MACROBLOCK-BASED PROGRESSIVE FINE GRANULARITY SCALABLE (PFGS) VIDEO CODING WITH FLEXIBLE TEMPORAL-SNR SCALABLILITIES

Xiaoyan Sun[*1], *Feng Wu*[2], *Shipeng Li*[2], *Wen Gao*[1], *Ya-Qin Zhang*[2]

[1] Department of Computer Application, Harbin Institute of Technology, Harbin, 150001
[2]Microsoft Research China, Beijing, 100080

ABSTRACT

In this paper, we proposed a flexible and efficient architecture for scalable video coding, namely, the macroblock (MB)-based progressive fine granularity scalable video coding with temporal-SNR scalabilities (PFGST in short). The proposed architecture can provide not only much improved coding efficiency but also simultaneous SNR scalability and temporal scalability. Building upon the original frame-based progressive fine granularity scalable (PFGS) coding approach, the MB-based PFGS scheme is first proposed. Three INTER modes and the corresponding mode selection mechanism are presented for coding the SNR enhancement MBs in order to make a good trade-off between low drifting errors and high compression efficiency. Furthermore, temporal scalability is introduced into the MB-based PFGS, which forms the MB-based PFGST scheme. Two coding modes are proposed for coding the temporal enhancement MBs. Since it would not cause any error propagation if using the high quality reference in the temporal enhancement MB coding, the coding efficiency of the PFGST is highly improved by always choosing the most suitable reference for the temporal scalable coding. Experimental results show that the MB-based PFGST video coding scheme can significantly improve the coding efficiency up to 2.8dB compared with the FGST scheme adopted in MPEG-4, while supporting full SNR, full temporal, and hybrid SNR-temporal scalabilities according to the different requirements from the channels, the clients or the servers.

1. INTRODUCTION

The fast growing networked video applications introduce a new video delivery system and also some new challenges, such as bandwidth fluctuation and different Quality-of-Service (QoS) which the conventional video coding techniques are not required to deal with [2]. To respond to the challenges of the new applications, scalable coding schemes are proposed to optimize the quality of video for a wide range of bit rates rather than a fixed bit rate [1]. The MPEG-4 Fine Granularity Scalable (FGS) video coding scheme [3][7] is one of such promising techniques which can easily adapt to the channel bandwidth fluctuation by using bit plane coding in the enhancement layer. However, the coding efficiency of FGS is low since only the lowest quality base layer is used in its motion prediction. The Progressive Fine Granular Scalable (PFGS) coding scheme proposed in [4] is an improvement over the FGS scheme. Unlike the FGS scheme, the PFGS scheme provides much higher coding efficiency by using high quality reference frames in enhancement layer coding. However, since the references selected are used for the whole frame in the original PFGS scheme, it is very difficult to provide a good trade-off between high coding efficiency and low drifting errors.

A limitation of those scalable coding schemes mentioned above is that only the image quality of each individual frame can be scalable, but not the frame rate. Considering the different access capabilities of users and the lack of QoS guarantee over the Internet, a better trade-off between the temporal resolution (frame rate) and image quality of each individual frame should be made. The Fine Granularity Scalable-Temporal (FGST) scheme proposed in [6][8], combining with the FGS coding scheme, can provide both SNR and temporal scalabilities for streaming video applications. However, in the FGST scheme, since the references for temporal frame prediction are always derived from the lowest quality base layer at the frame level, the coding efficiency of FGST scheme is consistently lower than that of traditional temporal scalable coding.

In this paper, a flexible and efficient MB-based Progressive Fine Granularity Scalable video coding scheme with Temporal-SNR scalabilities (PFGST) is proposed. Firstly, three INTER modes with the corresponding mode-selection criteria are presented for the SNR enhancement MB coding. The proposed scheme achieves a better balance between drifting errors reduction and coding efficiency improvement while preserving the excellent property of fine granularity SNR scalability. Then, the temporal scalability is introduced into the MB-based PFGS. Two prediction modes are presented for the temporal enhancement MB coding. By choosing the proper reference for each temporal enhancement MB from either high quality reference or low quality reference, the coding efficiency of the proposed scheme is greatly improved. Moreover, just like in the FGST scheme, since the temporal scalability only applies to the B-frames, no addition drifting errors will be introduced into the PFGST video if using the high quality reference in the temporal scalable coding. With the proposed solution, the MB-based PFGST can achieve fine granularity scalabilities for both SNR and temporal while providing a very high coding efficiency.

