A NEW SEAMLESS BITSTREAM SWITCHING SCHEME FOR H.264 VIDEO ADAPTATION WITH ENHANCED CODING PERFORMANCE

Shih-Ming Hsu[*], *Chia-Wen Lin*[*], *and I-Hsien Lee*[†]

[*]Department of Computer Science & Information Engineering
National Chung Cheng University
Chiayi 621, Taiwan
[†]Information & Communications Research Labs
Industrial Technology Research Institute
Hsinchu 310, Taiwan

ABSTRACT

In this paper, we propose a new seamless bitstream switching scheme to improve the coding performance of H.264 SP-frames for rate adaptation. Our method removes one of the two re-quantization blocks in the SP-frame encoders so as to significantly improve coding performance. The seamless switching property of SP-frames is retained by properly restructuring the primary and secondary switching frame codecs. Experimental results show that our proposed scheme achieves close coding performance to that of regular H.264 P-frames and significantly better performance than that of SP-frames. The proposed method also provides the advantage of using a single secondary switching bitstream for both switching-up and switching-down processes.

Index Terms**—** video coding, video signal processing, multimedia communication

1. INTRODUCTION

With the proliferation of online multimedia contents, the popularity of multimedia streaming technologies, and the establishment of video coding standards, people are able to ubiquitously access and retrieve various multimedia contents via the Internet, promoting networked multimedia services at an extremely fast pace. In streaming video, users may access videos from heterogeneous networks such as Local Access Network (LAN), Digital Subscriber Line (DSL), cable, wireless networks, and dial-up. The different access networks have different channel characteristics such as bandwidths, bit error-rates, and packet loss-rates. At the users' end, network appliances including handheld computers, Personal Digital Assistants (PDA), set-top boxes, and smart cellular phones are slated to replace personal computers as the dominant terminals for accessing the Internet. These network terminals vary significantly in resources such as computing power and display capability. To flexibly deliver multimedia data to users with different available resources, access networks, and interests, the multimedia contents may need to be adapted dynamically according to the usage environment. For example, the notion of Universal Multimedia Access (UMA) calls for the provision of different presentations of the same multimedia content, with more or less complexity to suit the different usage environments in which the content is consumed.

Video adaptation [1] is an emerging field that offers a rich body of knowledge and techniques for handling the huge variation of resource constraints (e.g., bandwidth, display capability, processing speed, and power consumption) and the diversity of user tasks which often arise in pervasive media applications. Dynamic bitstream switching [2]-[6] is an efficient means for realizing video adaptation which has been widely deployed in commercial streaming services to deal with bandwidth variation in a standard compliant way. With bitstream switching, the server provides multiple bitstreams with different bitrates/resolutions for each client to switch over to choose the bitstream which matches the client's channel bandwidth the most for rate adaptation. For instance, clients with high channel bandwidths can subscribe to higher-rate bitstreams for better video quality, whereas low-bandwidth clients need to subscribe to lower-rate bitstreams with worse perceptual visual qualities. There are some issues with bitstream switching schemes to concern about. For example, when the available channel bandwidth of a client drops, the client has to switch from one higher-rate bitstream to another lower-rate one (a "switching-down" process), and vice versa (a "switching-up" process). Because general video coding schemes use the temporal predictive coding, switching at any predictive frame would cause different reference frames at the encoder and the decoder. This mismatch leads to drift which will cause serious error propagation to subsequent predictive frames until reaching the next intra frame [2][3].

In order to mitigate the quality drift caused by bitstream switching, a pioneering work in [2] proposed to use a new-type intermediate switching frame (S-frame) to compensate for the switching drift at predictive frames. The S-frames can effectively reduce the switching drift but cannot eliminate the drift completely if they are not coded losslessly. Recently, the H.264 standard has proposed a new picture type, the SP-frames [4], which supports drift-free switching at predictive frames. Like normal predictive frames (P-frames), SP-frames use motion compensated predictive coding to remove the temporal redundancy, while allowing identical reconstruction of the frames at switching points even when they are predicted with different reference frames [4]. The SP-frames can provide seamless switching points just like intra frames, but their frame sizes are much smaller than intra frame due to the predictive coding.

