

SEAMLESS SWITCHING OF SCALABLE VIDEO BITSTREAMS FOR EFFICIENT STREAMING

Xiaoyan Sun^{*1}, Feng Wu², Shipeng Li², Wen Gao¹, Ya-Qin Zhang²

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

²Microsoft Research Asia, Beijing, 100080

ABSTRACT

In this paper, we propose a seamless switching scheme for scalable video bitstreams that fully takes advantage of both the high coding efficiency of non-scalable bitstreams and the flexibility of scalable bitstreams. Small bandwidth fluctuations are accommodated by the scalability of the bitstreams, while large bandwidth fluctuations are tolerated by switching between scalable bitstreams. The major contribution of this paper is a flexible and effective scheme for seamless switching between scalable bitstream operated at lower rates to one operated at higher rates without sending any overhead bits. Additional bits are only necessary when switching from a scalable bitstream operated at lower rates to one operated at higher rates. Experimental results show that the proposed scheme significantly outperforms both the approach with a single scalable bitstream and the approach of switching among multiple non-scalable bitstreams.

1. INTRODUCTION

With steady growth of 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 dynamical best-effort network, channel bandwidth usually fluctuates in a wide range from bit rate below 64kb/s to well above 1Mbits/s. This brings great challenges to video coding and streaming technologies in providing a smooth playback experience and best available video quality to the users. To deal with the network bandwidth variations, two main approaches, namely, switching among multiple non-scalable bitstreams and streaming with a single scalable bitstream, have been extensively investigated in recent years.

In the first approach, a video sequence is compressed into several non-scalable bitstreams at different bit rates. Some special frames, known as *key frames*, are either compressed without prediction or coded with an extra switching bitstream [3][4]. Key frames provide access points to switch among these bitstreams to fit in the available bandwidth. The advantage of this method is the high coding efficiency with non-scalable bitstreams. However, due to limitation in both the number of bitstreams and switching points, this method only provides coarse and sluggish capability in adapting to channel bandwidth variations.

In the second approach, a video sequence is compressed into a single scalable bitstream, which can be truncated flexibly to adapt to bandwidth variations. Among numerous scalable coding techniques, MPEG-4 Fine Granularity Scalable (FGS) coding has become prominent due to its fine-grain scalability [5]. Since the enhancement bitstream can be truncated arbitrarily in any frame, FGS provides a remarkable capability in readily and precisely adapting to channel bandwidth variations. However, low coding efficiency is the vital disadvantage that prevents FGS from being widely deployed in video streaming applications. Progressive Fine Granularity Scalable (PFGS) coding scheme [6][7] is a significant improvement over FGS by introducing two prediction loops with different quality references. On the other hand, since only one high quality reference is used in enhancement layer coding, most coding efficiency gain appears within a certain bit rate range around the high quality reference. Generally, with today's technologies, there is still a coding efficiency loss compared with the non-scalable case at fixed bit rates.

A seamless switching scheme is proposed in this paper to significantly improve the efficiency of scalable video coding over a broad bit rate range by using two scalable bitstreams. Each scalable bitstream has a base layer with different bit rate and can best adapt to channel bandwidth variations within a certain bit rate range. If the channel bandwidth is out of this range, the scalable bitstream can be seamlessly switched from one to another with better coding efficiency. We will refer to switching from a scalable bitstream operated at lower bit rates to one operated at higher bit rates as *switching up* and the reversion as *switching down* hereafter in this paper. The key problem we try to solve in this paper is how to flexibly and effectively switch up and down between scalable bitstreams. In particular, when channel bandwidth somehow drops, the server has to rapidly switch from high bit rate bitstream to low bit rate bitstream to reduce packet loss ratio and maintain smooth video playback. Therefore, there are three basic requirements in the proposed scheme: (1) the scalable bitstreams could be switched down at any frame; (2) there should be no drifting errors introduced by switching; (3) overhead bits should be avoided during switching down since they will increase network traffic and may further deteriorate network conditions.

This paper is organized as follows. Section 2 describes how to encode a video sequence into two scalable bitstreams for the proposed seamless switching scheme. Seamless switching between scalable bitstreams is 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.

