Enhanced Video Stream Switching Schemes for H.264

Xiaosong Zhou and C.-C. Jay Kuo
Integrated Media Systems Center and Department of Electrical Engineering
University of Southern California, Los Angeles, California, U.S.A.

***Abstract*— Two enhanced video stream switching schemes for H.264 are proposed aiming at prompt switching among compressed video streams coded at different quality levels and bit rates. The original stream switching scheme using primary and secondary SP/SI pictures in H.264 is effective in switching, but degrades the compression efficiency of the transmitted stream even if the switching operation does not take place. In the two proposed schemes, no modification is made to the original compressed streams (namely, no primary SP picture is adopted) so that compression efficiency is maintained at switching points. As a result, switching points can be assigned flexibly in our enhanced schemes according to the need of applications and network conditions without worrying about the bit rate overhead. When switching occurs, the proposed schemes use a new picture type - DIFF (difference picture) in different ways to compensate the mismatch of reference frames. The first proposed scheme is designed to simplify the implementation while the second scheme is designed to further reduce the bit rate overhead. It is shown by experimental results that both of the proposed schemes outperform the switching scheme defined in H.264 and they are able to realize prompt stream switching without noticeable quality degradation and have a small bit rate overhead only when the actual switching operation occurs.**

Topic area - Multimedia Processing.

I. Introduction

Video streaming is one of the major methods of distributing video data for entertainment and information services. Many communication networks (wired, wireless and cellular networks) start to provide sufficient bandwidths for transmission of high-quality video data. At the same time, emerging video coding standards such as H.264 [1] help reduce the bit rate of compressed video substantially. As more and more electronic devices and portable devices have the built-in video display capability, there will be a tremendous growth in these applications in the near future.

In H.264, new picture types - primary and secondary SP/SI pictures are introduced by Karczewicz and Kurceren [2] to facilitate stream switching. Video frames are encoded into SP pictures at switching points instead of intra-coded pictures. The coding efficiency of the SP pictures is much higher than that of intra-coded frames, but they are still less efficient than normal P pictures. Therefore, the overall coding efficiency is still degraded if many switching points are assigned.

In this work, two enhanced stream switching methods are proposed that do not require any modification to the original video stream at the switching points. Therefore, no matter how many switching points are assigned, the overall coding efficiency will not be affected. It is shown by experimental results that the proposed schemes are able to realize flexible stream switching and have a small bit rate overhead only when switching occurs.

The rest of the paper is organized as follows. Some related previous work is reviewed in Sec. II. The stream switching method in H.264 is discussed in Sec. III. The two proposed schemes are presented in Sec. IV. It is followed by experimental results in Sec. V and concluding remarks in Sec. VI.

II. Review of Previous Work

For a video streaming system, guaranteed end-to-end QoS (quality of service) over the entire streaming period may be difficult to achieve. For example, the bandwidth available for video streaming may change during the streaming period. To utilize the bandwidth efficiently and avoid congestion caused by the overflow, the outgoing bit rate of the streaming application needs to be adjusted to fit the available bandwidth in real time.

One way to enable real-time bit rate adjustments during the streaming period is the use of scalable video coding [3]. That is, a video source is encoded into different quality levels or layers. During the transmission, the sender can choose not to transmit data in some of the layers to adjust the outgoing bit rate. The method is suitable for streaming over heterogeneous channels, but it degrades the overall compression efficiency and increases the computational complexity of both the encoder and the decoder. There are some on-going efforts for scalable H.264 video development [4].

A simple method to realize bit rate adjustment is to encode multiple versions of the same video sequence. These versions have different quality levels and bit rates. During the streaming, when there is a need to adjust the outgoing bit rate, the stream to be transmitted can be switched from one version to the other in order to fit the bandwidth requirement. This is called "stream switching". However, in existing video coding standards, most video frames in the stream are compressed using motion-compensated coding. Directly switching between streams at these inter-coded frames (P or B frames) tends to cause the mismatch of reference frames and results in incorrect reconstruction. The quality of reconstructed video may be degraded significantly. One method to solve the problem is to encode video frames into intra-coded pictures at pre-selected switching points. It is desired that stream switching can be processed immediately after the bandwidth change is detected. As switching can take place at these switching points only, the switching points need to be assigned frequently in the bit stream in order to realize prompt stream switching. However, it will introduce a substantial bit rate overhead to have many intra-coded frames in the video stream due to their low coding efficiency. Thus, a new stream switching scheme is proposed in H.264 as described in the next section.

