Proceedings of 2009 APSIPA Annual Summit and Conference, Sapporo, Japan, October 4-7, 2009

Cross-Layer Error Resilient Mechanism in Scalable Video Coding

Shu-Jhen Fan-Jiang[1], Chia-Hung Yeh[1] and Mei-Juan Chen[2]
[1]Department of Electrical Engineering, National Sun Yat-Sen University, 804 Taiwan, R.O.C.
[2]Department of Electrical Engineering, National Dong-Hua University, 974 Taiwan, R.O.C.
E-Mail: {m9621056@ems.ndhu.edu.tw,yeh@mal.ee.nsysu.edu.tw,cmj@mail.ndhu.edu.tw}

Abstract— **Scalable video coding (SVC), which is the scalable extension of H.264/AVC, is designed to transmission over heterogeneous networks and provide scalabilities and high compression ratio. However, such kinds of compressed videos are sensitive to transmission error, resulting in quality degradation seriously. In this paper, a cross-layer error resilient method is proposed to prevent transmission error in SVC. In the proposed algorithm, the reversible data embedding technique is adopted to hide the essential information of lower layer such as IntraBL and the predicted motion vector (MV) for the reconstruction of higher layer. Experimental results show that the proposed method can provide better PSNR performance than frame copy by 6.76 dB on average.**

I. INTRODUCTION

Scalable video coding (H.264/AVC SVC) [1] that is the extended version of H.264/AVC has been finalized in 2007. The objective of the SVC standardization is to encode a high-quality video that consists of one or more subset videos. The complexity, reconstruction quality as well as the quantity of data in the subset bitstream is also similar to those of H.264/AVC. SVC has several possible scalability modalities such as temporal, spatial, SNR/quality/fidelity scalabilities and a combined scalability that is a combination of the previous three modalities. In SVC, the multilayer video coding structure is adopted, called pyramid coding scheme shown in Fig. 1.

Although SVC could accomplish good coding performance, it cannot avoid transmission error or loss, degrading video quality seriously. Due to the inter-layer prediction in SVC, error propagation not only appears in the following frames but also the frames of different layers when errors occurring in the lower layers. In the case of transmission error, different mechanisms could be set up to correct errors and improve quality of service.

Retransmission is one of these mechanisms, retransmitting loss packets. However, such kinds of mechanisms cause information overhead that requires extra bandwidth, slowing downing the transmission rate. Error concealment is a decoder-based error correction tool, which could provide well performance when packet loss rate is low. In [2], frame copy (FC) and temporal direct (TD) motion vector generation is adopted when the base layer (BL) is lost. While the enhancement layer (EL) is lost, motion and residual upsampling (BLSkip) and reconstruction base layer upsampling (RU) are chosen to conceal the error. In [3], several decision rules are used to choose TD or BLSkip for concealment. In [4], the authors proposed an error tracking model to estimate the concealment error and error propagation for the choice of BLSkip or RU. However, the PSNR will decrease rapidly when packet loss rate gets high. In this situation, error resilience has better performance than error concealment.

