass subband of reference frame which is stored in the delay buffer. The mode classification is based on the comparison of mean-squared error (MSE) of the DPCM residue and the variance of current block. The mode selection can be described as:

$$\text{mode} = \begin{cases} \text{intra} & if: VAR(X_{intra}) < \alpha MSE(X_d) \\ \text{DPCM} & else \end{cases} \quad (1)$$

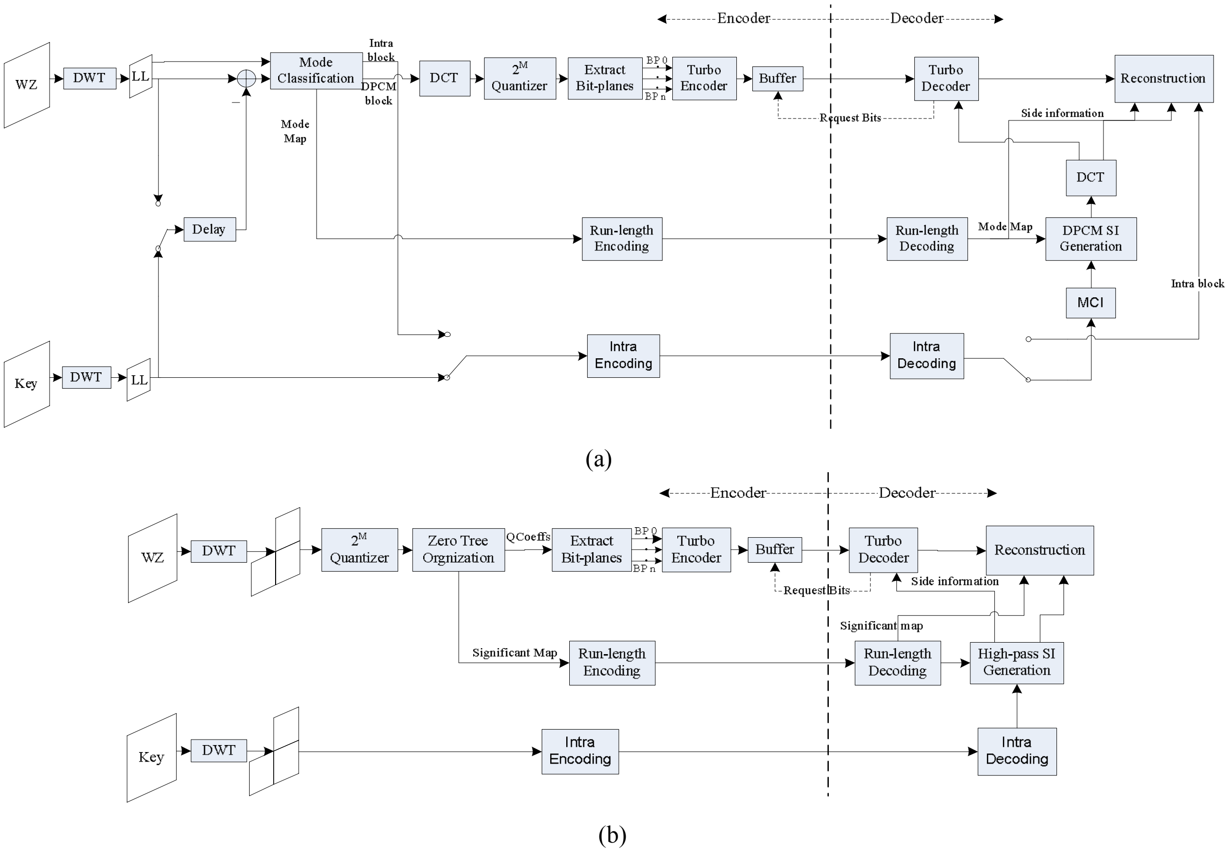


Fig. 1 (a) Diagram of low-pass subband WZ coding (b) Diagram of high-pass subbands WZ coding

where X_d is the DPCM residue and α is a constant number. The reference frame could be its previously reconstructed frames, backward reconstructed frames, or the average of two adjacent reconstructed frames. After the mode classification, a map which identifies the coding mode of each block is generated and is encoded by run-length coding. The DPCM blocks are WZ encoded and the procedure are the same with the scheme proposed in [3]. Firstly, the DPCM residue of each block is transformed with 4x4 DCT. Then, the coefficients are quantized and extracted into bit-planes. At last, all bit-planes are sequentially encoded by Turbo encoder.