III. Stream Switching in H.264

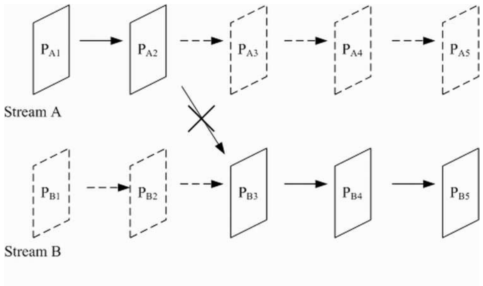

Fig. 1. Example of stream switching.

A stream switching scenario is shown in Fig. 1. Stream A and stream B are two compressed streams encoded from the same video sequence but with different bit rates. The third frame is selected as a switching point. Suppose that stream A's bit rate is higher and it is selected to be transmitted initially. However, a change of the bandwidth condition is detected prior to transmission of the third frame and it is decided that stream B is more suitable for the new bandwidth. In this case, stream switching will take place at the switching point (*i.e.* the third frame). However, P_{B3} can not be transmitted directly in order to switch the streams because it is coded using P_{B2} as the reference frame that was never transmitted. Only the reconstruction of P_{A2} exists at the decoder side and it is different from that of P_{B2}. If P_{B3} is transmitted, it cannot be reconstructed correctly. The errors introduced in the reconstruction not only degrades its own quality, but also affects the reconstruction of subsequent frames for a long period of time due to motion-compensated coding.

To solve the problem, a stream switching scheme using primary and secondary SP pictures was introduced in H.264 [2]. As shown in Fig. 2, the switching point is encoded into primary SP pictures SP_{A3} and SP_{B3} instead of the normal P pictures in both streams. If switching from A to B does not take place, the transmitted sequence for these five frames will be $P_{A1} \rightarrow P_{A2} \rightarrow SP_{A3} \rightarrow P_{A4} \rightarrow P_{A5}$.

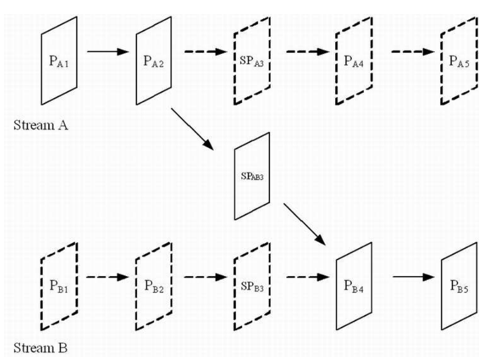

Fig. 2. Stream switching in H.264.

When switching is required, a secondary SP picture, denoted by SP_{AB3} that is coded using P_{A2} as the reference frame, will be transmitted instead of SP_{B3}. Thus, the five frames to be transmitted will be $P_{A1} \rightarrow P_{A2} \rightarrow SP_{AB3} \rightarrow P_{B4} \rightarrow P_{B5}$. Obviously, SP_{AB3} can be correctly reconstructed as its reference frame is available at the decoder side. At the same time, the encoding of primary SP picture SP_{B3} and secondary SP pciture SP_{AB3} ensures that they have the identical reconstruction. Thus, the subsequent frame transmitted after switching - SP_{B4} can also be correctly reconstructed since its reference frame can be obtained by decoding SP_{AB3}. With this method, the two streams are switched successfully without introducing any reconstruction errors. Similarly, if the above example changes to switching from stream B to A, another secondary SP picture SP_{BA3} has to be created and transmitted.