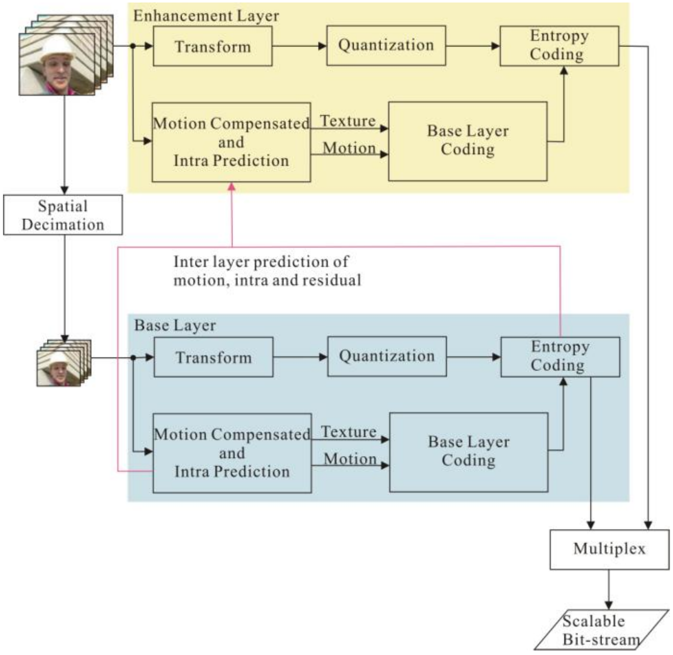

Fig. 1. Architecture of SVC encoding process for spatial scalability

The papers [5] and [6] are two error resilient methods used in SVC based on a multiple description coding and redundant coded picture, respectively. In comparison with above methods, date embedding is another way used in error resilience. Kang *et al.* [7] embedded the direction of edge, mode or motion vector for error resilience. The odd-even data embedding scheme [8] is one of the most popular methods for data embedding. However, most embedded schemes change the original data and cause quality degradation even when packet does not lose. A reversible data embedding scheme proposed in [9] can restore the original data without any distortion. Lie *et al.* [10] used this scheme to embed required indexes of wavelet coefficient for error concealment. Lin *et al.*

[11] also used this scheme to embed the downsampling pixel values.

In this paper, an error resilience based on reversible data embedding technique is proposed in SVC. The proposed scheme embeds the required information of the lower layer in its above layer when the macroblock (MB) in EL is inter-layer predicted.

The rest of the paper is organized as follow: the reviews of the inter-layer prediction in SVC and the reversible data embedding scheme are introduced in Sec. II. Section III describes the proposed algorithm in detail. Simulation results are presented in Section IV. Finally, concluding remarks are given in Sec. V.

II. BACKGROUND REVIEW

A. *Inter-Layer Prediction in SVC*

SVC further supports three new inter-layer prediction modes to improve the rate-distortion efficiency of the ELs for EL encoding process and they are inter-layer motion prediction, inter-layer residual prediction and inter-layer intra prediction. It is worth to mention that the quality scalability can be regarded as a special case of spatial scalability. This is because each layer has the same resolution in the quality scalability. Two inter-layer predictions cannot be employed in quality scalability: the upsampling operations and the inter-layer deblocking for intra-coded reference layer MB [1].

1) *Inter-Layer Motion Prediction:* The MB partition of the higher layer is obtained by upsampling corresponding partition of the 8x8 block of the lower layer and the motion vector of the higher layer is also derived by scaling corresponding motion vector of lower layer by a factor 2.
2) *Inter-Layer Intra Prediction:* When the BL is intra-coded, the MB of the EL can be predicted by upsampling the reconstructed sub-MB of the BL. The reconstructed sub-MB of the BL is upsampled by the four-tap filter for luminance component and bilinear filter for chrominance component.
3) *Inter-Layer Residual Prediction:* Inter-layer residual prediction can be employed for all inter-mode blocks in the EL. The residual data of the corresponding 8×8 block of the BL is block-wise upsampled by a bilinear filter. The difference between the upsampling residual and the predicted one of EL block needs to be coded.

B. *Reversible Data Embedding Scheme*

The reversible data embedding could provide embedding data without any distortion compared to the original data, which implies that the proposed scheme would not cause damage in visual quality of the original bitstream. The reversible data embedding schemes are classified into two categories, pixel-domain and DCT-domain, that are proposed in [9] and [10].

To embed m bits into two quantized coefficients f_0^o and f_1^o, we should calculate d^o that means the difference between two of them. Then, d^h is calculated by expanding the difference d^o and add the m bits in Eqs. (1) and (2). $\bar{f}$ is the average of f_0^o and f_1^o. Then, the two coefficients, f_0^h and f_1^h, are modified as the subtraction and addition between the low-bound and up-bound of the half of d^h, respectively, shown in Eqs. in (3) and (4).