2.1.2 High-pass subbands WZ coding

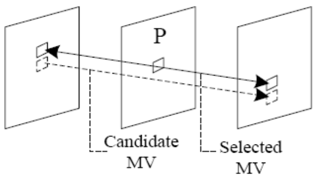
To exploit the high-order statistical correlation in wavelet domain, we use zero-tree to reorder the high-pass coefficients. The quantization and organization of coefficients are similar to [6]. The coding procedure is shown as Fig.1 (b). Firstly, for different subbands, different M level uniform scalar quantizers are applied. Secondly, the quantized coefficients are reordered to zerotree structure and the significance map is obtained. At last, the significance map is run-length encoded and the bitplanes of quantized high-pass subbands are sequentially encoded by Turbo encoder.

2.2. Wavelet domain DVC decoding

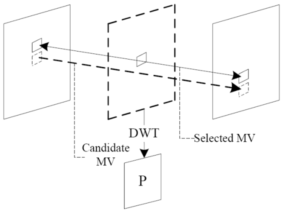
The key frames are decoded by intra frame decoding. There are two decoding strategies for WZ frames at reduced-resolution and full-resolution respectively.

2.2.1 Reduced-resolution WZ Decoding

The WZ decoding of reduced-resolution frames is depicted as Fig.1 (a). Firstly, a prediction frame Y for reduced-resolution is generated by MCI. As shown in Fig.2 (a), if the decoder can not access the high-pass subbands, the prediction Y at time *t* is interpolated by the decoded reduced-resolution frames at time *t*-1 and *t*+1. The MCI is the same as that proposed in [5]. Secondly, the mode map of current WZ frames is decoded by run-length decoder. According to the mode map, the DPCM part Y_D of Y is extracted and the DPCM residues Y_d can be obtained. $Y_d = Y_D - \hat{X}_{ref}$, where $\hat{X}_{ref}$ is the reference frame. Furthermore, DPCM residuals Y_d are DCT transformed and serve as SI in DCT domain. With DCT domain WZ decoding strategy proposed in [3], the DPCM residuals are reconstructed. At last, the conventional decoded intra blocks and reconstructed DPCM parts are combined in accordance with the mode map to reconstruct the reduced-resolution layer of WZ frames.



(a)



(b)

Fig.2 (a) SI generation for reduced-resolution layer WZ decoding (b) SI generation for full-resolution layer WZ decoding

2.2.2 Full-resolution WZ Decoding

1). Low-pass subbands decoding

The WZ decoding procedure of low-pass subbands of full-resolution is similar to the previous reduced-resolution WZ decoding. The only difference is the manner of SI generation. The generation of prediction frame for low-pass subband is shown in Fig.2 (b). Firstly, a full-resolution prediction frame Y_f is interpolated by the reconstructed full-resolution key frames. Secondly, DWT is applied to Y_f and the low-pass subband of Y_f is used as the reduced-resolution prediction frames Y_r. Experimental results show that Y_r is better than Y which is interpolated by reduced-resolution reference frames. This improvement attributes to that the full-resolution reconstruction frames comprise higher frequency information. By using the information from higher frequency spatial-subbands, the type II leakage exists in reduced-resolution layer can be compensated [2].

2). High-pass subbands decoding

The high-pass subbands WZ decoding is described as Fig.1 (b). The significance map is decoded firstly. Then high-frequency subbands of SI are reordered to a zero tree structure according to the significant map. Finally, the high-pass subbands are WZ decoded.

