

ADOPTING SP/SI-FRAMES IN DUAL-BITSTREAM VIDEO STREAMING WITH VCR SUPPORT

Tak-Piu Ip, Yui-Lam Chan and Wan-Chi Siu

Centre for Multimedia Signal Processing
Department of Electronic and Information Engineering
The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

ABSTRACT

Digital video cassette recording (VCR) operations such as fast-forward and fast-reverse playbacks enable quick and user-friendly browsing of video. However, the predictive techniques adopted in current video standards severely complicate these operations. One approach to implement the fast-forward/reverse playback is to store an additional reverse-encoded bitstream into the server. Once a client requests a fast-forward/reverse operation, the server can select an appropriate frame for the client from either the forward or reverse-encoded bitstreams to reduce the network traffic and the decoder complexity. Unfortunately, the forward and reverse-encoded bitstreams are encoded separately. The frame that has previously decoded by the client may not be exactly identical to the reference of the current selected frame and the mismatch problem occurs frequently. In this paper, a novel H.264 dual-bitstream scheme aiming at providing fast-forward/reverse playback based on SP/SI-frames is proposed to eliminate mismatch errors during switching between the forward and reverse-encoded bitstreams. As a result, the proposed scheme enhances the performance of the conventional dual-bitstream scheme.

1. INTRODUCTION

With the proliferation of online video, it is also highly desirable that video streaming systems should have the capability of providing fast and effective browsing of video content. A key technique that facilitates this feature is to provide full VCR functionality. The set of VCR operations include forward, reverse, random access, fast-forward and fast-reverse. However, the prediction scheme adopted in current video coding standards [1-2] is mainly designed for normal playback in which frames are played back in a pre-determined order and it severely complicates the fast-forward/reverse playback since a predicted frame can only be decoded when its reference frame has already been decoded.

Recently, some works on the implementation of fast playback operations for video streaming have been introduced [3-4]. In [3], these operations are implemented by using frame-skipping video transcoding. But, it

introduces much computation in the server and will cause the drift due to the transcoding process. *Lin et al.* [4] recently proposed to store the forward-encoded bitstream (FB) and the reverse-encoded bitstream (RB) in the server. The idea behind is to switch frames between the FB and RB based on a frame-selection scheme which is used to minimize the transmitted frames over the network for any speed-up factors. This dual-bitstream technique can alleviate the decoder complexity while maintaining the low network bandwidth requirement in fast playback operations. However, this approach would cause the mismatch problem because the P-frame in the RB may have to be approximated by the I-frame in the FB and vice versa.

In this paper, we provide an efficient solution to perform bitstream switching in the dual-bitstream streaming system to eliminate the problem arising from the mismatch between the FB and RB. Our scheme borrows ideas from the SP/SI-frame concept [4-5] and makes some modification so that it serves as fast playback operations in digital VCR. Our scheme is designed specifically for H.264 and can be adapted to any other coding standards that provide SP/SI or similar coding standards.

The organization of this paper is as follows. Section 2 of this paper presents a study of the problems on the dual-bitstream scheme. In Section 3, a brief introduction of SP/SI-frame coding for H.264 is given and the proposed scheme is discussed in detail, where we describe how to adopt SP/SI-frames in the dual-bitstream scheme. Experimental results are presented in Section 4. Finally, some concluding remarks are provided in Section 5.

2. THE DUAL-BITSTREAM SCHEME

Figure 1 shows an illustrative example of the conventional dual bitstreams in which the video is coded in I/P-frames with a GOP size of 14 frames. Besides the FB, a reverse-encoded bitstream (RB) is stored. In this example, the coding arrangement of I-frames in the RB are interleaved between I-frames in the FB. Upon the server receives the VCR command or the requested frame number from the client, the server employs the frame-selection scheme to determine which frames in either the FB or RB should be transmitted to the client by minimizing the cost of decoding

