THE IMPROVED SP FRAME CODING TECHNIQUE FOR THE JVT STANDARD

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

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

[2] Microsoft Research Asia, Beijing, 100080

ABSTRACT

An efficient and flexible coding technique is proposed in this paper inspired by the SP frame in the H.26L standard, which can achieve a drift-free bitstream switching at the predicted frame. The proposed scheme improves the coding efficiency of the SP frames in the H.26L standard by limiting the mismatch between the references for the prediction and reconstruction with two DCT coefficient coding modes and the rate-distortion optimization. Furthermore, the proposed scheme allows independent quantization parameters for up-switching and down-switching bitstreams. It further reduces the switching bitstream size while keeping the coding efficiency of the normal bitstreams. More rapid and frequent down-switching than up-switching and much smaller size of down-switching bitstream can be achieved with the proposed SP technique. These are very desirable features for any TCP-friendly protocols. Compared with the original SP method for H.26L, the proposed SP method improves the coding efficiency up to 1.0dB. This SP technique has been officially accepted by the JVT standard.

1. INTRODUCTION

Due to the explosive growth and great success of the Internet, as well as the increasing demands on video services, streaming video over the Internet has drawn tremendous attention in both academia and industry [1]. However, the Internet, which was initially designed for data transmission and communication among computers, is inherently a heterogeneous and dynamic network. The channel bandwidth may usually fluctuate in a wide range from bit rates below 64 kbps to well above 1 Mbps. The traditional video coding technologies usually generate bitstreams at fixed bit rates, therefore 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 [2][3]. Such a scheme is extensively used in many commercial video streaming systems.

In general, the switching between these bitstreams is restricted only on *key frames*, at which the encoding does not depend on information from any previous frames, e.g., at I frames, to avoid severe drifting problem. However, the more I frames, the worst coding efficiency. As a result, key frames are normally encoded periodically far apart from each other and bitstream switching can not take place at any desired frames. This greatly reduces the flexibility of existing streaming systems on bandwidth adaptation.

H.26L adopts a SP coding method that allows seamless switching between bitstreams with different bit-rate at predictive frames [4]~[6]. Figure 1 illustrates the general scenario for the SP method. 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. Normally, the streaming system either transmits bitstream 1 or bitstream 2 depending on the current available bandwidth. Assume that Bitstream 1 is being transmitted to the user. When there is a switch to Bitstream 2, S_{12} instead of S_1, is transmitted at time t. After S_{12} decoded, the decoder can obtain exactly the same reference as normally S_2 decoded at time t, therefore it can continually decode Bitstream 2 at time t+1 seamlessly.

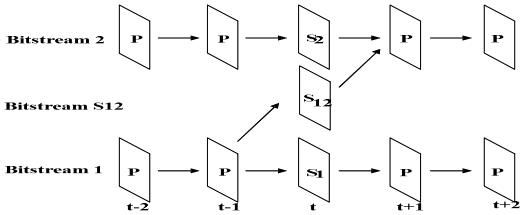

Figure 1: The Switching from Bitstream 1 to Bitstream 2 through SP frames.

Since the temporal redundancy is exploited by motion compensated predictive coding in the H.26L SP method, the bits needed for switching through an SP frame are far less than that through an I frame or an extra bitstream. Moreover, it is the identical decoder in processing S_1, S_2 and S_{12}, which makes the original SP scheme very simple. However, there are some potential problems with the original H.26L SP scheme preventing it from further improved performance [7], especially on coding efficiency and streaming quality.

As an improvement, we proposed an alternative coding scheme for seamlessly switching between video bitstreams in [7][8]. The preliminary results showed a promising gain in coding efficiency and flexibility. Taking advantages of the original SP coding method and combined with our previous ideas, a new scheme is proposed in this paper that not only overcomes the above drawbacks existing in the H.26L SP coding scheme but also improves the SP coding from many aspects, such as providing better coding efficiency with two coefficient predictive coding modes and the rate-distortion optimization, separating the quantization parameters 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. The proposed SP scheme in this paper has been accepted by the JVT standard.

This paper is organized as follows. Section 2 describes the normal SP frame coding in the proposed scheme in detail. The generation and decoding of the switching SP frame in the proposed scheme are discussed in Section 3. Experimental results are given in Section 4. Finally, Section 5 concludes this paper.