2. SCALABLE VIDEO CODING FOR SEAMLESS SWITCHING

Either MPEG-4 FGS or PFGS coding can be used in the proposed scheme. For better coding efficiency, the macroblock-based PFGS (MPFGS) is chosen as the basic scalable video codec in this paper [7]. The MPFGS codec compresses a video sequence into two bitstreams. In each frame, the base layer bitstream is first generated by traditional non-scalable coding technique, and then the residue between original/predicted DCT coefficients and dequantized DCT coefficients of the base layer forms the enhancement layer bitstream with bit-plane coding technique. The bit rate of the base layer is the lower bound of the channel bandwidth covered by this scalable bitstream. The enhancement layer bitstream provides fine-grain scalability to adapt to channel bandwidth variations.

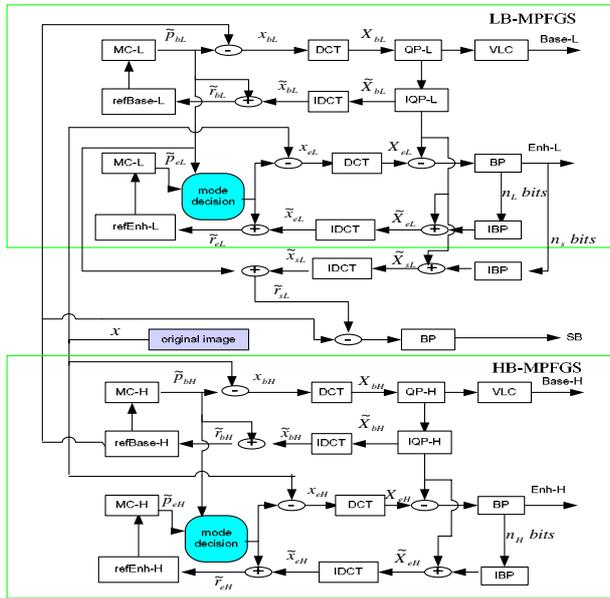


Figure 1: The proposed coding framework with two MPFGS encoders.

The framework for encoding two scalable bitstreams in the proposed scheme is illustrated in Figure 1. Motion estimation modules are omitted for simplicity. There are two MPFGS encoders outlined by the dashed boxes in Figure 1. The upper one is denoted as LB-MPFGS since it generates a scalable bitstream with a lower bit rate base layer, whereas the lower one is denoted as HB-MPFGS for higher bit rate base layer case accordingly. The middle part between the two MPFGS encoders is used to generate an extra bitstream for switching up. For convenience in discussion, lowercase letter denotes an image in pixel domain, and the corresponding uppercase letter denotes an image in DCT domain. The subscript “b” and “e” indicate the base layer and the enhancement layer, respectively. The hat “ \sim ” denotes reconstructed image or DCT coefficients. The symbols “-H” and “-L” are used to distinguish the modules in HB-MPFGS and LB-MPFGS, respectively.

To ensure that the MPFGS bitstreams are able to be seamlessly switched from one to the other, the base layer bitstreams of these two MPFGS encoders are actually not generated independently. Firstly, motion vectors are estimated in HB-MPFGS and are

applied to both HB-MPFGS and LB-MPFGS. Secondly, the video frames to be encoded for the base layer of LB-MPFGS are the reconstructed base layer frames from HB-MPFGS instead of the original video frames. Now we will discuss how to generate the two MPFGS bitstreams with the proposed framework.

The original video is first input to HB-MPFGS. Since the motion vectors estimated will be used for both MPFGS encoders, original video frames are used as reference for estimating integer motion vectors, whereas fractional parts of motion vectors are still estimated by referencing the reconstructed base layer of HB-MPFGS to maintain the coding efficiency. There are two references in each MPFGS codec. The low quality reference \tilde{r}_{bH} stored in *refBase-H* frame buffer is reconstructed from the base layer, whereas the high quality reference \tilde{r}_{eH} stored in *refEnh-H* frame buffer is reconstructed from both the base layer and the enhancement layer. The base layer only uses low quality reference for prediction and reconstruction, while the enhancement layer can select either the low quality reference or the high quality reference, which is decided by mode decision algorithm proposed in [7]. The base layer bitstream and the enhancement layer bitstream are generated using MPEG-4 non-scalable coding and bit plane coding, respectively.

