EFFICIENT AND FLEXIBLE DRIFT-FREE VIDEO BITSTREAM SWITCHING AT PREDICTIVE FRAMES

Xiaoyan Sun[*1], *Shipeng Li*[2], *Feng Wu*[2], *Guobin Shen*[2], *Wen Gao*[1]

[1] Department of Computer Application, Harbin Institute of Technology, Harbin, 150001

[2]Microsoft Research Asia, Beijing, 100080

ABSTRACT

In this paper, we propose an efficient and flexible coding scheme inspired by the SP picture technique in H.26L TML, which can achieve drift-free bitstream switching at predictive frames. Firstly, the proposed scheme improves the coding efficiency of the SP frames in H.26L TML by (1) reducing the number of quantization modules in the encoding path; (2) eliminating the mismatch between references for the prediction and the reconstruction; (3) outputting a high quality image for display purpose before the quantization step in the reconstruction loop. Secondly, the proposed scheme allows independent quantization parameters for up-switching and down-switching bitstreams. It can further reduce the switching bitstream size while keeping the coding efficiency of the normal bitstreams. It allows more rapid and frequent down-switching than up-switching. Furthermore, the size of the down-switching bitstream can be much smaller than that of the up-switching one. This is a very desirable feature for any TCP-friendly protocols currently used in most existing streaming systems.

1. INTRODUCTION

With the steady growth of the access bandwidth, more and more Internet applications start to use streaming audio and video contents [1][2]. Since the Internet is inherently a heterogeneous and dynamic best-effort network, the effective channel bandwidth usually fluctuates in a wide range from bit rate below 64kbps to well above 1 Mbps. Since the traditional video coding technologies usually generate bitstreams at fixed bit rates, a simple method to achieve bandwidth adaptation in the streaming applications is to produce multiple and independent bitstreams at different bit rates and dynamically switch among them to accommodate the bandwidth variations. Such a scheme is extensively used in many commercial video streaming systems.

In general, the switching between these bitstreams is only allowed at key frames where the encoding does not depend on information from any previous frames, e.g., at I frames. This would ensure that no drifting error is introduced from bitstream switching. However, it is well known that I frames cost much more bits than the predictive frames to achieve the same decoded quality. To achieve rapid and random switching between bitstreams by frequently inserting I frames would inevitably decrease greatly the coding efficiency of each of the bitstreams. Therefore, an interesting research topic is how to flexibly switch from one bitstream to another without causing any drifting error while keeping the high coding efficiency. A natural question comes out: can we switch at predictive frames?

Since a P frame is always encoded using the prediction from the previous reconstructed reference, without any special treatment, the switching between bitstreams at a P frame would lead to drifting errors due to the mismatch of the reconstructed references at that frame. Furthermore, such mismatch errors could propagate and be accumulated in the subsequent P frames until another I frame. Such errors are often referred to as drifting errors. Drifting errors would rapidly deteriorate the decoded visual quality as the number of frames increases. In order to eliminate such mismatch, one method is to losslessly compress the difference between two reconstructed references into an extra switching bitstream at the switching P frames [3][4]. Drifting-free switching can be achieved at these frames by transmitting the extra bitstream.

A novel switching technique was proposed in [5][6] to achieve drifting-free switching at predictive frames. It has been accepted by H.26L TML as a new frame type: SP. Figure 1 illustrates the process of seamlessly switching from one bitstream to another through SP frames. There are two bitstreams: Bitstream 1 and Bitstream 2. In general, the bit rate of Bitstream 1 is different from that of Bitstream 2. The frames t-1 and t are compressed as normal P frames in both Bitstream 1 and Bitstream 2. At the time t, S_1, S_2 and S_{12} are compressed as SP frames, where a switching point is provided for switching from Bitstream 1 to Bitstream 2 and vice versa. Assume that Bitstream 1 is being transmitted to the user. When there is a need to switch to Bitstream 2, S_{12} instead of S_1, however, is transmitted at time t. After S_{12} decoding, the decoder can obtain exactly the same reference as normally decoding S_2 at time t, therefore it can continue to decode Bitstream 2 at time t+1 seamlessly.