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

0-7803-7750-8/03/$17.00 ©2003 IEEE.

2. THE PROPOSED CODING SCHEME FOR NORMAL SP FRAME

The block diagram of the normal SP frame coding in the original H.26L encoder is shown in Figure 2. There are some problems in this scheme.

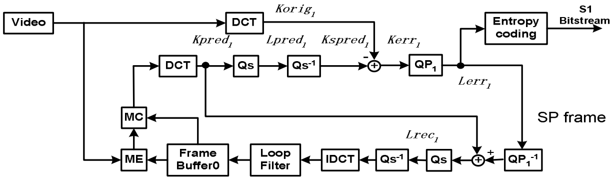

Figure 2: The normal SP frame coding in H.26L.

Firstly, there are many quantization and de-quantization processes in the signal flow in the H.26L SP encoder. Each quantization operation could potentially degrade the coding efficiency of Bitstreams 1 and 2. Moreover, the mismatch between the prediction reference and reconstruction reference in the SP scheme will hurt the coding performance of Bitstream 1 and 2. Additionally, the display image at an SP frame is output after a quantization operation. The decoded quality is further degraded.

Secondly, in the H.26L SP scheme, *Qs*, the quantization parameter for switching purpose, is always included in the prediction and reconstruction loop. It only allows the same *Qs* for both the down-switching bitstream and the up-switching bitstream. Therefore, if the *Qs* is too small, the coding efficiency for both Bitstreams S1 and S2 is good, but meanwhile resulting in a very large down-switching bitstream. If *Qs* is too large, a very compact switching Bitstream S12 is generated but the coding efficiency of Bitstreams S1 and S2 will be severely degraded. It is very difficult for the original SP scheme to solve this contradiction.

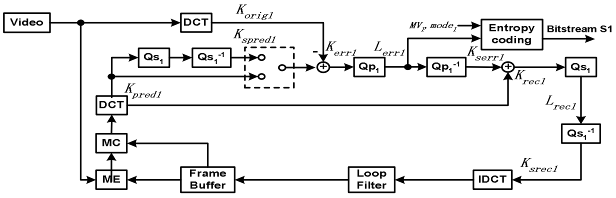

Figure 3: The proposed encoder for normal SP.

For providing higher coding performance and better flexibility, an improved SP coding scheme is proposed in which the same method to achieve bitstream switching in H.26L SP is utilized except that the picture type at the switching point for the original 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 3 gives the block diagram of compressing the normal SP frames, S_1 and S_2, using the proposed scheme.

How to generate and decode the Bitstreams S1 and S2 for inter-macroblocks is described as follows. For intra-macroblocks, simple "copy" operation can be used. *Qp* is the quantization parameter as same as that for normal P frame, while *Qs* is the quantization parameter used for SP switching.

The proposed encoding process (as shown in Figure 3)

The encoding of normal SP frame S_1 or S_2 in a normal bitstream (Bitstream S1 or Bitstream S2) (use S_1 as an example):

a) Perform DCT transform on the macroblock of the original video, and denote the obtained coefficients as K_{orig1}.

b) After motion compensation, perform DCT transform to the predicted macroblock, and denoted the obtained coefficients as K_{pred1}.

c) Quantize and de-quantize K_{pred1} using Qs_1, and denoted the obtained coefficient as K_{spred1}

d) Either K_{pred1} or K_{spred1} is subtracted from K_{orig1}, according to the current coefficient predictive coding mode determined by a rate-distortion optimized mode decision. The obtained error coefficients K_{err1} is

$$K_{err1} = K_{orig1} \text{- } Mode\ decision\ (K_{spred1}, K_{pred1}).$$

e) Quantize K_{err1} using QP_1 and obtain prediction errors L_{err1}.

$$L_{err1} = QP_1\ (K_{err1}).$$

Perform entropy encoding on L_{err1} and obtain bitstream $S_{1.}$

f) Levels L_{err1} are de-quantized using QP_1^{-1} and obtain reconstructed residual coefficients K_{serr1}.

$$K_{serr1} = QP_1^{-1}(L_{err1}).$$

g) The reconstructed coefficients K_{rec1} are obtained by :

$$K_{rec1} = K_{pred1}+K_{serr1.}$$

h) The reconstructed coefficients K_{rec1} are quantized by Qs_1 to obtain reconstructed levels L_{rec1},

$$L_{rec1} = Qs_1\ (K_{rec1}).$$

i) The levels L_{rec1} are de-quantized using Qs_1^{-1} and the inverse DCT transform is performed to obtain the reconstructed image. The reconstructed image will go through a loop filter to smooth block artifacts and output to the frame buffer for the next frame encoding.