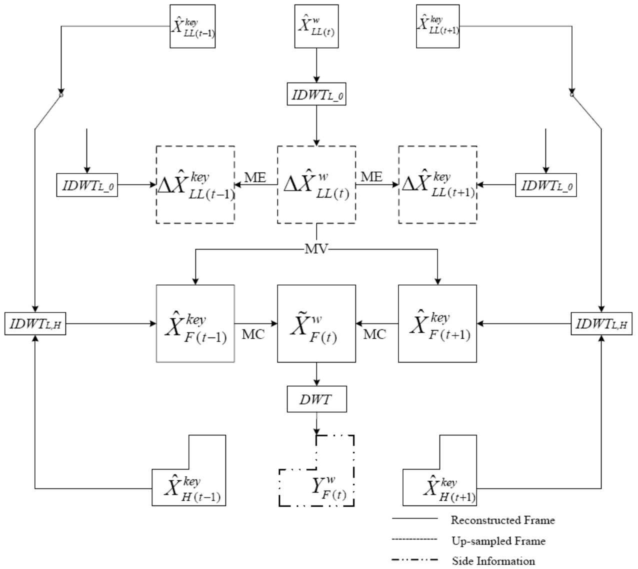


Fig.3 Refined SI generation for high-pass subbands

A refined SI generation scheme is proposed for high-pass subbands WZ decoding. As we know, the motion information of the reduced-resolution layer and full-resolution layer are highly correlated. Moreover, there are inter-band correlations among DWT subbands. In the scheme, the inter-band correlations of current frame are exploited by using the temporal correlations of upsampled reduced-resolution layers. The refined SI generation scheme for two levels DWT decomposition is illustrated as Fig.3. The low-low pass subband (LL) is the reduced-resolution layer. Details are described as follows:

Step 1: Up-sample the decoded reduced-resolution frames.

$IDWT_{L\_0}$ operator is used to up-sample the decoded LL frames and it is defined as follows,

$$\Delta\hat{X}_{LL(t)} = IDWT_{L\_0}\{\hat{X}_{LL(t)}, 0\} \tag{2}$$

where $\hat{X}_{LL(t)}$ is the decoded LL frame at time t and $\Delta\hat{X}_{LL(t)}$ is the up-sampled LL frame. $IDWT_{L\_0}$ operator is an inverse DWT in which the LL subband is $\hat{X}_{LL(t)}$ and the high-pass sub-bands are set to zeros.

Step 2: Perform motion estimation.

The motion estimation is performed between the up-sampled WZ frame $\Delta\hat{X}^{w}_{LL(t)}$ and its references, the up-sampled LL band of key frame $\Delta\hat{X}^{key}_{LL(t)}$. The motion vectors are calculated by minimizing

$$SAD = \min_{mv}\{\sum_{p\in S, i=1,-1} |\Delta X^{w}_{LL(t)}(p) - \Delta X^{key}_{LL(t+i)}(p+mv)|\} \tag{3}$$

where p denotes the pixel position and it belongs to the sample position set S of current block. $\Delta X^{w}_{LL(t)}(p)$ represents the sample intensity at position p in $\Delta\hat{X}^{w}_{LL(t)}$. Here only up-sampled key frames at time instant (t-1) and (t+1) are used as forward and backward references.

Step 3: Perform motion compensation.

As the MVs of two adjacent spatial resolution layers are highly correlated, the MVs of upsampled reduced-resolution layer are used as the MVs of full-resolution layer. The motion compensation is performed on the reconstructed full-resolution key frames $\hat{X}^{key}_{F(t)}$ by using the MVs generated in step2.

$$\tilde{X}^{w}_{F(t)}(p) = \hat{X}^{key}_{F(t+i)}(p+mv) \qquad i = 1, -1 \tag{4}$$