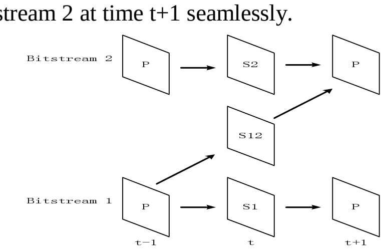


Figure 1: The switching from Bitstream 1 to Bitstream 2 through SP frames

Similar to P frame, SP frame exploits temporal redundancy by using motion compensated predictive coding. The bits needed for switching through an SP frame are far less than that through

[*] This work has been done while the author is with Microsoft Research Asia.

an I frame or an extra bitstream. Bitstreams for S_2 and S_{12} in Figure 1 are generated with different references, but their reconstructions are identical. Therefore, the SP coding scheme can switch between bitstreams without any drifting errors. However, there are drawbacks in the SP coding scheme used in TML preventing it from further improved performance [7]. Firstly, there are three quantization operations in the TML SP encoder. Each quantization operation could potentially degrade the decoded video quality. Secondly, the SP encoder uses different references for prediction and reconstruction. Although this does not cause error propagation, it could reduce the coding efficiency. Thirdly, the display image at an SP frame is output after a quantization operation; therefore, its quality is also degraded.

As an improvement, we proposed an alternative coding scheme for seamlessly switching between video bitstreams in [8]. The preliminary results showed a promising gain in coding efficiency. Combined with our previous studies, a new scheme is proposed in this paper that not only overcomes all the above drawbacks in TML SP coding scheme but also improves the SP coding from other aspects, such as separating the quantizations for up-switching and down-switching, decoupling the up-switching and down-switching points, further reducing the switching bitstream size and removing unnecessary quantization processes, etc.

This paper is organized as follows. Section 2 describes the proposed SP coding scheme in detail. The advantages provided by the proposed scheme are discussed in Section 3. Experimental results are given in Section 4. Finally, Section 5 concludes this paper.

2. THE PROPOSED SP CODING SCHEME

As shown in Figure 1, the proposed SP coding scheme uses the same method to achieve bitstream switching except that the picture type at the switching point for the origin bitstream can also be P picture now. In other words, S_1 in Figure 1 can be encoded as an SP frame or a normal P frame. Figure 2 gives the block diagram of compressing S_1, S_2 and S_{12} using the proposed scheme, where all of them are encoded as SP frames.

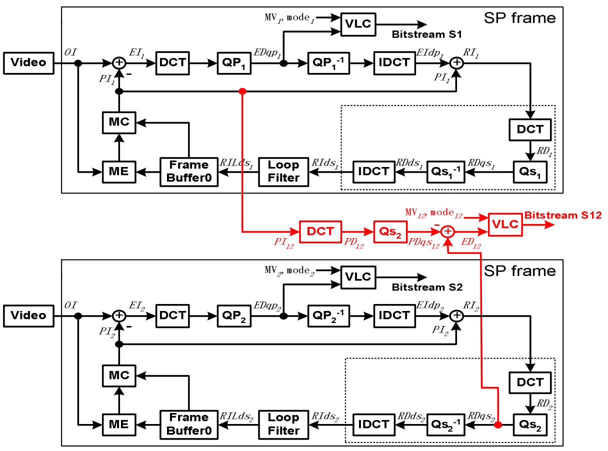


Figure 2: The block diagram of compressing S_1, S_2 and S_{12}.

For convenience, we use the terminology with two mandated letters and three optional letters to denote the signals at each processing stage and it is specified as follows:

- First mandated letter
  - *O*: Original picture, macroblock and block
  - *P*: Predicted picture, macroblock and block
  - *R*: Reconstructed picture, macroblock and block
  - *E*: Predicted error picture, macroblock and block
- Second mandated letter
  - *D*: DCT domain
  - *I* : Pixel domain
- First optional letter
  - *L*: Reconstructed picture, macroblock and block after loop filtering
- Second optional letter
  - *q*: Quantization
  - *d*: Dequantization
- Third optional Letter
  - *p*: Quantization parameter specified by QP
  - *s*: Quantization parameter specified by Qs

For easy understanding of the above definitions, two examples are given here. "*OD*" means the original picture, macroblock or block after the DCT transform. "*RDLqp*" means the reconstructed picture, macroblock or block after the DCT transform, loop filtering and quantization with QP.

How to generate and decode the Bitstreams in S1, S2 and S12 is described as follows.