Fig. 1 illustrates an example of using SP-frames to switch form one bitstream to another. Suppose one sequence is encoded into two bitstreams with different bit-rates. As shown in Fig. 1, the SP-frames can be classified into primary SP-frames (e.g., S_1 and S_2) and secondary SP-frames (e.g., S_{12}), respectively. The secondary SP-frames are the special frames which can be used for switching up or down without drift just like switching at intra frames. They will be transmitted only when switching between two bitstreams. For example, as shown in Fig. 1, if the server wants to switch from bitstream 1 to bitstream 2, it sends S_{12} instead of S_1 or S_2 to the decoder at the switching point.

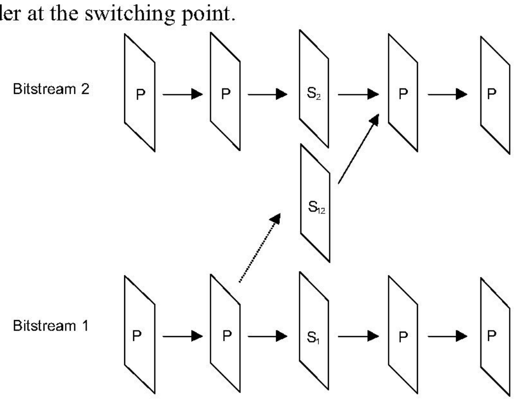


Fig. 1. Illustration of bitstream switching using SP-frames.

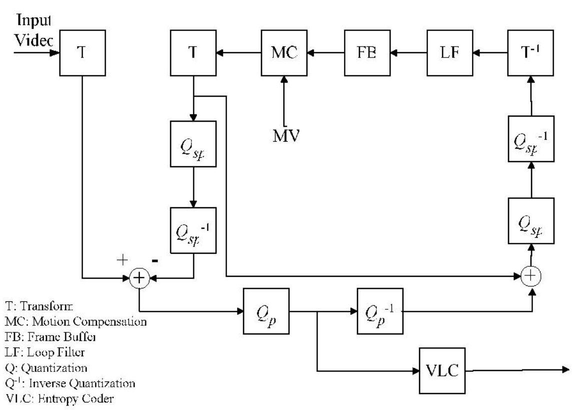


Fig. 2. Block diagram of H.264 primary SP-frame encoder.

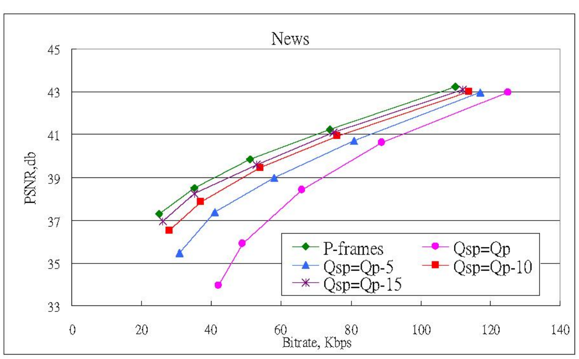


Fig. 3. PSNR performance comparison of H.264 P-frames and SP-frames for the *News* sequence at different bit-rates.

Fig. 2 depicts the encoder block diagram for generating H.264 primary SP-frames [4]. Compared to the P-frame encoding process, the primary SP-frame encoding process involves two extra re-quantization (also followed by an inverse quantization) blocks with the same quantization step-size of Q_{sp} on the temporal prediction loop and the reconstruction loop, respectively. Using the additional quantizations, for the example shown in Fig. 1, the reconstructed S_{12} frame can be exactly identical to the reconstructed S_2 frame, thereby achieving seamless bitstream switching [4][5] without introducing mismatch error between S_{12} and S_2. However, such re-quantizations reduce the coding efficiency significantly as analyzed in [6], in which analytical models for the rate-distortion performance of SP-frames were derived from the power spectral density (PSD) of the image signal and the prediction error. Fig. 3 compares the average PSNR performances of H.264 P-frames and SP-frames for the *News* sequence at different bit-rates, showing that performance degradation caused by the re-quantizations is significant.