the next requested frame. Generally, a larger number of frames to be sent requires much heavier network traffic and higher decoding complexity. To reduce all these in various VCR trick modes, the cost can be approximated to the number of frames to be sent over the network [3]. To illustrate the scheme, let us use the structure of dual bitstreams in Figure 1 again. Assume that the previous mode was in the normal forward play at frame 20 and the requested mode is fast-backward playback with a speed-up factor of 6. This operation requires to display frames 14, 8, etc. If the requested frame is an I-frame in one of two bitstreams, the frame can be decoded by itself. Thus, in the above example, frame 14 will be decoded from the FB directly since it is an I-frame. Then, the next frame to be decoded is frame 8. Since the requested frame is a P-frame in both bitstreams, the current displayed frame, or the nearest I-frame either in the FB or the RB is first selected to initiate the decoding of the requested frames. In this example, frame 8 will be decoded from frame 7 of the RB (an I-frame). It implies the nearest I-frame of the RB (frame 7) is the closest reference to frame 8. Note that frame 7 of the RB (an I-frame) is used as an approximation of frame 7 of the FB (a P-frame) to reconstruct frame 8 of the FB, as depicted in Figure 1. This I-to-P approximation will lead to the drift problem due to the mismatch between the reference frames. Furthermore, the drift will not only be confined to the frame at the switching location, but will further propagate in time if it is used as reference for other frames. This is the case when the current fast-backward mode is switched back to normal playback at frame 8. In this situation, frame 9 will be decoded by using the drifted frame 8. This drift continues occurring until the next I-frame in the FB.

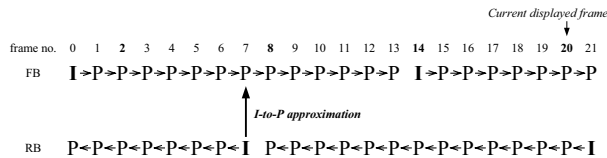


Figure 1. Fast-backward operation with the speed-up ratio of 6 in the conventional dual-bitstream video streaming system.

3. THE PROPOSED DUAL-BITSTREAM VCR SCHEME BASED ON SP/SI-FRAMES

In this paper, a novel dual-bitstream scheme aiming at avoiding the problem of reference mismatch at bitstream switching based on SP/SI-frames is proposed to maintain the quality of reconstructed frames during fast playback operations. The main feature of SP/SI-frames is that identical reconstruction can be achieved even when different reference frames are used for prediction. This property motivates us to adopt SP/SI-frames in the dual-bitstream scheme to prevent reference mismatch during switching from FB to RB or vice versa. In the proposed scheme, we make some modification to adopt an SP/SI-

frame pair at the switching point. This arrangement ensures that an I-to-P approximation will not happen at bitstream switching. Consequently, the new dual-bitstream scheme will not cause any mismatch at the decoder side.

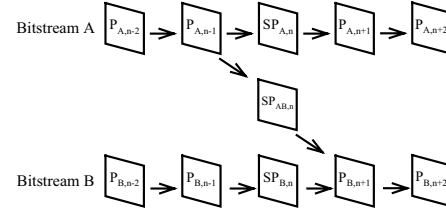


Figure 2. Bitstream switching using SP pictures.

3.1. Overview of SP/SI-Frame Coding in H.264

Bitstream switching is used in video streaming over a channel with bandwidth variation. This approach compresses each video sequence into multiple and independent bitstreams of different bitrates and then dynamically switches among these bitstreams. Owing to the temporal prediction used in current video standards, switching at a predicted frame (P-frame) would result in different references at the decoder, and such a mismatch would bring the so-called drifting error which could propagate and be accumulated in the subsequent frames until the next intra-frame (I-frame). For this reason, the newest H.264 standard introduces SP/SI-frame coding to enable seamless bitstream switching capability. The main feature of SP/SI-frames is that identical frame can be reconstructed even when different reference frames are used for their prediction. An example of bitstream switching using SP-frames is shown in Figure 2, where an image sequence is coded into two bitstreams (bitstream A and bitstream B) with different bit rates and quality level. Within each encoded bitstream, two SP-frames - SP_A and SP_B , as shown in Figure 2 are placed at the locations at which switching from one bitstream to another will be allowed. These SP-frames are referred to as primary SP-frames. Besides, for each primary SP-frame, a corresponding secondary SP-frame (SP_{AB} in Figure 2) is generated, which has the same reconstructed values as the primary SP-frame. Such a secondary SP-frame is sent only during bitstream switching. For normal transmission, either bitstream A or bitstream B is sent to the user depending on the current available bandwidth. When there is a need to switch the transmitting bitstream from bitstream A to bitstream B at frame n , $SP_{AB,n}$ instead of $SP_{B,n}$, is transmitted. After decoding $SP_{AB,n}$, the decoder can obtain exactly the same reference as normally $SP_{B,n}$ decoded at frame n , therefore it can continually decode bitstream B at frame $n+1$ seamlessly. The way to encode SP_{AB} ensures that an identical reconstruction as that of the SP_B can be obtained by decoding it so that the bitstream switching process will not introduce any mismatch between the encoder and decoder.