The proposed decoding process (as shown in Figure 4)

The decoding of the normal SP frame S_1 or S_2 in a normal bitstream (Use S_1 as an example.):

a) After entropy decoding of the Bitstream S1, the levels of the prediction error coefficients, L_{err1}, and motion vectors, are generated for the macroblock. Levels L_{err1} are dequantized using the quantizer QP_1^{-1}:

$$K_{serr1} = QP_1^{-1}(L_{err1}).$$

b) After motion compensation, perform forward DCT transform on the predicted macroblock and obtain K_{pred1}, the reconstructed coefficients K_{rec1} are obtained by:

$$K_{rec1} = K_{pred1}+K_{serr1}.$$

There is an option that a display video with better quality can be outputted. After performing an inverse DCT transform on K_{rec1}, the reconstructed image goes through a post filter and proceeds to display.

c) The reconstructed coefficients K_{rec1} are quantized by Qs_1 to obtain reconstructed levels L_{rec1},

$$L_{rec1} = Qs_1\ (K_{rec1}).$$

d) The levels L_{rec1} are de-quantized using Qs_1^{-1} and the inverse DCT transform is performed to obtain the reconstructed image. The reconstructed image will go through a loop filter to smooth block artifacts and output for next frame decoding. A display image can also be generated here.

It is clear that there are significant differences between the proposed SP method and the original H.26L SP method, due to which the proposed SP scheme provides several advantages over it. The main improvement is that two DCT coefficient coding modes are proposed for prediction to effectively limit the mismatch between the references used in prediction and reconstruction. The two DCT coefficient coding modes, namely *Q-mode* and *N-mode*, are determined with a rate-distortion optimization criterion for normal SP frame coding according to different references used for prediction.

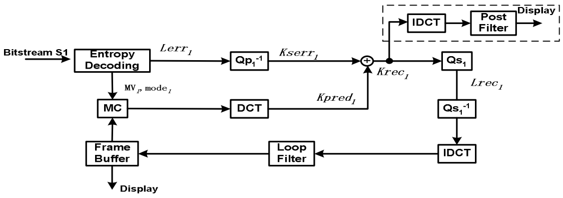

Figure 4: The proposed decoder for normal SP.

In *N-mode*, the reference used for the current coefficient prediction is the original prediction information K_{pred}; while in *Q-mode*, the quantized original prediction information K_{spred} is utilized as the prediction in the residue generation. It should be noted that no matter what reference is chosen as the prediction, the reference used in reconstruction is the original prediction information. Therefore, a notable advantage of this method is that no overhead bits are needed for informing the decoder in which predictive mode each SP DCT coefficient is encoded.

A correspond rate-distortion optimized decision-making mechanism is proposed for each SP DCT coefficient coding. The mode decision is done by minimizing the Lagrangian functional

$$C(K_{orig}, K_{pred}, m \mid Qs, Qp, \lambda_m) = SAD(K_{orig}, K_{pred}, m \mid Qs) + \lambda_m \cdot R(K_{orig}, K_{pred}, m \mid Qp)$$

with *Qp* is the normal quantization parameter, *Qs* is the SP quantization parameter, and *m* indicates a mode decision chosen from the set of potential prediction modes:

$$m \in \{N\text{-}mode,\ Q\text{-}mode\},$$

λ_m *is* the Lagrange multiplier for coefficient predictive mode decision. K_{orig} and K_{pred} represent the original and predictive values under *m* mode. The rate term $R(K_{orig}, K_{pred}, m \mid Qp)$ represents the number of actual bits associated with choosing *m* mode and *Qp*.