LB-MPFGS obtains the motion vectors directly from HB-MPFGS without motion estimation. Normally the bit rate of the base layer in HB-MPFGS is much higher than that in LB-MPFGS. To make seamless switching from HB-MPFGS bitstream to LB-MPFGS bitstream possible, the reconstructed high quality base layer from HB-MPFGS instead of the original video is input to the LB-MPFGS base layer encoder. In other words, the predicted error x_{bL} encoded at LB-MPFGS base layer is the difference between the reconstructed HB-MPFGS base layer \tilde{r}_{bH} and the prediction \tilde{p}_{bL} . This is similar to transcoding bitstreams from high bit rate to low bit rate. On the other hand, the predicted error x_{eL} is still calculated from the original video in order to maintain the coding efficiency of the enhancement layer in LB-MPFGS.

Directly switching up from LB-MPFGS bitstream to HB-MPFGS bitstream will cause severe drifting errors, because the references used in LB-MPFGS are \tilde{r}_{bL} and \tilde{r}_{eL} , whereas the references used in the HB-MPFGS base layer is \tilde{r}_{bH} . Normally the switching up happens when the available channel bandwidth is high enough to cover the HB-MPFGS base layer bit rate. There is a corresponding switching point at LB-MPFGS enhancement layer. When switching up, it is reasonable to assume that the DCT coefficients encoded in the LB-MPFGS enhancement layer up to the switching point are correctly transmitted to the decoder. Therefore, the reconstructed image \tilde{r}_{sL} , which is obtained by adding the low quality prediction \tilde{p}_{bL} and the current DCT coefficients encoded in LB-MPFGS up to the switching point, is available at both the encoder and the decoder. Differences between \tilde{r}_{sL} and \tilde{r}_{bH} are losslessly encoded with bit plane coding to form an extra bitstream, known as *switching bitstream*, to ensure a drifting-free switching up.

3. SEAMLESS SWITCHING UP AND DOWN BETWEEN SCALABLE BITSTREAMS

How to switch up and down between MPFGS bitstreams is discussed in this section. The procedures of both switching up and down are depicted in Figure 2.

experiment does not adopt the drifting control technique in LB-MPFGS, the reconstructed quality may have a little loss at low enhancement bit rates like as the result of News sequence at 48kbps. However, the coding efficiency of the proposed scheme can be 2.0dB higher than MPFGS and 3.0dB higher than FGS at higher bit rates.

A dynamic changing channel is used to verify the performance of these four different schemes in terms of bandwidth adaptation. The bit rate periodically switches from 72kbps to 152kbps. Each cycle starts at 72kbps for 1 second and then switches to 152 kbps for 3 seconds. The curves of PSNR versus frame number are shown in Figure 5. The proposed scheme switches up 3 times then switches down twice in order to adapt to channel bandwidth fluctuations. Clearly, the proposed seamless switching scheme can always achieve the best performance among these four schemes at both lower bit rates and higher bit rates.

5. CONCLUSIONS AND DISCUSSIONS

In this paper, we propose a seamless switching scheme for scalable video bitstreams that fully takes advantage of both the high coding efficiency of non-scalable bitstreams and the flexibility of scalable bitstreams. Small bandwidth fluctuations are accommodated by the scalability of the bitstreams, while large bandwidth fluctuations are tolerated by switching between scalable bitstreams. The proposed scheme can cover an extended bit rate range with significantly improved coding efficiency. The experimental results show that the proposed scheme outperforms the method of switching non-scalable bitstreams both in coding efficiency and bandwidth adaptation. The coding efficiency of the proposed scheme is also significantly higher than that of single MPFGS or MPEG-4 FGS bitstream over a wide range of bit rates.

One concern in the proposed scheme is the complexity. Since there are two MPFGS encoders in the proposed framework, more motion compensation and DCT transform modules are used. On the other hand, the complexity increase is manageable since still only one motion estimation module is needed and streaming applications often allow offline encoding. In the decoder, the proposed framework decodes either the LB-MPFGS bitstream or the HB-MPFGS bitstream but not simultaneously. The complexity of the LB-MPFGS decoder is the same as a single MPFGS decoder. The complexity of the HB-MPFGS decoder is higher than that of a single MPFGS decoder because of the additional transcoder-like structure. Our single MPFGS decoder can real-time decode a single MPFGS bitstream in CIF format at 1Mbps with PII 400 Mhz laptop.