$$d^o = f_0^o - f_1^o \tag{1}$$

$$d^h = 2d^o + m \tag{2}$$

$$\bar{f} = \left\lfloor \frac{f_0^o + f_1^o}{2} \right\rfloor \tag{3}$$

$$\begin{cases} f_0^h = \bar{f} - \left\lfloor \dfrac{d^h}{2} \right\rfloor \\ f_1^h = \bar{f} + \left\lceil \dfrac{d^h}{2} \right\rceil \end{cases} \tag{4}$$

The data extraction process is firstly to compute the difference d^h between the two coefficients. Then the embedded data could be extracted by divide 2 of d^h, in Eq. (5). Finally, the data could be recovered by Eq. (6).

$$m = d^h \% 2 \text{ ,and } d^h = f_0^h - f_1^h \tag{5}$$

$$\begin{cases} f_1^o = \bar{f} - \left\lfloor \dfrac{d^o}{2} \right\rfloor \\ f_2^o = \bar{f} + \left\lceil \dfrac{d^o}{2} \right\rceil \end{cases} \text{where } \bar{f} = \left\lfloor \frac{f_1^h + f_2^h}{2} \right\rfloor,\ d^o = \left\lfloor \frac{d^h}{2} \right\rfloor \tag{6}$$

III. PROPOSED ERROR RESILIENCY STRAEGY

As mentioned in the introduction, the proposed scheme aims to improve visual quality of higher layer when the lower layer is lost as shown in Fig 2. We explicate the proposed scheme in three parts: the inter-layer intra prediction, the inter-layer motion prediction and the inter-layer residual prediction. Here, we use two-layer SVC case (one BL and one EL) as one example to explain the proposed method in the following.

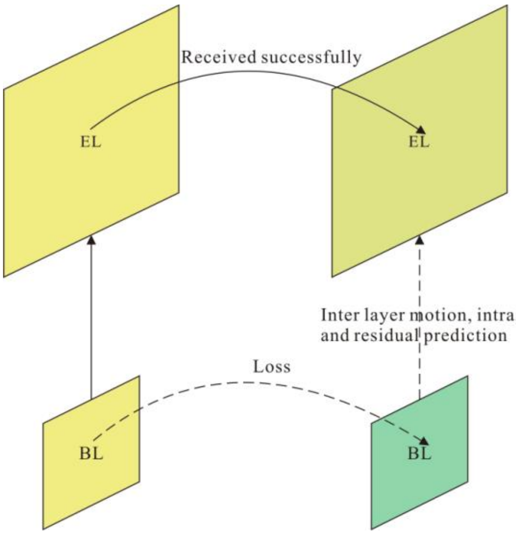

Fig 2. Illustration of the loss of BL and the inter-layer prediction of EL

A. *Inter-layer Intra Prediction*

The proposed scheme embeds the indexes of wavelet coefficient of BL in EL when the MB in the EL is coded as “IntraBL.” This is because the dynamic range of DCT coefficient is harder to estimate than that of wavelet coefficient. If the MB in the EL is encoded as “IntraBL”, the luminance component of the co-located block in BL is performed by wavelet transform. The average of chroma components of the first frame in BL is embedded in the first I frame of EL and other chroma components of other frames in BL are concealed by motion copy in decoder. If the resolution ratio between EL and BL is four, an MB in EL has a co-located 8 × 8 block in BL. Then, a three-level wavelet transform is performed in the 8 × 8 block of BL shown in Fig. 3. 8-bit is used to represent band 0 due to its significance, and bands 1-6 are embedded with their center values that are represented with their corresponding the indexes shown in Table I. Because the probability of occurrence of the wavelet coefficient in the range of -1 to 1 is close to 70%, one bit is used to represent and 4-bit for the rest. The compression ratio of this method is less than that of equal length encoding method (3-bit). Table I with 9 classes is trained by the blocks of the BL whose co-located MBs in EL is “IntraBL.” The training sequences include *Akiyo*, *Bus*, *City*, *Football*, *Foreman*, and *Mobile*. The clustering result is similar when the half of these sequences is used. In the decoder, the block is recovered by the center of the wavelet coefficient.