This paper is organized as follows. In Section 2, the MB-based PFGS scheme is presented. Three INTER modes and the mode-selection mechanism are proposed for SNR enhancement MB coding. In Section 3, the MB-based PFGS scheme with temporal scalability (PFGST) is proposed. Two prediction modes that can significantly improve the coding efficiency are introduced. Some 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 China.

0-7803-6725-1/01/$10.00 ©2001 IEEE

2. MACROBLOCK-BASED SNR PFGS

Instead of using only the base layer as the prediction reference as in the FGS scheme, the Progressive Fine Granularity Scalable (PFGS) coding scheme proposed in [4][5] improves the FGS scheme by using an additional high quality enhancement layer reference. A typical architecture of the PFGS is depicted in Figure 1. In the PFGS scheme, since the quality of the enhancement layer is higher than that of the base layer, the PFGS coding scheme provides more accurate motion prediction for the enhancement video layer to improve the coding efficiency by using as many high quality reference as possible. Moreover, by reconstructing the high quality references from the previous high quality reference and low quality reference alternatively, the PFGS scheme can effectively eliminate the error propagation that might be introduced when the high quality reference is not available at the decoder.

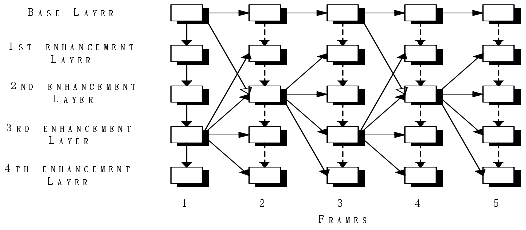

Figure 1: The architecture of the PFGS scheme.

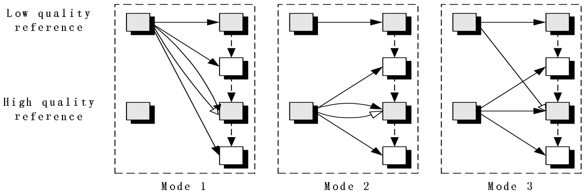

Figure 2: The coding modes for SNR enhancement macroblocks

In fact, since two references are used in the PFGS scheme, a more flexible trade-off between drifting errors and coding efficiency can be achieved if the references for prediction and reconstruction are chosen based on MBs in the SNR enhancement layer coding. Three INTER modes for coding the SNR enhancement layer MBs are depicted in Figure 2. Gray rectangular boxes in Figure 2 denote those layers to be reconstructed as references. Solid arrows with solid lines are for the temporal prediction, hollow arrows with solid lines are for reconstruction of high quality references, and solid arrows with dashed lines are for the prediction in DCT domain.

In the LPLR mode (Mode 1), the SNR enhancement MB is predicted and reconstructed from the previous low quality reference. There is no drifting error in this mode, but the coding efficiency of this mode is low due to low quality temporal prediction.

In the HPHR mode (Mode 2), the SNR enhancement MB is predicted and reconstructed from the previous high quality reference. This mode can provide high coding efficiency. If all SNR enhancement MBs are encoded with this mode, the PFGS scheme can provide the highest coding efficiency at high bit rates. However, it may cause drifting errors if the high quality reference in the previous frame is not available due to network bandwidth or transmission errors.

The HPLR mode (Mode 3) is an extension of the drifting error reduction technique used in the original PFGS scheme[4][5] at the MB level. In this mode, the SNR enhancement MB is predicted from the previous high quality reference while reconstructed from the previous low quality reference at both the encoder and the decoder. This mode may affect the coding efficiency of the PFGS scheme. However, since the encoder and decoder can always obtain the same temporal prediction, the error propagated from the previous frames can be effectively eliminated in this mode

A mode-selection mechanism is proposed to choose the proper coding mode of each SNR enhancement MB. If a MB is encoded with INTRA mode in the base layer, the corresponding SNR enhancement MB is also encoded with INTRA mode. If a MB in the base layer is encoded with INTER mode, the proposed mode-selection scheme has to determine which INTER coding mode should be best used in the corresponding enhancement MB.

If the absolute mean of the predicted DCT residues produced in the LPLR mode is less than that in the HPHR/HPLR mode, the SNR enhancement MB should be coded using the LPLR mode; otherwise, the mode-selection scheme further determines the coding mode between HPHR and HPLR.