The encoding process (as shown in Figure 2)

1. The encoding of SP frame S1 or S2 in a normal bitstream (Bitstream 1 or Bitstream 2) (use S1 as an example)
   a) After motion compensation, denote the obtained prediction macroblock as PI_1.
   b) Subtract PI_1 from the original macroblock OI_1, and denote the obtained predictive errors EI_1.
   c) Perform DCT transform and quantization on EI_1 with QP_1, and denote the obtained levels as $EDqp_1$. They are then compressed into Bitstream S_1 with entropy coding (UVLC or CABAC).
   d) Perform dequantization and inverse DCT on $EDqp_1$, and denote the reconstructed error macroblock as $EIdp_1$.
   e) Add $EIdp_1$ to PI_1, and denote the obtained reconstructed macroblock as RI_1.
   f) Perform DCT transform on RI_1, and denote the obtained DCT coefficient as RD_1.
   g) Perform quantization on RD_1 with Qs_1, and denote the obtained levels as $RDqs_1$.
   h) Perform dequantization on $RDqs_1$ with Qs_1, and denote the obtained coefficient as $RDds_1$.
   i) Perform inverse DCT transform on $RDds_1$, and denote the obtained macroblock as $RIds_1$.
   j) After loop filtering, the reconstructed macroblock $RILds_1$ is used to update the frame buffer.

   Note that except steps f) to i), the S_1/S_2 encoding is exactly the same as in a normal P picture encoder.
2. The encoding of switching bitstream S_{12} (switching from Bitstream 1 to Bitstream 2). The encoding of S_{12} is based on the encoding of S_1 and S_2.
   a) There are two cases. In most applications, the input PI_{12} can be directly obtained from the prediction PI_1 in the S_1 or P encoder, that is, $PI_{12}=PI_1$; in other applications, there might be a need to perform new

motion estimation and compensation referencing a previous frame in Bitstream 1, and we denote the obtained prediction as PI_{12}.

b) Perform DCT transform on PI_{12}, and denote the obtained coefficients as PD_{12}.

c) Perform quantization on PD_{12} with Qs_2, and denote the obtained levels as $PDqs_{12}$.

d) Subtract $PDqs_{12}$ from $RDqs_2$ in S_2 encoder, and denote the obtained error coefficients as ED_{12}.

e) Compress ED_{12} into Bitstream S_{12} with entropy coding (UVLC or CABAC).

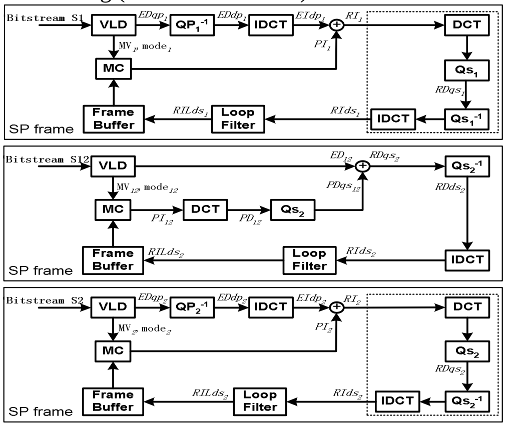


Figure 3: The decoder for SP frames S_1, S_2 and S_{12}.

The decoding process (as shown in Figure 3)

1. The decoding of SP frame S_1 or S_2 in a normal bitstream (Use S_1 as an example.):
   a) After entropy decoding of the Bitstream S_1, denote the obtained error levels as $EDqp_1$. Macroblock modes and motion vectors can also be obtained meanwhile.
   b) Perform dequantization on $EDqp_1$ with QP_1, and denote the obtain coefficients as $EDdp_1$.
   c) Perform inverse DCT transform, and denote the obtained macroblock as $EIdp_1$.
   d) After motion compensation, denote the obtained prediction as PI_1.
   e) Add $EIdp_1$ to PI_1, and denote the obtained macroblock as RI_1. RI_1 is going through an optional deblocking filter and output for display purpose.
   f) Perform DCT transform and quantization on RI_1 with Qs_1, and denote the obtained levels as $RDqs_1$.
   g) Perform dequantization with Qs_1 and inverse DCT transform on $RDqs_1$, and denote the obtained macroblock as $RIds_1$.
   h) After loop filtering, the obtained macroblock $RILds_1$ is used to update the frame buffer.

   Note that except steps f) and g), the S_1/S_2 decoding is exactly the same as in a P picture decoder. Also, the display image is output before further quantization processes, therefore, it has much better visual quality.