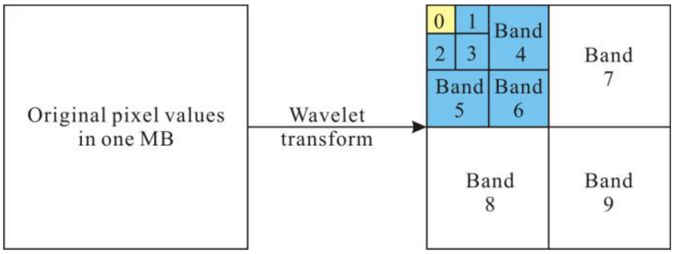

Fig. 3. Three-level wavelet transform of the block.

Table I The index table of wavelet coefficient

Wavelet coefficient (value)	Center	index
value <= -41	-52	1111
-40 <= value <= -23	-30	1110
-22 <= value <= -12	-16	1101
-11 <= value <= -6	-8	1100
-5 <= value <= -2	-3	1011
-1 <= value <= 3	0	0
4 <= value <= 11	6	1000
12 <= value <= 26	16	1001
value >= 27	36	1010

B. *Inter-layer Motion Prediction*

If the predicted motion vector in the EL is from the BL, the half-pixel precision motion vector of the BL is embedded in the quantized coefficient of EL. The embedded data includes 6 bits that contain the half pixel precision motion vector and its significant bit for each horizontal or vertical motion vector. The horizontal and vertical predicted motion vectors of list_0 and list_1 are embedded in the blocks, as shown in Fig. 4. Each block embeds at most two bits, which means at most 4 quantized coefficients are changed in one 4 × 4 block.

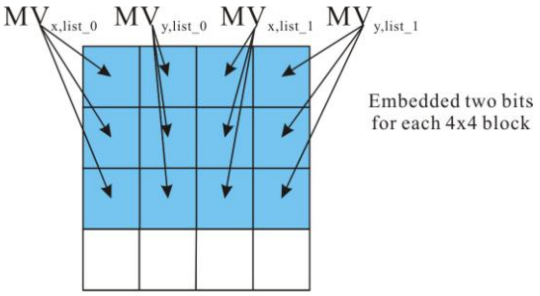

Fig. 4. The blocks for embedding the predicted motion vectors

C. *Inter-layer Residual Prediction*

Intra and motion vector prediction is much more important than residual prediction for loss frame recovery. The embedded information of predicted residual causes dramatically the increase of bit rate. Therefore, the predicted residual of the current frame is replaced by motion copy from the previous coded frame in BL.

The flowchart of the proposed algorithm in the decoder side is shown in Fig. 5. The bold block represents down-sampling the EL to BL.