The proposed mode decision only concerns on the selection of prediction information. Since it is the original prediction K_{pred} used in the reconstruct of all INTER coded DCT coefficients, the mismatch between the references of prediction and reconstruction still exist by using the two proposed modes. However, because the mode decision criterion is based on the rate-distortion optimized function, the side effect of such mismatch on references is effectively reduced. Furthermore, the more valuable candidates can be used in prediction, the more quality gain may be achieved. Proper selection between these two prediction references at coefficient level can certainly improve the coding efficiency. Additionally, a high quality display image can be outputted before the quantization process in the proposed SP reconstruction loop. This will further improve the decoded video quality especially when many switching points are inserted in the normal bitstream.

3. THE PROPOSED CODING SCHEME FOR SWITCHING SP FRAME

Figure 5 shows how the switching SP is coded in the original H.26L SP technology. Assume that the switching is taken place from S1 to S2, based on the encoding of normal SP frame, the encoding of switching bitstream S12 is to first subtract L_{pred1} in S1 from the reconstruct level L_{rec2} in S2, and then perform entropy coding of the obtained difference. Since the quantization parameter *Qs* is always included in the normal SP encoding loop, the size of the switching bitstream S12 is smaller compared with an I frame.

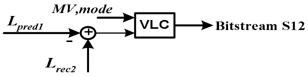

Figure 5: the original H.26L encoder for switching SP frame.

On the other hand, due to the separation of the switching quantization parameter *Qs*, the proposed switching SP coding scheme can provide adjustable size of switching bitstream. The encoding process in the proposed scheme for switching SP is shown in Figure 6. Different from the original SP method, the quantization on the prediction coefficients is drawn out from the normal SP encoding loop but performed in the switching SP encoding. The prediction coefficients K_{pred1} of S1 is quantized using Qs_2 to obtained coefficients $L_{pred12.}$ The difference between L_{pred12} and the reconstructed level L_{rec2} in S2 is entropy encoded to generate the bitstream S12.

If the current switching frame is frame t, then, in this processing, K_{pred1} is the predicted DCT coefficient derived from the reconstructed reference in the Bitstream S1 at frame t-1, and L_{rec2} is the reconstructed DCT coefficient obtained in the Bitstream S2 at frame t. Knowing that the switch is performed from S1 to S2, since L_{rec2} is already quantized with parameter *Qs* in bitstream S2, the same parameter is also used in K_{pred1} to reduce the size of Bitstream S12. It provides the important feature of unequal bitstream size in up-switching and down-switching, because the quantization parameter for up-switching and down-switching can be difference.

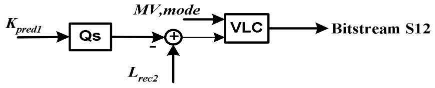

Figure 6: the proposed SP encoder for switching SP frame.

The proposed decoding of switching bitstream S12 (switching from Bitstream S1 to Bitstream S2).

a) After entropy decoding of the bitstream S12, the levels of the prediction error coefficients, L_{err12}, and motion vectors, are generated for the macroblock. Levels L_{err12} are de-quantized using the quantizer Qs_2^{-1}:

$$K_{serr12} = Qs_2^{-1}(L_{err12}).$$

b) After motion compensation, perform forward DCT transform on the predicted macroblock and obtain K_{pred1}, the reconstructed coefficients K_{rec12} are obtained by:

$$K_{rec12} = K_{pred1} + K_{serr12}.$$

e) The reconstructed coefficients K_{rec12} are quantized by Qs_2 to obtain reconstructed levels L_{rec2},

$$L_{rec2} = Qs_2\,(K_{rec1}).$$

f) The levels L_{rec2} are de-quantized using Qs_2^{-1} and the inverse DCT transform is performed to obtain the reconstructed image. The reconstructed image will go through a loop filter to smooth block artifacts and output to the frame buffer to display and for next frame decoding.

The resultant picture is exactly as same as that decoded from S2. Thus a drifting-free switching from Bitstream S1 to Bitstream S2 is achieved.

In the proposed SP method, the quantization with *Qs* on the prediction is released from the encoding loop. That is, bitstream S12 is only related to Qs_2, while bitstream S21 for switching from Bitstream S2 to Bitstream S1 only depends on Qs_1. By optimizing the parameters Qs_1 and Qs_2, the proposed SP scheme completely solves the contradiction between reducing switching bitstream size and improving coding efficiency of the normal bitstream. Moreover, in the proposed SP method, the switching points for up-switching and down-switching can be decoupled according to the real streaming requirements. In other words, it allows more very efficient down-switching points than up-switching ones to suit the TCP-friendly protocols and provide smooth streaming.