2. The decoding of switching bitstream S_{12} (switching from Bitstream 1 to Bitstream 2).
   a) After entropy decoding of the bitstream S_{12}, denote the obtained error levels as ED_{12}. Macroblock modes and motion vectors can also be obtained meanwhile.
   b) After motion compensation, perform forward DCT transform for the predicted macroblock and obtain PI_{12}.
   c) Perform DCT transform on PI_{12}, and denote the obtained coefficients as PD_{12}.
   d) Perform quantization on PD_{12} with Qs_2, and denote the obtained levels as $PDqs_{12}$.
   e) Add ED_{12} to $PDqs_{12}$, and denote the obtained levels as $RDqs_2$.
   f) Perform dequantization on $RDqs_2$ with Qs_2, and denote the obtained coefficients as $RDds_2$.
   g) Perform inverse DCT transform on $RDds_2$, and denote the obtained macroblock as $RIds_2$.

   After loop filtering, the obtained macroblock $RILds_2$ is used to update the frame buffer and output for display.

3. THE ADVANTAGES OF THE PROPOSED SCHEME

The proposed scheme provides a few significant advantages over the SP scheme in H.26L TML. Firstly it simplifies the implementation. The encoder and decoder structures are similar to that for a normal P frame, thus many common modules can be shared with P frames. Due to 4x4 integer DCT transform, H.26L adopts integrated DCT/quantization and dequantization/IDCT modules to reduce rounding errors. In the proposed SP coding scheme, each DCT module is closely coupled with a quantization module; each IDCT module is closely coupled with a dequantization module. The integrated implementation in H.26L can be readily applied here.

Secondly, the proposed scheme improves the coding efficiency of the normal bitstream. It reduces the number of quantization modules in the encoding path and thus the quantization errors can be reduced. It eliminates the mismatch between references for the prediction and the reconstruction, therefore, the quality of the reconstructed image can be improved. Furthermore, a more important advantage of the proposed scheme that distinguishes itself from others is that it can output a high quality display image before the quantization process in the reconstruction loop. This will significantly improve the video quality if many switching points are inserted in the normal bitstream.

In real streaming applications, it is much desired to be able to switch down from a high bit-rate bitstream to a low bit-rate one very quickly. This is a desirable feature for any TCP-friendly protocols currently used in most existing streaming systems. On the other hand, switching up from a low bit-rate video bitstream to a high bit-rate one does not have to be done as quickly as switching down. This is again a feature of the TCP-friendly protocols. Therefore, we would expect more rapid and frequent down-switching than up-switching, that is, more down-switching points than up-switching points. Moreover, the very reason for down-switching is because that the channel bandwidth is not enough. Therefore, the size of the down-switching bitstream should be much smaller than that of the up-switching one. Since the high bit-rate bitstream should contain most of the information of a low bit-rate one, in theory, we could find a scheme that can make the size of switching bitstream very small.

However, the current TML SP coding scheme only allows the same parameter Qs for both the down-switching and up-switching bitstreams. Since Qs is included in the prediction and reconstruction loop, this will inevitably degrade the coding efficiency compared with the original bitstreams without SP frames. If we set Qs too small, we could get high coding efficiency for both Bitstreams 1 and 2. However, the difference for down-switching is also fine-grain quantized and it would result a very large down-switching bitstream. On the other hand, if we set Qs too large, although we can obtain a very compact switching bitstream, the coding efficiency of Bitstreams 1 and 2 will be severely degraded, which is not desired either. The contradiction can not be solved by the current TML SP coding scheme. It always has to make a compromise between coding efficiency and the size of the switching bitstream.

However, in the proposed scheme, different quantization parameters Qs can be used in Bitstreams 1 and Bitstream 2. Furthermore, Bitstream S_{12} for switching from Bitstream 1 to Bitstream 2 is only related to Qs_2. Similarly, Bitstream S_{21} for switching from Bitstream 2 to Bitstream 1 is related to Qs_1. By optimizing the parameters Qs_1 and Qs_2, the proposed scheme completely solves the contradiction so that the down-switching bitstream can be encoded to have minimal size while the coding efficiency of Bitstreams 1 and 2 is well preserved.

Another significant advantage of such a scheme is that the switching points for up-switching and down-switching can be decoupled too. This means that we can encode more down-switching points than up-switching points to suit the TCP-friendly protocols. Moreover, such a decoupling design allows us to further improve the coding efficiency of the origin bitstream which the system is switched from by independently setting the Qs in the reconstruction loop to a very small value or simply encoding it as a P frame whenever a frame is not a switching point from other bitstreams.