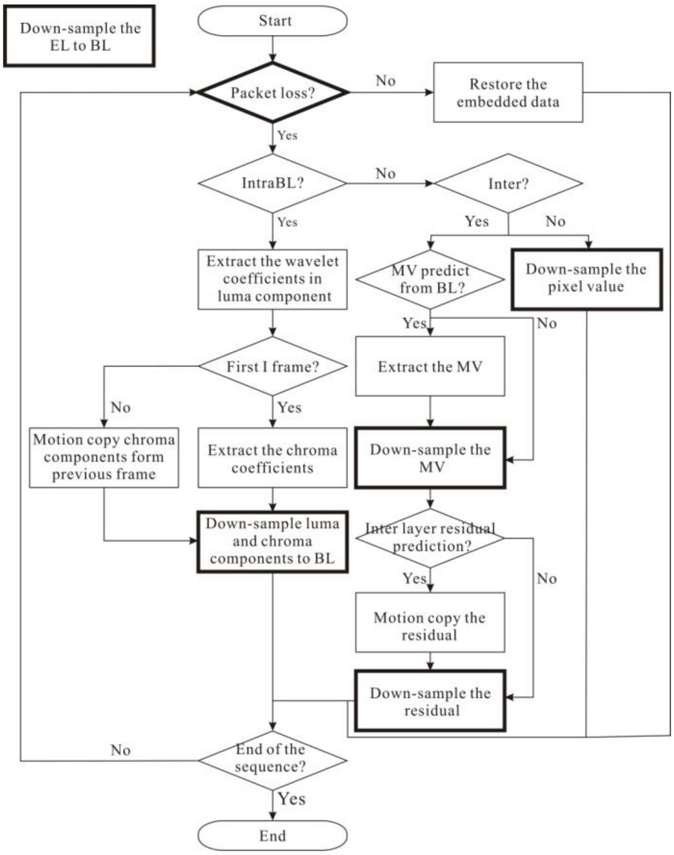

Fig. 5. Flowchart of the proposed scheme in decoder

IV. SIMULATION RESULTS

The algorithm proposed in this paper and frame copy method are simulated on scalable video coding reference software JSVM 9.14. Simulation settings are listed in Table Ⅱ. The 2-layer simulation tests seven benchmarks, including *Bus*, *City*, *Crew*, *Football*, *Foreman*, *Harbour* and *Mobile*.

Bus and *Football* are encoded with 150 and 260 frames separately, and others are encoded with 300 frames. The packet loss rates (PLR), 5%, 10%, 15%, 20%, are simulated in BL, and EL is lossless. The average PSNR is calculated by repeating 10 times for each packet loss rate.

Table Ⅱ Simulation settings

Number of layers	2
Number of reference frame	1
GOP size	16
Intra refresh	Only the first frame
Frame rate	30
Resolution of BL	QCIF
Resolution of EL	CIF

Table Ⅲ shows the average Y-PSNR and gain in EL of seven test sequences. As can be seen in Table III, the proposed method outperforms frame copy by 6.67 dB on average, and the average increase of bit rate is 9.49%. The frame copy method only copies the previous frame in BL. Without the information of motion vector from the BL, the predicted motion vector in the EL is assigned to zero motion vectors. Fig. 6-9 show that the proposed method provides better subjective quality than frame copy. The proposed method handles well both on object moving (Fig. 6-8) and camera moving (Fig. 9). Fig. 10 shows that the proposed method could do well in the “IntraBL” mode. In addition, when the first frame in BL losses, the proposed method could have coarse quality in EL and frame copy will fail when the MB is inter-layer intra prediction.

Table Ⅲ Performance comparisons of the proposed algorithm and frame copy

Sequence	PLR (%)	PSNR		
		Frame Copy	Proposed	Gain
Bus	5	28.80	33.02	4.22
	10	25.09	31.80	6.71
	15	24.00	30.78	6.78
	20	22.26	30.56	8.30
City	5	29.27	35.55	6.28
	10	27.23	34.92	7.69
	15	25.27	33.49	8.22
	20	23.78	32.90	9.12
Crew	5	32.95	36.09	3.14
	10	31.31	35.89	4.58
	15	29.34	35.66	6.32
	20	27.08	35.32	8.24
Football	5	31.68	34.83	3.15
	10	29.27	34.57	5.30
	15	27.79	34.04	6.25
	20	26.26	33.50	7.24
Foreman	5	31.32	36.25	4.93
	10	27.46	35.54	8.08
	15	26.44	35.23	8.79
	20	24.44	34.72	10.28
Harbour	5	26.89	31.88	4.99
	10	25.26	31.56	6.30
	15	23.73	30.09	6.36
	20	22.23	29.16	6.93
Mobile	5	23.31	30.33	7.02
	10	22.42	29.88	7.46
	15	20.73	28.07	7.70
	20	17.31	26.19	8.88
Average		26.16	32.92	6.76


(a) Frame copy (b) Proposed

Fig. 6. Comparison of the subjective quality of the 154th frame on *Foreman* sequence under 10% packet loss ratio.


(a) Frame copy (b) Proposed

Fig. 7. Comparison of the subjective quality of the 19th frame on *Football* sequence under 5% packet loss ratio.


(a) Frame copy (b) Proposed

Fig. 8. Comparisons of the subjective quality of the 53th frame on *Crew* sequence under 20% packet loss ratio.