In this paper, we propose a new seamless bitstream switching scheme to improve the coding performance of H.264 SP-frames by removing one of the two re-quantization blocks in the SP-frame encoders. The drift-free switching property of SP-frames is still retained by properly restructuring the primary and secondary switching frame codecs. The proposed method also provides the advantage of using a single secondary switching bitstream for both switching-up and switching-down processes, thereby reducing the storage cost significantly.

2. PROPOSED SEAMLESS SWITCHING SCHEME

As mentioned above, the extra re-quantizations of SP-frames lead to coding performance degradation of primary SP-frames. To address the problem, we propose a new seamless bitstream switching scheme called SS-frames. Fig. 4 illustrates an example of using SS-frames to switch from higher bit-rate bitstream to lower one. The dotted box is the switch point at which the server performs bitstream switching. The broken curves between frames indicate that the encoder uses the higher-rate reconstructed frame, instead of the incoming video frame, as input video to encode the lower bit-rate frame.

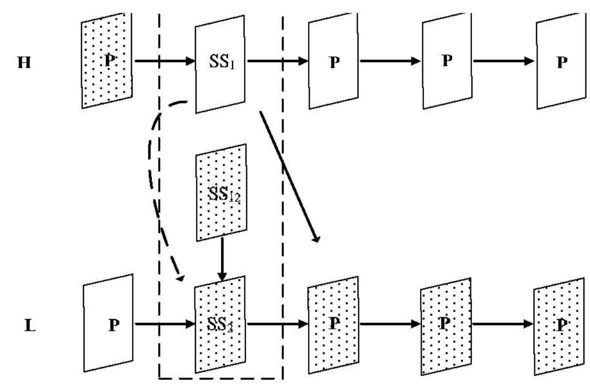


Fig. 4. Illustration of bitstream switching using SS-frames.

The gray blocks represent the frames that are actually transmitted to the client while performing switching. In our SS-frame method, at the switch point, the server sends the secondary bitstream (e.g., SS_{12}), the residual of the lower bit-rate bitstream (e.g., SS_2), and the motion vectors of the higher bit-rate bitstream. With these data, the lower-rate frame (e.g., SS_2) can be reconstructed form the higher-rate primary residues plus the secondary switching frames, as will be elaborated in the following.

2.1. Encoding process

Fig. 5 depicts the block diagrams of the proposed primary and secondary SS-frame encoders. The upper half of Fig. 5 illustrates the architecture of the higher-rate primary SS-frame encoder. Compared to the SP-frame encoder shown in Fig. 2, the re-quantization and de-quantization blocks in the reconstruction loop of the encoder have been removed to increase the coding efficiency, whereas the re-quantization block (e.g., Q_s which is operated on HK_{pred}) and the subsequent de-quantization block prior to motion compensation are kept to avoid the drifting error while performing switching. The architecture of the lower-rate primary SS-frame encoder, as shown in the lower half of Fig. 5, is about the same with the higher bit-rate one, but it employs a coarser quantization step-size, Q_{p2}. Note that, we use the corresponding reconstructed higher bit-rate SS-frame (e.g., H_{rec}) as the input video to the lower bit-rate encoder.

The secondary SS-frame encoder depicted in Fig. 5 for drift-free bitstream switching is similar to the secondary SP-frame encoder in H.264. However, in our scheme the difference between the higher-rate quantized prediction residues, HL_{pred}, and the corresponding lower-rate version, LL_{pred}, is used to generate the secondary bitstream (e.g., SS_{12}). Instead of using two different secondary SP-frames bitstreams for switching up and down, respectively, from one bitstream to another one of different bit-rate in the SP-frame scheme, our method provides the advantage of using the same secondary SS-frame for switching up or down at a switching point. This thus reduces half of storage space for the secondary bitstream compared to the SP-frame scheme.