The forward and backward reconstructed full-resolution key frames are adopted as reference frames. The reconstructed full-resolution key frames $\hat{X}^{key}_{F(t)}$ comprise both the low-pass and the high-pass information. It is reconstructed by

$$\hat{X}^{key}_{F(t)} = IDWT_{L,H}(\hat{X}^{key}_{LL(t)}, \hat{X}^{key}_{H(t)}) \tag{5}$$

where $\hat{X}^{key}_{LL(t)}$ and $\hat{X}^{key}_{H(t)}$ are reconstructed LL subband and high-pass subband of key frames respectively. Because of the inter-band correlation of DWT transformed coefficients, the high-pass subbands of current WZ frame are also obtained through the motion compensation procedure besides the low-pass subband. Consequently, a full resolution prediction of current WZ frame $\tilde{X}^{w}_{F(t)}$ is generated by (4).

Step 4: Decompose the full-resolution prediction frame to get SI.

$\tilde{X}^{w}_{F(t)}$ is decomposed by DWT and the high-pass subbands of $DWT\{\tilde{X}^{w}_{F(t)}\}$ are used as the SI $Y^{SI}_{H(t)}$ to decode the high-pass subbands of WZ frame. Simulation result shows that this scheme has better performance than MCI

III. EXPERIMENTAL RESULTS

In order to verify the coding efficiency of proposed wavelet domain DVC, we implement the above scheme. Foreman and Tempete sequences (CIF) are used in the test. In each sequence, 100 frames are selected and the coding structure is I-W-I-W. The DWT was implemented with biorthogonal 9/7 filter up to 2 decomposition levels. H.263+ is applied to intra frame coding parts. Rate-compatible punctured turbo codes (RCPT) are adopted as Slepian-Wolf codec.

Fig. 4 shows the rate-distortion (RD) performance comparison of two WZ coding methods for the reduced-resolution layer: one is the adaptive DPCM WZ coding and intra mode coding; the other is DCT domain WZ coding. Because we keep the RD performance of the key frames in these two schemes the same, the RD of luminance components of WZ frames are only included and the frame rate of WZ frames is 15 Hz. We can conclude that the DPCM and intra mode based WZ coding method is more efficient than the DCT based WZ coding especially when irregular motions exist in the video sequence. Fig.4 shows the proposed our scheme achieves a PSNR increment of up to 1.6dB.

In Fig. 5 the RD results of DCT domain and DWT domain full-resolution WZ coding are compared. For the DCT domain WZ coding, the scheme proposed in [3] is used for comparison. The SI for DCT domain WZ coding is interpolated by adjacent full-resolution reference frames. It is evident that the wavelet domain scheme has better performance especially for high end. The improved performance is from the effectively adaptive DPCM WZ coding and intra mode coding of reduced-resolution layer and the improved SI for both of low-pass subband and high-pass subbands.

IV. CONCLUSION

In this paper, we have presented an improved DWT domain distributed video coding scheme which support spatial scalability. The reduced-resolution layer is encoded by an adaptive DPCM WZ and Intra mode coding scheme. The coefficients of high-pass subband are orgnized by zerotree and coded by PCM WZ encoder. In full-resolution WZ decoding, an aliasing effect compensated SI generation method is adopted for reduced-resolution layer. Furthermore, a refined SI generation scheme for high-pass subband decoding is proposed. The proposed scheme reduces the MCI aliasing effects in wavelet domain and the performance of the proposed scheme is promising compared to the DCT domain WZ coding while with better spatial scalability support.

V. ACKNOWLEDGEMENTS

This work was supported by National Science Foundation of China (60672088, 60736043).

REFERENCES

[1] A. D. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” IEEE Transactions on Information Theory, vol. 22, pp.1-10, Jan.1976.

[2] Mehrseresht, N. and Taubman, D. “An efficient content-adaptive motion compensated 3D-DWT with enhanced spatial and temporal scalability,” IEEE Transactions on Image Processing, vol. 15, pp.740-753, Mar.1976.