How to support more than two scalable bitstreams to cover greater bandwidth range, and how to optimally losslessly compress the difference frame when switching up are two open questions need to be further investigated.

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, special issue on streaming video, Vol. 11, no 3, 301-317, 2001
- [3] B. Girod, N. Farber, and U. Horn, "Scalable codec architectures for Internet video on demand," in Proc. 1997

- Asilomar Conf. Signals and Systems, Pacific Grove, CA, Nov, 1997.
- [4] M. Jarczewicz and R. Kurceren, "A Proposal for SP-frames," document VCEG-L-27, ITU-T Video Coding Experts Group Meeting, Eibsee, Germany, Jan, 2001.
- [5] W. Li, "Fine granularity scalability in MPEG-4 for streaming video," ISCAS 2000, vol 1, 299-302, Geneva, Switzerland, May 28-31, 2000.
- [6] F. Wu, S. Li and Y.-Q. Zhang, "A framework for efficient progressive fine granularity scalable video coding," IEEE Trans. Circuits and Systems for Video Technology, special issue on streaming video, Vol. 11, no 3, 332-344, 2001.
- [7] X. Sun, F. Wu, S. Li, W. Gao, and Y.-Q. Zhang, "Macroblock-based progressive fine granularity scalable video coding," ICME2001, Tokyo, Aug. 22-25, 2001.

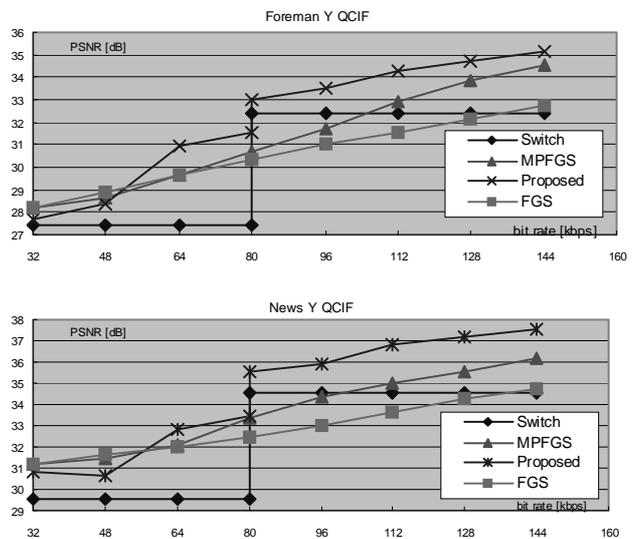


Figure 4: The curves of average PSNR versus bit rates.

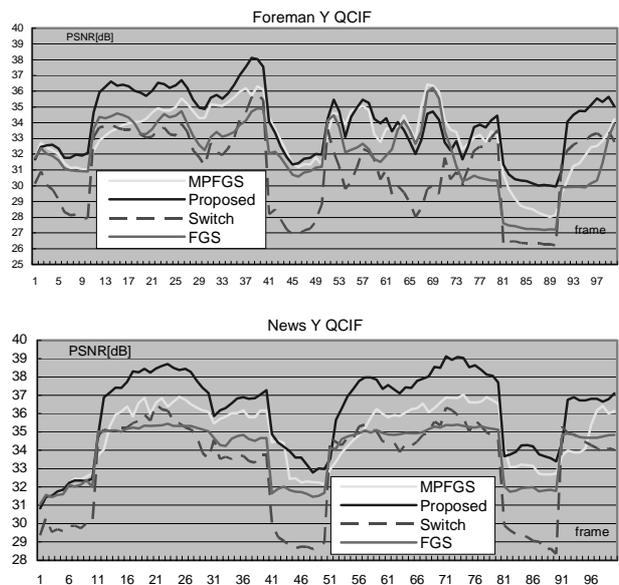


Figure 5: The curves of PSNR versus frame number in channel with bandwidth periodically varying at 72kbps and 152kbps.