3.2. Motivation of using SP/SI-Frames in Dual-bitstream VCR scheme

As mentioned in Section 2, the conventional dual-bitstream scheme suffers the drift problem due to the I-to-P frame approximation when bitstream switching is needed. We consider in this paper by adopting the concept of SP/SI-frame coding in the dual-bitstream scheme to eliminate the problem of reference mismatch at switching between the FB and RB.

frame no.	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
FB	SI	P	P	P	P	P	P	P	P	P	P	P	P	P	P	SI	P	P	P	P	P	P
RB	SP	P	P	P	P	P	P	P	P	P	P	P	P	P	P	SP	P	P	P	P	P	P

Figure 3. The structure of the proposed dual bitstreams.

The idea of utilizing SP/SI-frame coding for the switching purpose in VCR operations is shown in Figure 3. In the original dual bitstreams, I-frames represent the points of access to decode the sequence from any arbitrary position. In order to avoid the problem of reference mismatch, instead of using I-frames in the conventional dual bitstreams, SI-frames are adopted at the switching point. An SI-frame can be placed either in the FB or RB which is called as a forward-encoded SI-frame (SI_{FB}) or a reverse-encoded SI-frame (SI_{RB}) respectively. For each SI_{FB} , there is a corresponding reverse-encoded SP-frame (SP_{RB}) and these two frames form an SI_{FB}/SP_{RB} pair. Similarly, there is a corresponding forward-encoded SP-frame (SP_{FB}) for each SI_{RB} and they form an SP_{FB}/SI_{RB} pair. Identical frame reconstruction can be allowed at these SI_{FB}/SP_{RB} and SP_{FB}/SI_{RB} pairs when switching between the FB and RB. This facilitates the correctly predicted frame to be used if switching from FB to RB or vice versa is needed in any VCR operations and hence no drift error will be occurred. To illustrate the proposed scheme, we use the example in Section 2 again in which the current VCR is in the fast-backward mode and this operation requires to display frames 14, 8, etc. After decoding and displaying frame 8, then a normal-play request is launched. Frame 8 will be decoded from frame 7 of the RB, which is an SI_{RB} . After decoding this SI_{RB} , the decoder can obtain exactly the identical reconstruction as normally decoding SP_{FB} at frame 7 in the FB, therefore it can decode the FB at frame 8 without any mismatch error and continue to decode the subsequent frames seamlessly when normal playback is requested at frame 8. Therefore, the proposed dual-bitstream scheme does not suffer the drift problem because approximation between the SI-frame and the SP-frame on the dual bitstreams do not exist.

3.3. Encoding of SP_{FB}/SI_{RB} and SI_{FB}/SP_{RB} Pairs

In this section, we will provide a detailed description of how to encode the SP_{FB}/SI_{RB} and SI_{FB}/SP_{RB} pairs. All forward SP/SI-frames (SP_{FB} and SI_{FB}) are encoded as primary SP/SI-frames. For instance, the encoding of SP_{FB} is the same as that of the P-frame except the primary SP-frame is

undergone an additional quantization process, as depicted in Figure 4. This additional quantization process provides a footprint for the reconstruction of SP_{FB} at a certain quantization level Q_s . By finding the difference between SP_{FB} and the prediction of SI_{RB} at the quantization Q_s domain, the exact reconstruction of SP_{FB} is achieved by using the prediction of SI_{RB} . Figure 5(a) shows the simplified block diagram for encoding and decoding of SI_{RB} . By using the same example in Figure 3, the exact reconstruction of the forward SP-frame ($SP_{FB,n}$) is being coded instead of the original video frame in the RB. Note that the prediction is formed using the 4×4 Intra prediction modes from previously-decoded samples of the reconstructed frame. Then the bitstream of $SI_{RB,n}$ becomes

$$Bitstream(SI_{RB,n}) = VLC\{SP_{FB,n} - Q_s[T[Intrapred(SI_{RB,n})]]\} \quad (1)$$

