Error Concealment Using Inter-layer Correlation for Scalable Video Coding

Chun Su Park, Tae Shick Wang, and Sung Jea Ko

ABSTRACT—In this paper, we propose a new error concealment (EC) method using inter-layer correlation for scalable video coding. In the proposed method, the auxiliary motion vector (MV) and the auxiliary mode number (MN) of intra prediction are interleaved into the bitstream to recover the corrupted frame. In order to reduce the bit rate, the proposed method encodes the difference between the original and the predicted values of the MV and MN instead of the original values. Experimental results show that the proposed EC outperforms the conventional EC by 2.8 dB to 6.7 dB.

Keywords—Error concealment, auxiliary motion vector, scalable video coding.

I. Introduction

Scalable video coding (SVC) is the powerful state-of-the-art scalable video standard which was finalized as an amendment of the H.264/AVC [1]. In the SVC standard, the most fundamental information used to reconstruct a video frame is referred to as the base layer (BL). The other layers, called enhancement layers (ELs), produce more refined information when combined with the lower layers.

Many error concealment (EC) techniques have been proposed to cope with transmission error [2], [3]. Basically, these methods have been used to reconstruct corrupted frames of non-scalable bitstreams. Although EC can be directly applied to a scalable bitstream, the application in which EC is performed at each separate layer can produce unsatisfactory results since the inter-layer correlation is not considered. In the case of a scalable bitstream, a corrupted frame can be recovered using the information of the lower layers as well as the current layer.

In this paper, we propose an EC method which uses inter-layer correlation for a scalable bitstream. The proposed method interleaves the auxiliary motion vector (MV) and mode number (MN) of the current frame into the next frame. In order to reduce the bit rate further, the difference between the original and predicted values of the MV and MN of the EL is encoded. At the decoder, when a transmission error is detected, the difference is utilized along with the information of the BL to reconstruct the corrupted block of the EL.

II. Proposed EC Method

1. Proposed Encoding Method

For simplicity of explanation, we consider a bitstream with two layers. The auxiliary MV (MN) is coded when the corresponding block in the BL is coded in inter (intra) mode. Let F_n be the n-th frame of the EL. In the proposed encoding scheme, the MV and MN of the current frame, F_n, are interleaved into the bitstream of the next frame F_{n+1} as shown in Fig. 1. Thus, the MV and the MN of F_n can be preserved
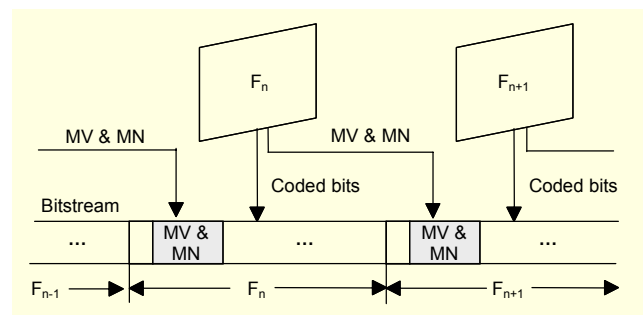

Fig. 1. Concept of the proposed method.

Chun Su Park (phone: + 82 2 3290 3790, email: cspark@dali.korea.ac.kr), Tae Shick Wang (email: tkwang@dali.korea.ac.kr), and Sung Jea Ko (email: sjko@dali.korea.ac.kr) are with Department of Electronic Engineering, Korea University, Seoul, Korea.

even when a whole frame is corrupted by transmission error.

Next, to reduce the bit rate further, we propose a new coding scheme that uses the residual MV (RMV) and the residual MN (RMN) instead of the original MV and MN. Because the only difference between the picture in the BL and that in the EL is the size or quality, the MV and MN of the EL are strongly correlated with those of the BL. Thus, the number of coded bits for the MV and MN can be significantly reduced by encoding the difference between the original value and the predicted value. Let the original and predicted values for the i-th block be X_o^i and X_p^i, respectively. Then the proposed scheme is represented by

$$\Delta X^i = X_o^i - X_p^i, \qquad (1)$$

where X_o^i is obtained by encoding the block in the EL, and X_p^i can be determined by using the correlation between BL and EL. Let the MV and MN of the BL be V_b^i and N_b^i, respectively. Then, X_p^i is simply determined by

$$X_p^i = \begin{cases} \alpha \cdot V_b^i & \text{for RMV,} \\ N_b^i & \text{for RMN,} \end{cases} \qquad (2)$$

where α is the scaling factor given by the proportion of the spatial resolution of the picture in the BL to that in the EL.