The above scheme is effective for stream switching. However, in order to make its reconstruction identical to the secondary SP picture SP_{AB3}, the coding efficiency of the primary SP picture SP_{B3} is poorer in comparison with a normal P picture P_{B3}. As the primary SP pictures at switching points are transmitted if switching does not take place, the overall coding efficiency of the transmitted stream is degraded.

For most video streaming applications, video sequences are compressed off-line. By using the above switching scheme, switching points need to be assigned during the off-line encoding process. The bandwidth change for the streaming channel happens randomly and it is difficult to predict the change beforehand. Prompt stream switching is required to adjust the outgoing bit rate quickly to fit the time-varying bandwidth. Thus, switching points need to be assigned frequently in the video stream so that stream switching can be processed immediately after a bandwidth change is detected. On the other hand, the actual number of bandwidth changes that require switching during the streaming period may be very limited.

To conclude, there exists a dilemma in the above switching scheme adopted by H.264. A large number of video frames are coded into primary SP pictures but switching may not take place in most of these frames. The existence of the primary SP pictures degrades the overall coding efficiency of the transmitted video stream with only a small portion of them really functioning for switching.

IV. Two Enhanced Stream Switching Schemes

In order to overcome the problem described above, two enhanced stream switching schemes are introduced in this section. The proposed schemes have two common characteristics. First, they do not require modification of the originally compressed stream so that the coding efficiency of the transmitted stream will not be degraded if no switching occurs. Second, when a switching operation is required, prompt switching can still be achieved without introducing noticeable errors and only a small bit rate overhead will be needed.

Fig. 3 illustrates the first proposed switching scheme. In the given example, the third frame is selected as a switching point. When switching is not required, five original P pictures in stream A are transmitted as $P_{A1} \rightarrow P_{A2} \rightarrow P_{A3} \rightarrow P_{A4} \rightarrow P_{A5}$. The original compression efficiency of stream A is preserved. The reason that P_{B3} cannot be transmitted

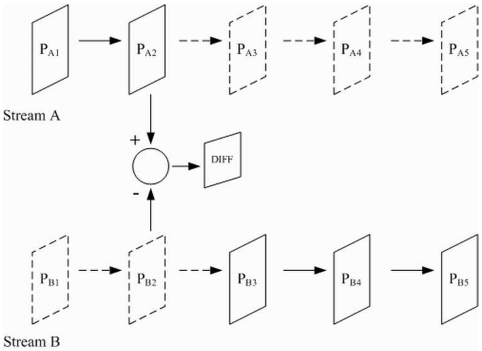

Fig. 3. The first enhanced stream switching scheme.

directly when switching is required is that the reconstruction of P_{A2} available at the decoder side is different from the reference frame needed to reconstruct P_{B3}, which is the reconstructed P_{B2}. Therefore, we can simply calculate the difference of the reconstructions of P_{A2} and P_{B2}, encode it into a difference picture (DIFF) and transmit the picture to the decoder to compensate the mismatch. The transmitting sequence for switching will then be $P_{A1} \rightarrow P_{A2} \rightarrow DIFF \rightarrow P_{B3} \rightarrow P_{B4} \rightarrow P_{B5}$. Upon receiving the DIFF picture, the decoder simply adds its values to the reconstruction of P_{A2} in the reference buffer. By using this method, the two streams are switched successfully.

It is clear that no modification is made to the originally compressed frames at the switching points in the above scheme. Therefore, the switching points can be assigned freely without worrying about degradation of coding efficiency. The DIFF picture encodes the difference between the two differently coded versions of the same video frame and such a difference is usually very small. Thus, it does not require a large number of bits to encode it. As the DIFF picture is only transmitted when switching occurs, the bit rate overhead introduced by the above scheme is small.

However, as high-precision compression is required to encode the DIFF picture to avoid errors during switching, it is desirable that the number of bits required to encode DIFF picture is further reduced. To achieve this goal, we propose another switching scheme that utilizes the secondary SP picture described in Sec. III. Similarly as the first scheme, original P pictures are transmitted when switching is not required in the proposed scheme. When switching takes place at the third frame, a pre-coded secondary SP picture SP_{AB3} is transmitted.