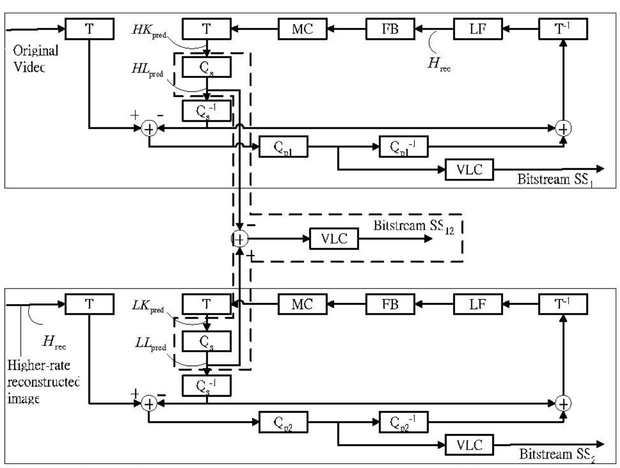


Fig. 5. Block diagram of primary and secondary SS-frame encoders.

2.2. Decoding process

Fig. 6 depicts the block diagram of the primary SS-frame decoder. The incoming bitstream is decoded by VLD, and is then de-quantized. The motion-compensated temporal prediction obtained from the frame memory is subsequently added to the dequantized residual signal. Similar to the primary SP-frame decoder, an extra re-quantization/de-quantization pair with a quantization step-size of Q_s is performed on the temporal prediction prior to the motion compensation to achieve drift-free switching. However, same with the proposed encoder, the extra re-quantization process on the reconstructed frame has been removed.

Fig. 7 shows the decoding process for switching down. As mentioned above, while switching down, the server needs to send the secondary SS-frame, the motion vectors of the higher-rate bitstream, and the residual of the lower-rate bitstream to the client decoder. The decoder first uses the received motion vectors to obtain the temporal prediction from the higher-rate reconstructed frame buffer. The decoded secondary SS-frame is then added to the temporal prediction to obtain exactly the same content as the lower-rate temporal prediction. Finally, the lower-rate primary residual is added to the de-quantized lower-rate temporal prediction, followed by the inverse transform and loop filtering, to reconstruct the lower-rate video without drift. As for switching up, the decoder architecture is the same, but the two frame memories (FB-H and FB-L) are exchanged. For simplicity, the explanation about the switching up process is omitted in this paper.

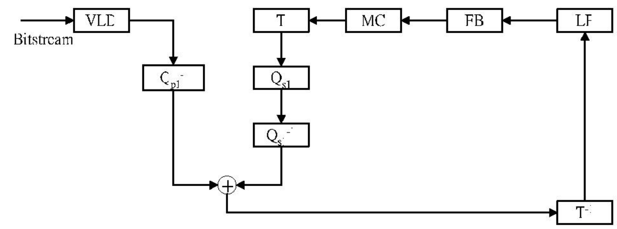


Fig. 6. Block diagram of primary SS-frames decoder.

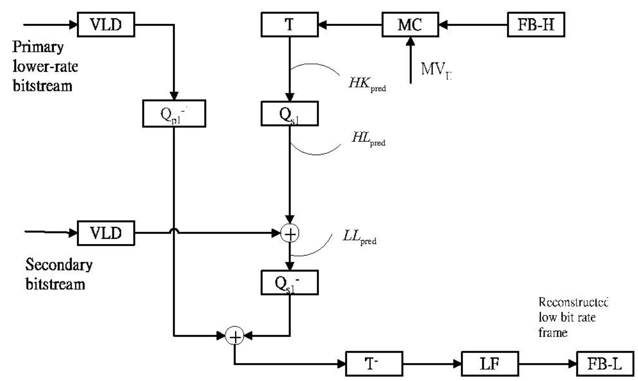


Fig. 7. Block diagram for decoding lower-rate video from primary and secondary SS-frames in a switching down process.