2. EC Using RMV and RMN

An EC method to recover the corrupted block using the RMV at the decoder and the reconstruction process using the RMN are presented here. Let R_e (R_b), P_e (P_b), and D_e (D_b) be the reconstructed, prediction, and residual blocks of the EL (BL), respectively. Then, the temporal prediction in the EL is represented by

$$R_e = S(P_e + D_e), \qquad (3)$$

where $S(\cdot)$ is a deblocking filter. Suppose that the current frame F_n has been lost in transmission. In this case, P_e can be obtained by using the motion vector V_e of the block in the EL, where V_e is given by

$$V_e = \alpha \cdot V_b + \Delta X. \qquad (4)$$

In (3), since D_e is not available when the current picture is lost, we estimate D_e as

$$D_e \approx U(D_b) = U(R_b - P_b), \qquad (5)$$

where $U(\cdot)$ represents an optional upsampling filter. As shown in Fig. 2, the prediction block P_b should be obtained by using the scaled MV of the EL instead of V_b, where the scaled MV, V_e', is given by
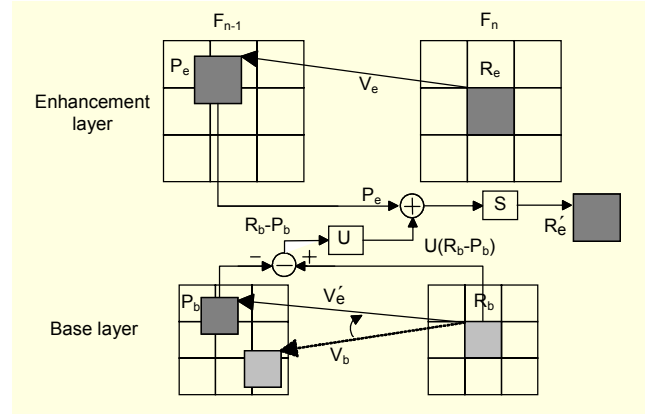

Fig. 2. EC using the RMV.

$$V_e' = V_e / \alpha = V_b + \Delta X / \alpha. \qquad (6)$$

Then, from (3) and (5), the recovered block R_e' can be obtained by

$$R_e \approx R_e' = S(P_e + U(R_b - P_b)). \qquad (7)$$

The proposed EC method requires minor modification of the syntax of the SVC standard. At the end of the slice header, we add a flag, *redundant_inf_flag*, which indicates whether the current slice corresponding to F_n includes ΔX's (*redundant_inf_flag* = 1) or not (*redundant_inf_flag* = 0). At the decoder, the following parsing process is performed:

- If *redundant_inf_flag* = 0, the decoder skips the parsing process for ΔX.
- If *redundant_inf_flag* = 1, the decoder parses the ΔX's just after parsing the slice header. The decoder stops the parsing process after the ΔX's for all the blocks in the previous frame F_{n-1} are completely parsed. The number of blocks in F_{n-1} can be obtained by using the block partitioning information at the BL.

III. Experimental Results

To evaluate the performance of the proposed method, we use four scalable sequences, "Foreman," "City," "Harbour," and "Soccer," with QCIF at 15 fps for the BL and CIF at 15 fps for the EL. The performance of the proposed EC scheme is compared with that of the reconstruction base layer upsampling (RU) technique, one of four EC methods adopted in the SVC standard [4]. We calculate the average redundant bits and PSNR for video sequences of 130 frames.

Table 1 shows the comparative results associated with redundant bits and PSNR. We measure the average coded bits B_T for the original frames without redundant information in the

Table 1. Performance of the proposed EC method.