4. EXPERIMENTAL RESULTS

The coding efficiency of the proposed scheme and the H.26L SP coding scheme are evaluated. The TML 9.4 software is used in this experiment. The sequences Foreman and Container in QCIF format are encoded at 10Hz. In this experiment, only the first frame is encoded as I frame, and other frames are encoded as SP frames. The parameters are set as follows at encoder:

• 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 7. Each curve consists of five points corresponding to the *Qp* equals to 12, 16, 20, 24 and 28. Compared with the original H.26L SP method, with same decoding complexity, the proposed scheme can get better reference quality; when the high quality display image is outputted, the proposed scheme can improve the coding efficiency up to 1.0dB.

5. CONCLUSION

An efficient and flexible scheme is propose in this paper to achieve seamless video bitstream switching at predictive frames without drifting error. Compared with the original H.26L SP method, the presented method provides more advantages and higher coding performance. By introducing two rate-distortion optimal DCT coefficient modes for prediction, the mismatch between references for prediction and reconstruction is effectively limited. Together with the reconstruction of the high quality display image, the coding efficiency of the proposed SP method is significantly improved.

Moreover, by decoupling different quantization parameters for switching up and switching down, efficient and flexible drifting-free bitstream switching for video streaming can be achieved with the proposed SP scheme. It can provide 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.

The proposed SP scheme has been officially accepted by JVT standard.

6. 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] D. Wu, Y. T. Hou, W. Zhu, Y.-Q. Zhang, J. M. Peha, “Streaming video over the Internet: Approaches and Directions”, IEEE trans. on Circuits and Systems for Video Technology, vol 11, no3, 282-300, 2001.

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

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

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

[6] R. Kurceren, M. Karczewisz, “Synchronization-Predictive coding for video compression: the SP frames design for JVT/h.26L”, ICIP2002, 497-500, USA, Sep. 2002.

[7] X. Sun, F. Wu, S. Li, W. Gao, Y.-Q. Zhang, “Improved SP coding technique”, JVT-B097, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, Geneva, January 2002.

[8] X. Sun, F. Wu, S. Li, R. Kurceren, “The improved JVT-B097 SP coding scheme”, JVT-C114, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, Fairfax, May 2002.

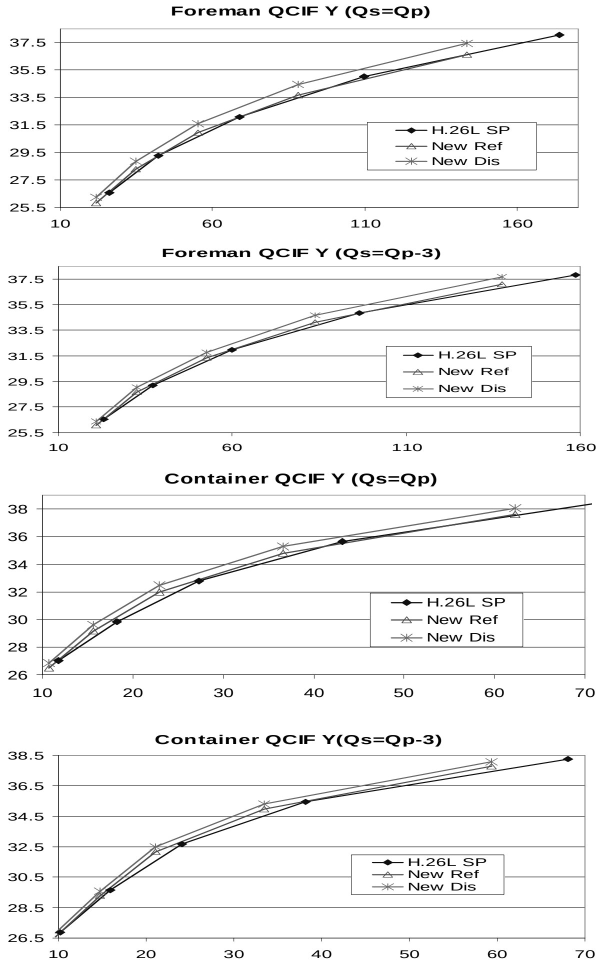

Figure 7: The experimental results between the proposed scheme and H.26L SP scheme.