2.3. Analysis of drift-free switching property

Although removing a re-quantization block from the reconstruction loop of the primary SS-frame codec can significantly improve the coding efficiency of primary switching frames, it will also lead to serious drift error while performing switching should the codec not be properly redesigned. To address the problem, we modify the quantization process as enclosed by the broken lines in Fig. 5. First, the quantization processes with Q_s are separately performed on the higher-rate and lower-rate temporal predictions, HK_{pred} and LK_{pred}, to obtain $HL_{\text{pred}} = \mathrm{Q}[HK_{\text{pred}}]$ and $LL_{\text{pred}} = \mathrm{Q}[HK_{\text{pred}}]$, respectively. The secondary SS-frame is obtained by calculating the difference between the two quantized temporal predictions as follows.

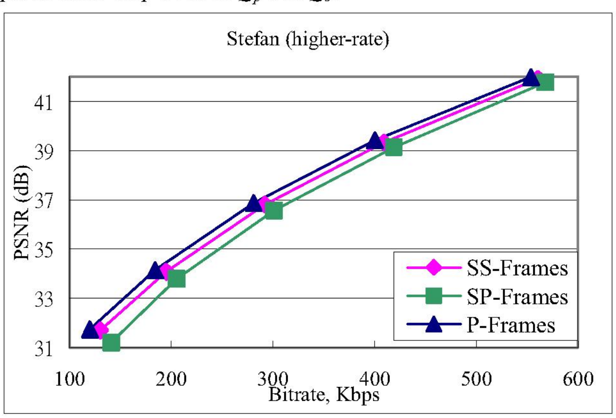
$$SS_{12} = LL_{\text{pred}} - HL_{\text{pred}} \quad (1)$$

In the switching-down decoder shown in Fig. 7, the decoded SS_{12} frame is added to the quantized higher-rate temporal prediction, HL_{pred}, to obtain LL_{pred}, which is exactly the same temporal prediction as used in the lower-rate primary bitstream encoder. As a result, we can achieve drift-free switching by adding the lower-rate primary residual frames to the reconstructed temporal prediction. This justifies the drift-free switching property of the proposed method.

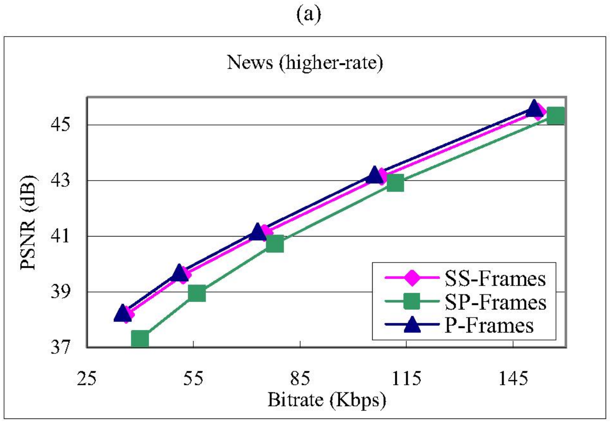
3. EXPERIMENTAL RESULTS

We implemented the proposed SS-frame codec and the SP-frame codec using the H.264 reference software (JM 7.3). Three QCIF (176×144) test sequences, *Coastguard*, *News*, and *Stefan*, are encoded with a GOP size of 30 frames and a frame rate of 30 fps. For simplicity of experiments but without loss of generality, we encoded each test sequence into two different bit-rate bitstreams with two fixed quantization step-sizes, respectively, for switching. The quantization step-size, Q_{p1}, for the higher bit-rate (i.e., higher-quality) bitstream is 22, and the quantization step-size, Q_{p2}, of the lower bit-rate bitstream is 28, respectively. The GOP structure of the two primary bitstreams is IPPP…

While switching down using SS-frames, the server needs to send a secondary switching frame, SS_{12}, the residual of the corresponding lower bit-rate primary switching frame, SS_2, and the motion vectors of the higher bit-rate primary switching frame, SS_1. The total bit-count sent at a switching point is thus the sum of the bit numbers of these data. Our experimental results show that the total bit-count required for switching using SS-frames is very close to that for switching using SP-frames for a wide range of quantization step-sizes of Q_p and Q_s.