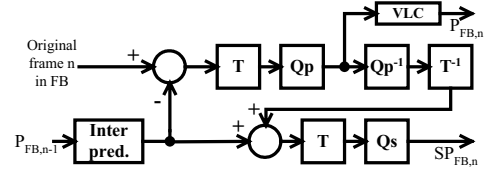


Figure 4. Encoding of the forward SP-frame.

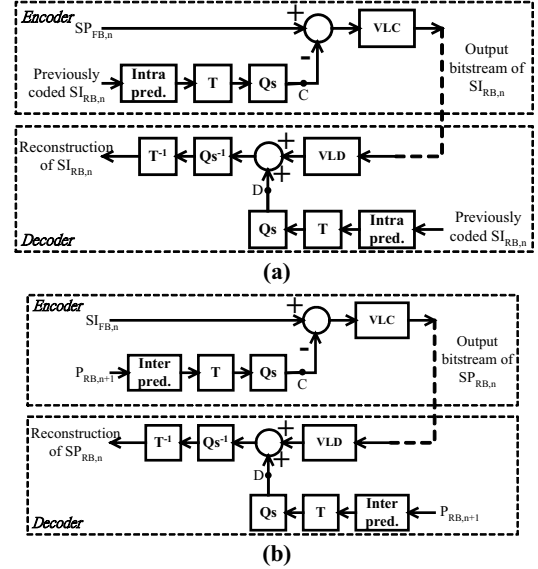


Figure 5. Encoding and decoding of (a) SI_{RB} in SP_{FB}/SI_{RB} and (b) SP_{RB} in SI_{FB}/SP_{RB} .

As compared with general H.264 coding, another difference of this new coding method is that quantization is done before calculating the residuals. The quantization used has to be equal to Q_s set by the forward SP-frame coding as shown in Figure 4. This special arrangement to encode macroblocks in $SI_{RB,n}$ is to guarantee that its reconstruction can be identical to the reconstruction of $SP_{FB,n}$, which can be explained as follows:

1. The transformed and quantized different predictions will not introduce any mismatch, that is, the value at the

position C in the encoder and at the position D in the decoder are exactly the same.

2. The reconstruction value can be quantized and de-quantized without any loss at the quantization level, Q_s , no quantization error will be introduced in the process.

For generating SI_{FB} at frame n ($SI_{FB,n}$) in SI_{FB}/SP_{RB} pair, it is similar to encode $SP_{FB,n}$ and the only difference is that $SI_{FB,n}$ does not use any reference frames, the macroblocks are coded by using intra prediction, which means that in the encoding diagram shown in Figure 4, the prediction is generated by intra prediction from neighboring blocks. To achieve identical reconstruction, Figure 5(b) shows how to encode SP_{RB} in the SI_{FB}/SP_{RB} pair, the reference it uses is the reconstructed frame at frame $n+1$ in the RB and the reconstruction value of $SI_{FB,n}$ is the input. Hence, the bitstream of $SP_{RB,n}$ is written as

$$Bitstream(SP_{RB,n}) = VLC\{SI_{FB,n} - Q_s[T[Interpred(P_{RB,n+1})]]\} \quad (2)$$

4. EXPERIMENTAL RESULTS

In this section, we present some experimental results. All the test sequences have a length of 200 frames and were encoded at different bitrates. Each test sequence is encoded into two bitstreams, FB and RB, and I- or SI-frames in the RB are interleaved between I- or SI-frames in the FB in the conventional and proposed dual-bitstream schemes respectively. For all test sequences, the frame-rate of the video stream was 30 frames/s and the GOP length is 60 without B-frames.