The encoding and decoding diagrams for the blocks in the secondary picture SP_{AB3} are shown in Fig. 4. Its encoding process has two major features. First, the input to the encoder is the reconstructed P_{B3} instead of the raw video data. This is to ensure that the reconstruction of the SP_{AB3} is very close to that of P_{B3}. Second, the transform and quantization are processed before the residual is calculated. SP_{AB3} uses P_{A2} as the reference frame and it is different from the reference frame of P_{B3}. Thus, it is important that the different prediction used does not affect the desired reconstruction. As shown in the figure, by using this encoding method the 'P1' value in encoding process and the 'P2' value in the decoding process are identical and, as a result, the different prediction used does not affect the reconstructed value.

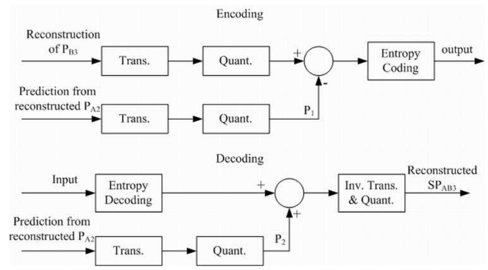

Fig. 4. The encoding and decoding diagrams of SP_{AB3}.

By combining the encoding and the decoding processes of the secondary SP picture SP_{AB3}, the overall diagram can be simplified to the one shown in Fig. 5. As shown in the figure, the only possible difference of the input (*i.e.* the reconstruction of P_{B3}) and the output (*i.e.* the reconstruction of SP_{AB3}) are introduced in the forward/inverse transform and the quantization process. If the forward/inverse transform process is lossless, errors introduced in the above process are only quantization errors. Since such errors are very small, the reconstruction of SP_{AB3} can be used directly to decode P_{B4} after switching without introducing large errors.

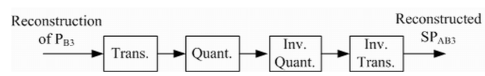

Fig. 5. The combined encoding/decoding process of SP_{AB3}.

When needed, the small mismatch between reconstructions of SP_{AB3} and P_{B3} can be further reduced by transmitting another difference picture DIFF that encodes their difference. Upon receiving the DIFF picture, the decoder adds it to the reconstruction of SP_{AB3} in the reference buffer. As the energy of the difference values to be coded are extremely small, the number of bits needed to encode this DIFF picture is usually much smaller than that of the DIFF picture in the first proposed scheme. As the bit consumption of SP_{AB3} is similar to that of P_{B3}, the overall bit rate overhead of the second scheme is further reduced.

In summary, both of the proposed stream switching schemes do not require any modification to the originally compressed stream at switching points. The original coding efficiency is maintained regardless of the assignment of switching points. Therefore, switching points can be assigned as frequently as possible depending on application requirements and network conditions. Prompt and correct stream switching can be achieved using both schemes without introducing large bit rate overhead. The first scheme does not require SP pictures and is very simple in implementation. The second approach further reduces the bit rate overhead when switching takes place at the expense of higher implementation complexity.

V. Experimental Results

In order to demonstrate the performance of the two enhanced stream switching schemes, stream switching is simulated in our experiments. The switching points are assigned with different frequencies to show their influence on the streaming performance. They are also tested when different numbers of actual switching takes place during the streaming period. The proposed schemes are compared with the stream switching method based on primary and secondary SP frames, which is adopted by H.264 as described previously.

In the experiments, the H.264 reference software (JM8.6) is modified to test the proposed schemes and the QCIF Foreman sequence is used. The bit rate of our proposed schemes and the switching scheme in H.264 are compared in Fig. 6. The quantization step size of the primary SP frames in the later approach is adjusted so that the overall PSNR of bit stream is close to the original stream (set as QP-6). As shown in the figure, the bit rates of the two proposed schemes do not change at all due to the fact that the number of switching points assigned does not affect their bit rates. Their bit rates actually equal to the original bit rate of the compressed stream. The bit rate overhead of the stream switching in H.264 becomes larger when more switching points are assigned since more primary SP pictures are transmitted and they demand a larger number of bits as the overhead.