[3] A. Aaron, S. Rane, E. Setton and B. Girod, "Transform-domain Wyner-Ziv codec for video", Proc. Visual Communications and Image Processing, VCIP-2004, San Jose, CA, Jan. 2004.

[4] R. Puri and K. Ramchandran, “PRISM: a new robust video coding architechture based distributed compression principles,” in Allerton Conference on Communication, Control, and Computing, Oct. 2002.

[5] A. Aaron, D. Varodayan and B. Girod, "Wyner-Ziv residual coding of video," Proc. Picture Coding Symposium, PCS-2006, Beijing, China, April 2006.

[6] Xun Guo, Yan Lu, Feng Wu, Wen Gao, "Distributed Video Coding Using Wavelet," IEEE International Symposium on Circuits and Systems (ISCAS’2006), Island of Kos, Greece, pp5427-5430, May21-24, 2006.

[7] C. Tang, Ngai-Man Cheung, Antonio Ortega, C. S. Raghavendra, “Efficient inter-band prediction and wavelet based compression for hyperspectral imagery: A distributed source coding approach,” in Proc. IEEE Data Compression Conference (DCC) 2005.

[8] J. E. Fowler, M. Tagliasacchi, B. Pesquet-Popescu, “Wavelet-Based Distributed Source Coding of Video,” in Proc. the European Signal Processing Conference, Antalya, Turkey, Sept 2005.

[9] Marco Grangetto Enrico Magli Gabriella Olmo, “Context-Based Distributed Wavelet Video Coding,” in Proc. IEEE 7th Workshop on Multimedia Signal Processing, October, 2005.

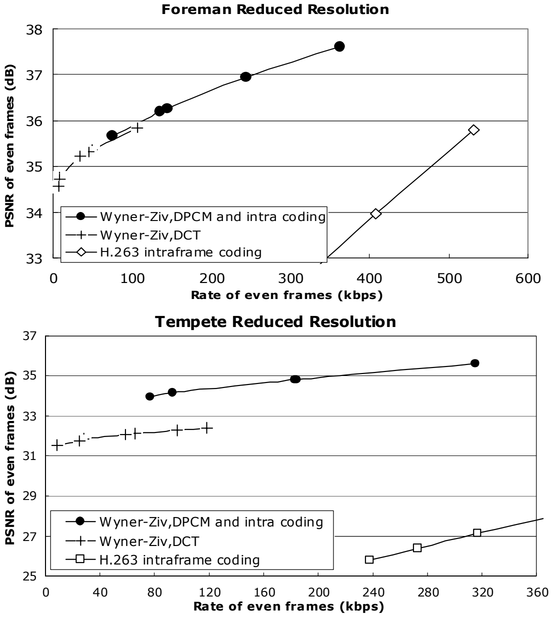


Fig. 4 Simulation result of reduced-resolution DWT domain Wyner-Ziv coding

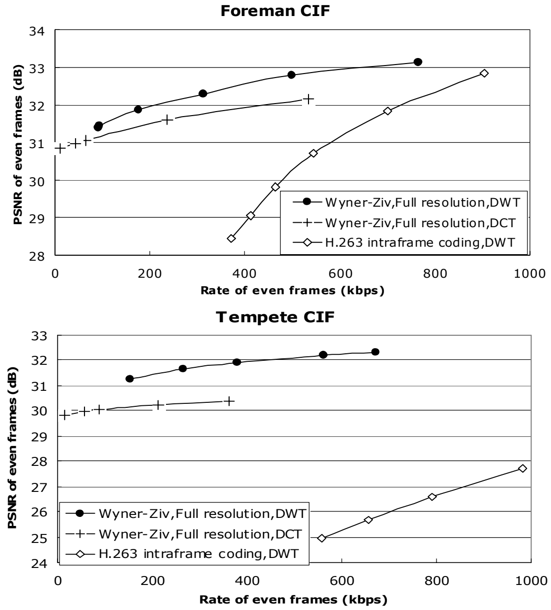


Fig.5 Simulation results for full-resolutions DWT domain Wyner-Ziv coding