(a) Frame copy (b) Proposed

Fig. 9. Comparison of the subjective quality of the 42th frame on *Mobile* sequence under 5% packet loss ratio.

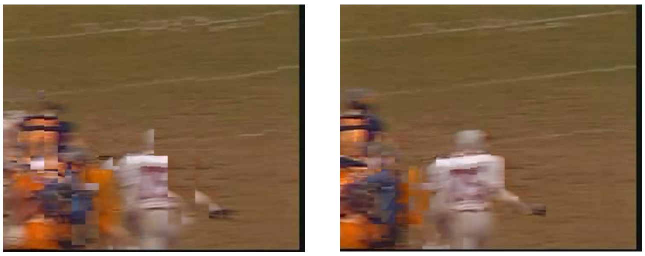
(a) Frame copy (b) Proposed

Fig. 10. Comparison of the subjective quality of the 81th frame on *Football* sequence under 5% packet loss ratio.

V. CONCLUSIONS

A cross-layer error resilient scheme based on reversible data embedding in SVC is proposed in this paper. The predicted motion vector and the indexes of wavelet coefficient are embedded in the higher layer for improvement of visual quality when packets are lost. Experiment results show that a PSNR improvement is 6.76 dB in average and the maximum improvement is up to 10.28 dB when packet loss rate is equal to 20%.

ACKNOWLEDGMENTS

The authors would like to thank the National Science Council of the Republic of China for financially supporting this research under Contracts No. NSC-98-2221-E-110-057.

REFERENCES

[1] H. Schwarz, D. Marpe and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 17, No. 9, pp.1103 – 1120, September 2007.

[2] Y. Chen, K. Xie, F. Zhang, P. Pandit and J. Boyce, "Frame loss error concealment for SVC," *Journal of Zhejiang University - Science A*, Vol. 7, pp. 677-683, May 2006.

[3] L.-N. Zhang, C. Yuan and Y.-Z. Zhong, "A novel SVC VOD system with rate adaptation and error concealment over GPRS/EDGE network," *in Proceedings of Congress on Image and Signal Processing, 2008(CISP '08),* Vol. 1, pp. 349-354, May 2008.

[4] C.-S. Park, M.-S. Yoon, S.-T. Kim, H.-C. Choi and S.-J. Ko, "Error tracking based error concealment strategy for scalable video coding," *in Proceedings of 2008 International Conference on Consumer Electronics*, pp. 1-2, January 2008.

[5] H. Mansour, P. Nasiopoulos and V. Leung, "An efficient Multiple Description Coding scheme for the scalable extension of H.264/AVC (SVC)," *in Proceedings of IEEE International Symposium on Signal Processing and Information Technology,* pp. 519 – 523, Aug. 2006.

[6] J. Jia, H.-K. Kim and H.-C. Choi, "Polyphase downsampling based redundant picture coding for SVC error resiliency," *in Proceedings of Congress on Image and Signal Processing*, Vol. 2, pp. 90 – 94, May 2008.

[7] L.-W. Kang and J.-J. Leou, "An error resilient coding scheme for H.264 video transmission based on data embedding," *in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 3, pp. iii - 257- iii - 260, May 2004.

[8] P. Yin, B. Liu and H.-H. Yu, "Error concealment using data hiding," *in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing,,* Vol. 3, pp. 1453 – 1456, May 2001.

[9] J. Tian, "Reversible data embedding using a difference expansion," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 13, No. 8, pp. 890 – 896, August 2003.

[10] W.-N. Lie, T. C.-I. Lin, D.-C. Tsai and G.-S. Lin, "Error resilient coding based on reversible data embedding technique for H.264/AVC," *in Proceedings of IEEE International Conference on Multimedia and Expo*, pp. 1174 – 1177, July 2005.

[11] S.-D. Lin, H.-C. Meng, C.-Y. Chuang, J.-Y. Huang, "An error resilient technique using reversible data embedding in H.264/AVC," *in Proceedings of 3rd International Conference on Innovative Computing Information and Control, 2008 (ICICIC '08)*, pp. 26 – 30, June 2008.