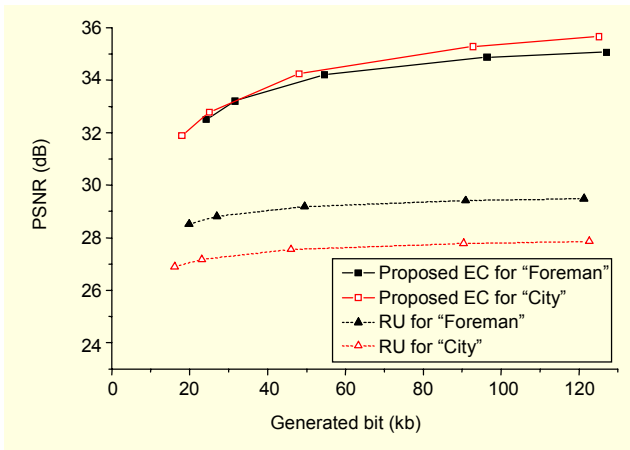| Sequence	Average bitrate (bits)			Average PSNR (dB)		
	B_D	B_O	B_T	O_{PSNR}	P_{PSNR}	R_{PSNR}
Foreman	5,034	11,297	36,060	39.93	34.22	29.20
City	2,105	8,459	40,855	38.76	34.24	27.57
Harbour	2,869	5,523	113,417	38.86	29.76	23.25
Soccer	5,907	13,334	51,829	39.48	32.71	29.87

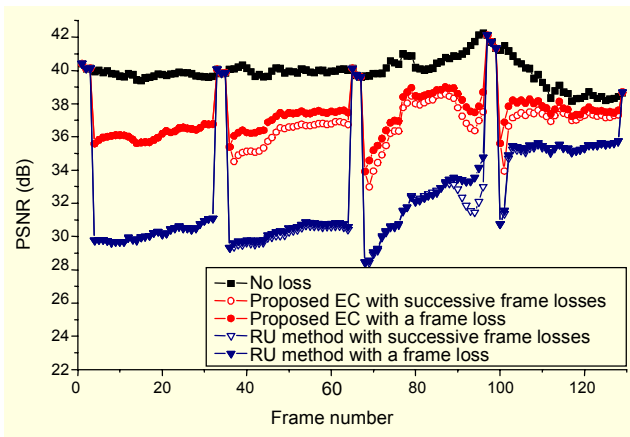Fig. 3. Rate-distortion curves for "Foreman" and "City."

Fig. 4. Performance of the proposed EC for "Foreman."

EL. In Table 1, B_O (B_D) is the redundant bits per frame, which are generated by inserting the original MV and MN (the RMV and RMN). Clearly, the use of the RMV and RMN can significantly reduce the redundant bits. The "City" and "Harbour" sequences produce lower overhead than "Foreman" and "Soccer" because the MV and MN of the EL for the former sequences are more similar to those of the BL. In Table 1, O_{PSNR} represents the PSNR of the original uncorrupted sequence, and P_{PSNR} and P_{PSNR} are the PSNRs of the proposed

method and the RU method, respectively. Note that the PSNR of the proposed method is approximately 2.8 dB to 6.7 dB higher than that of the RU method for different sequences. Figure 3 shows the rate-distortion curves which are obtained by averaging the PSNR of the recovered frame when a frame is lost in every GOP with 32 frames.

Figure 4 shows the PNSR performance at the first 130 frames for "Foreman." Since the recovered frame is used as the reference for the posterior frames, drift error is observed. In order to show the effects of drift error, we remove the third frame in every GOP with 32 frames. The PSNR is periodically refreshed because the first frame in GOP is coded as intra mode. Figure 4 also shows the results of the case in which two successive frames (the third and fourth) are lost from the second GOP. Although the measured PSNR of the proposed method is slightly degraded, it is always higher than that of the RU method over the entire sequence.

As shown in (7), the complexity of the proposed EC method is almost the same as that of the normal decoding operation. We also measured the processing time for all sequences. The results indicate that the complexity is slightly increased by 8 % as compared with the normal decoding operation.

References

[1] T. Wiegand et al., Joint Draft 6, ISO/IEC JTC1/SC29/WG11, Doc. JVT-S201, Apr. 2006.

[2] J.S. Lee et al., "Residual Motion Coding Method Supporting Error Concealment in Low Bitrate Video Coding," *Electronics Letters*, vol. 41, no. 7, Mar. 2005, pp. 406-408.

[3] M.K. Seo et al., "Piggyback Packetization of Duplicate Packets for Packet-Loss Resilient Video Transmission," *IEICE Comm.*, vol. E89-B, no. 10, Oct. 2006, pp. 2802-2808.

[4] C. Ying et al., "Frame Loss Error Concealment for SVC," *Journal of Zhejiang University*, vol. 7, Apr. 2006, pp. 677-683.