4. EXPERIMENTAL RESULTS

The coding efficiency of the proposed scheme and the H.26L TML SP coding scheme are evaluated. The TML 9.4 software is used in this experiment. The sequences Foremen and Coastguard with QCIF format are encoded at 10Hz. Only the first frame is an I frame and all other frames are SP frames. The quantization parameter Qs is equal to QP. Other parameters are given as follows:

- RD optimization: Enable
- Hadamard transform: Enable
- Search Range: 16
- MC: ¼ pixel
- Reference number: 1
- B frame: No
- Inter and Intra mode: All
- Entropy coding: UVLC

The experimental results are given in Figure 4. Each curve consists of six points corresponding to the quantization parameters 13, 16, 19, 22, 25 and 28. The proposed SP coding scheme can improve the coding efficiency up to 1.0dB for Foreman and up to 0.5 dB for Coastguard.

5. CONCLUSIONS

This paper proposed an efficient and flexible scheme to achieve seamless video bitstream switching at predictive frames without drifting errors. By reducing the number of quantization modules in the encoding path, eliminating the mismatch between references for prediction and reconstruction and outputting a high quality display image before the quantization step in the reconstruction loop, the proposed scheme can improve the coding efficiency up to 1.0dB for Foreman and up to 0.5dB for Coastguard. Furthermore, by decoupling different quantization parameters for switching up and switching down, the proposed scheme can encode more switching-down points than switching-up points. At the same time, the size of the down-switching bitstream can be much smaller than that of the up-switching one. These features are very desirable for the TCP-friendly protocols currently used in most existing streaming systems.

References

[1] J. Lu, “Signal processing for Internet video streaming: A review”, SPIE in Image and Video Communication and Processing 2000, Vol 3974, 246-258, 2000.

[2] W. Li, “Streaming video profile in MPEG-4”, IEEE trans. Circuits and Systems for Video Technology, Vol 11, no 3, 301-317, 2001

[3] G. Conklin, G. Greenbaum, K. Lillevold, A. Lippman and Y. Reznik, “Video Coding for Streaming Media Delivery on the Internet”, IEEE trans. Circuits and Systems for Video Technology, Vol. 11, no 3, 269-281, 2001

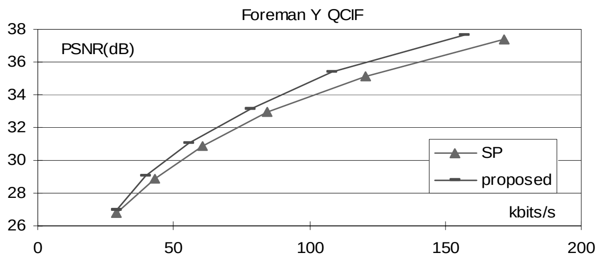
[4] B. Girod, N. Farber, and U. Horn, “Scalable codec architectures for Internet video on demand”, in Proc. 1997 Asilomar Conf. Signals and Systems, USA, Nov 1997.

[5] M. Jarczewicz and R. Kurceren, “A Proposal for SP frames”, ITU-T Q.6/SG 16, VCEG-L27, Eibsee, Germany, Jan 2001.

[6] R. Kurceren and M. Jarczewicz, “Improved SP-frame encoding”, ITU-T Q.6/SG 16, VCEG-M73, Austin, USA, April 2001.

[7] T. Wiegand, “H.26L test model long-term number 9 (TML-9)”, ITU-T Q.6/SG 16, VCEG-N83d1, Pattaya, Thailand, December 2001.

[8] X. Sun, F. Wu, S. Li, W. Gao, and Y.-Q. Zhang, “Improved SP coding technique”, ITU-T Q.6/SG 16, JVT-B097, Geneva, Switzerland, Jan 2002.



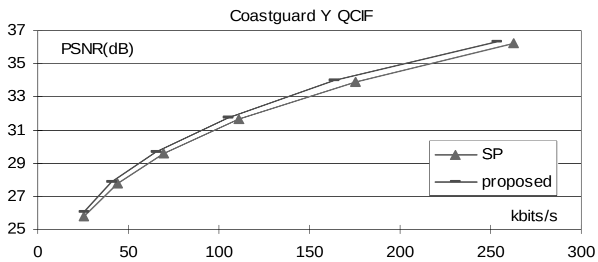


Figure 4: The comparisons between the H.26L SP coding scheme and the proposed scheme.