Table 1 gives the results of average PSNR values for different schemes on every possible switching points between the FB and RB. The experimental results indicate that our proposed scheme has significant improvement for all video sequences. The results are more noticeable for the sequences “Carphone” and “Foreman” which has PSNR improvement over 1dB. Figure 6 demonstrates the performance of two different schemes in the situation where the current VCR is in the fast-reverse mode and then a normal-play request is launched. Assume that the user requests a normal-play request at the start of each GOP in the RB. In this case, it has longest duration of drift propagation in the conventional dual-bitstream scheme due to the I-to-P approximation. As shown in Figure 6, the performance of the proposed scheme outperforms the conventional one and it follows the same PSNR of the FB. This result indicates that the visual quality of the reconstructed frames in the proposed scheme for various VCR operations is exactly identical to that of normal playback.

5. CONCLUSIONS

In this paper, we proposed to adopt the concept of SP/SI-frame coding to improve the performance of the dual-bitstream scheme. This new scheme can completely eliminate the problem of reference mismatch and then stop

the drift error. Experimental results show that the visual quality of the reconstructed frames in our new scheme during various VCR operations will be exactly the same as that of the normal playback. Therefore, the proposed dual-bitstream scheme with SP/SI-frames is a promising solution for H.264 video streaming with VCR functionality.

Table 1. Average PSNR performance for conventional and proposed dual-bitstream schemes on every possible switching between the FB and RB.

Sequences	Format	Conventional		Proposed	
		FB→RB	RB→FB	FB→RB	RB→FB
Salesman	(352×288)	33.997	34.161	34.441	34.66
Foreman	(352×288)	34.475	34.686	35.733	35.784
Football	(352×240)	33.107	33.212	33.85	33.916
TableTennis	(352×240)	32.176	32.335	32.894	32.988
Carphone	(176×144)	34.218	34.583	35.536	35.642

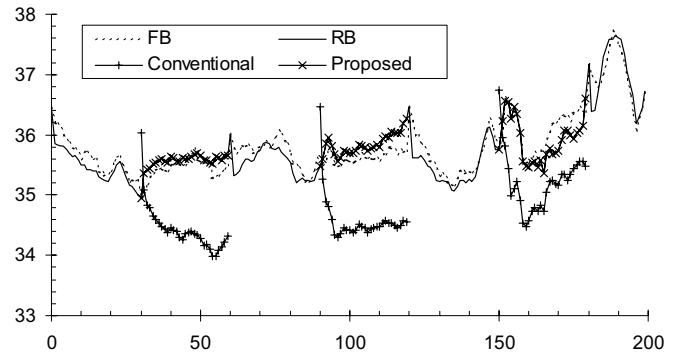


Figure 6. PSNR performance of two different schemes in the situation where the current VCR is in the fast-backward mode and then a normal-play request is launched at the start of each GOP in RB for the “Foreman” sequence.

6. ACKNOWLEDGMENTS

The work described in this paper is partially supported by the Centre for Multimedia Signal Processing, Department of Electronic and Information Engineering, Hong Kong Polytechnic University and a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China(PolyU 5204/04E). Tak-Piu Ip acknowledges the research studentships provided by the University.

7. REFERENCES

- [1] ISO/IEC 14996-2, “Coding of Audio-visual Objects – Part 2: Visual,” 2001.
- [2] ISO/IEC 14996-10 and ITU-T Rec. H.264, “Advanced Video Coding,” 2003.
- [3] Y.-P. Tan, Y. Liang, and J. Yu, “Video transcoding for fast forward/reverse video playback,” in Proc. International Conference on Image Processing, ICIP’2002, pp. 713-716, 22-25 Sept., 2002.
- [4] C. W. Lin, J. Zhou, J. Youn, and M. T. Sun, “MPEG video streaming with VCR functionality,” IEEE Trans. on Cir. Sys. for Video Tech., pp. 415-425., Vol. 11, No. 3, March 2001.
- [5] Marta Karczewicz and Ragip Kurceren, “The SP- and SI-Frames Design for H.264/AVC,” IEEE Trans. on Cir. and Sys. for Video Tech., Vol. 13, No. 7, pp. 637-644, July 2003.