(a)



(b)

Fig. 8. Average PNSR performance comparison of primary higher bit-rate bitstreams encoded using SS-frames, SP-frames and P-frames at different bit-rates for (a) *Stefan* and (b) *News*.

Fig. 8 compares the PSNR performance of primary higher bit-rate bitstreams encoded using SS-frames, SP-frames and P-frames for the *Stefan* and *News* sequences at different bit-rates. The re-quantization parameters for SP-frames and SS-frames, Q_{sp} and Q_s, are both set to be Q_{p1}-6. The quantization parameter for the lower-bit-rate primary bitstream $Q_{p2} = Q_{p1}+6$. As shown in Fig. 8, the coding efficiency for SP-frames scheme is significantly worse than those of the other two methods due to its two extra re-quantization processes, Q_s. By removing one re-quantization block and rearranging the switching frames, the proposed SS-frame scheme achieves close PSNR performance compared to regular P-frames, while retaining the property of seamless switching. Table I shows the average PSNR performance comparison of the three coding methods with $Q_{p1} = 22$, $Q_{p2} = 28$, and $Q_{sp} = Q_s = 16$ for three test sequences. The SS-frame method achieves PNSR performance improvement by 0.15~0.33 dB for the higher bit-rate primary bitstream (denoted as 'High'), and 0.17~0.38 dB for the lower bit-rate one (denoted as 'Low'), respectively. The PSNR improvement becomes more significant as the bit-rate is reduced.

Table I

Average PSNR performance comparison of three coding methods with $Q_{p1} = 22$, $Q_{p2} = 28$, and $Q_{sp} = Q_s = 16$ for three test sequences

Sequence	Average PSNR Performance (in dB)					
	P-frames		SP-frames		SS-frames	
	High	Low	High	Low	High	Low
Stefan	39.46	34.25	39.21	34.06	39.37	34.23
News	41.22	38.49	40.78	38.11	41.16	38.45
Coastguard	38.87	34.28	38.58	33.86	38.81	34.24

4. CONCLUSION

We proposed a new seamless bitstream switching scheme that can significantly improve the coding performance of SP-frames for H.264 rate adaptation. By removing one re-quantization block and rearranging the switching frames, the proposed SS-frame scheme achieves close PSNR performance compared to regular P-frames, while retaining the property of drift-free switching. The experiments show that our proposed scheme not only enhances the coding performance but also has about the same bit-count as SP-frames within a switching window.

REFERENCES

[1] S.-F. Chang and A. Vetro, "Video adaptation: Concepts, technologies, and open issues," *Proc. IEEE*, vol. 93, no. 1, pp. 148-158, Jan. 2005.

[2] N. Färber and B. Girod, "Robust H.263 compatible video transmission for mobile access to video servers," in *Proc. Int. Conf. Image Processing*, Oct. 1997, Santa Barbara, CA.

[3] B. Xie and W. Zeng, "Rate-distortion optimized aynamic bitstream switching for scalable video streaming,"in *Proc. IEEE Int. Conf. Multimedia & Expo,* June 2004, Taipei, Taiwan.

[4] M. Karczewicz and R. Kurceren, "The SP- and SI-frames design for H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol.13, no. 7, pp. 637-644, July 2003.

[5] X. Sun, F. Wu, S. Li, W. Gao, and Y.-Q. Zhang, "Seamless switching of scalable video bitstreams for efficient streaming," *IEEE Trans. Multimedia.*, vol.6, no. 2, pp. 291-303, Apr. 2004.

[6] E. Setton and B. Girod, "Rate-distortion analysis and streaming of SP and SI frames," *IEEE Trans. Circuits Syst. Video Technol.* vol. 16. no. 5, June 2006.