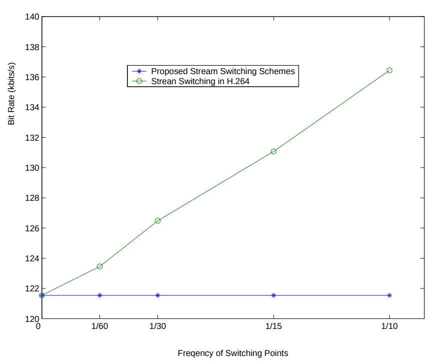

Fig. 6. The bit rate comparison as the number of switching points increases.

The R-D performances of the proposed schemes are also examined when switching does take place. Fig. 7 and Fig. 8 show the R-D curves of the proposed schemes when switching takes place every 300 frames and every 150 frames (10 seconds and 5 seconds for 30fps). They are compared with the stream switching scheme in H.264 and the ideal stream switching without any efficiency degradation. One switching point is assigned for every 15 frames in the video stream to ensure that prompt stream switching can be achieved in real applications. The R-D curve of the original compressed stream is also provided for comparison. As shown in these figures, both of the proposed schemes have better R-D performance when switching takes place. The proposed second scheme has a smaller bit rate overhead, and it R-D performance is very close to that of ideal switching.

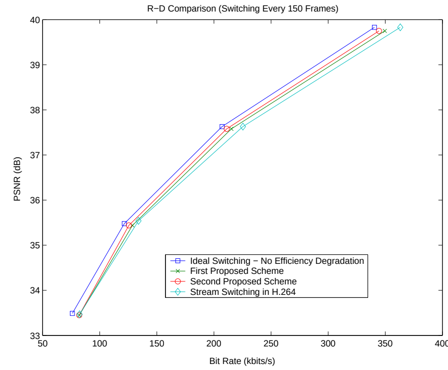

Fig. 7. The R-D performance comparison for switching at every 150 frames.

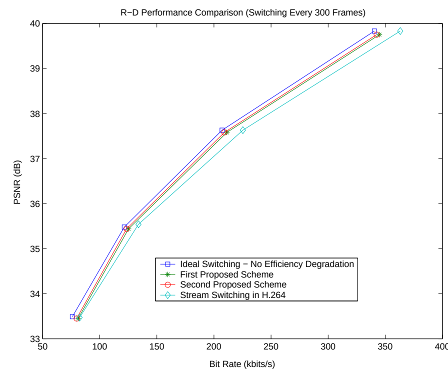

Fig. 8. The R-D performance comparison for switching at every 300 frames.

VI. Conclusion

Two enhanced stream switching schemes were proposed for streaming applications by adjusting the outgoing bit rate promptly to adapt to the available bandwidth efficiently. The enhanced schemes do not degrade the coding efficiency of received video when no switching takes places. Thus, switching points can be assigned freely based on application requirements and network conditions. It was shown by the experimental results that only a small number of overhead bit rate is introduced when switching occurs for both schemes. The first proposed scheme is very simple in implementation while the second approach further reduces the bit rate overhead at the expense of slightly higher complexity. The proposed schemes can serve as useful extensions for H.264 and most existing video coding standards.

References

[1] ITU-T Rec. H.264 - ISO/IEC 14496-10 AVC, Document JVT-G050, 8th Meeting: Geneva, Switzerland, May 2003.

[2] M. Karczewicz and R. Kurceren,”The SP- and SI-frames design for H.264/AVC”, *IEEE Trans. On Circuits and Systems for Video Technology*, vol. 13, no. 7, pp.637-644, July 2003.

[3] W. Li, ”Overview of fine granularity scalability in MPEG-4 video standard”, *IEEE Trans. On Circuits and Systems for Video Technology*, vol. 11, no. 3, pp.301-317, March 2001.

[4] K. Ugur and P. Nasiopoulos, ”Combining bitstream switching and FGS for H.264 scalable video transmission over varying bandwidth networks”, *Communications, IEEE Pacific Rim Conference on Computers and signal Processing*, Volume 2, pp.28-30 August 2003.