Obviously, the larger the difference between two temporal predictions is, the larger the quality loss will be caused when the previous high quality reference is not available. In order to control the possible quality loss, the mode-selection scheme defines a criterion as follows,

$$\|p_e(n) - p_b(n)\| > k \times \|x_o - r_e(n)\| \cdot \qquad (1)$$

Here x_o is the current original image, $p_b(n)$ and $p_e(n)$ denote the low and high quality prediction, respectively, $r_e(n)$ is the corresponding high quality reference for the next frame, and k is the acceptable loss factor. The factor k is an adjustable parameter, which controls the performance of the PFGS scheme at low bit rates and high bit rates. $\|x_o - r_e(n)\|$ is the mean squared error between the reconstructed high quality reference and the original image. When the difference between two temporal predictions is larger than the right-hand side value of Formula (1), this MB should be encoded with the HPLR mode, since the HPHR mode may cause a significant drifting error in this case. However, the coding mode of each MB is determined before the coding process, $r_e(n)$ in Formula (1) is generally not available during the process of mode decision. Therefore, the following criterion is applied to approximate Formula (1)

$$\|p_e(n) - p_b(n)\| > k' \times \|x_o - p_e(n)\| \cdot \qquad (2)$$

Where $r_e(n)$ is replaced by $p_e(n)$. Since the two temporal predictions in Formula (2) are already available, no additional computation is introduced to the PFGS encoder.

The coding mode information of each MB should be included in the MB header. Since the base layer bit-stream already provides the information about the INTRA mode, only the INTER mode information needs to be encoded. A simple VLC table is used to compress them into the enhancement bit-stream.

The significant coding efficiency improvement of the macroblock-based PFGS scheme is reported in [9]. It has been shown that the MB-based PFGS can achieve up to 2.0dB gain in PSNR over the FGS scheme in MPEG-4.

3. MACROBLOCK-BASED PFGS WITH TEMPORAL SCALABILITY

Due to the drastic variations of the Internet bandwidth and the dynamic changes in network conditions, it sometimes requires not only the image quality but also the frame rate should be scalable in the streaming video applications. As mentioned above, the MB-based PFGS is a highly efficient streaming video coding scheme. In order to cover an even wider range of bit rates and meet more requirements from the streaming applications, a corresponding scheme for MB-based PFGST is proposed in this paper. Besides the SNR scalability, the temporal scalability is also integrated into the MB-based PFGST. Similarly to FGST, the base layer of PFGST only contains the I-frames and P-frames coded at base frame rate, all B-frames are coded as temporal enhancement frames at the enhancement frame rate. The prediction errors of the temporal enhancement frames after bi-direction motion compensation are compressed with bit-plane coding technique for fine granularity scalability. However, different from the FGST scheme, the PFGST coding scheme makes use of the high quality SNR enhancement layer references for prediction in the temporal enhancement frames rather than always using the base layer.

Since there are two reconstructed references in each P-frame and I-frame, a temporal enhancement frame in the PFGST scheme can be predicted either from the reconstructed base layer or from the reconstructed enhancement layer. The B-frames that are not used in any prediction can benefit much from the PFGS structure because there are no drifting errors caused by using the high quality references in the temporal enhancement layer coding. Coding efficiency becomes the most important factor in the PFGST temporal scalable coding. Two coding modes, namely the HP mode (Mode 1) and the LP mode (Mode 2), are designed for the temporal enhancement macroblock coding corresponding to the different references used in the prediction. As shown in Figure 3, the layer above the dash-line denotes the base layer and those below the dash-line are the enhancement layers (including both SNR and temporal). Gray boxes indicate the reconstructed layers in the temporal base layer served as references, while white boxes indicate the temporal enhancement layers.

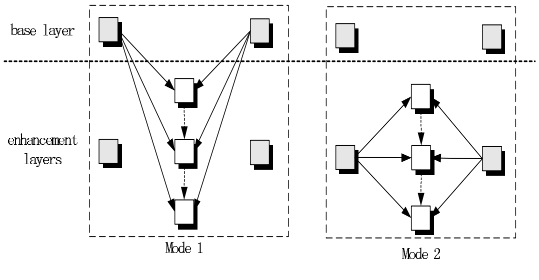

Figure 3: The coding modes for temporal scalable enhancement macroblock.

In the LP mode, the temporal enhancement MB is predicted from the low quality references in the temporal base layer, while in the HP mode, it is predicted from the high quality references in the temporal enhancement layer. Which reference should be used in the prediction of temporal enhancement MB is determined by the sum of the absolute values of the predicted residues in the DCT domain. If the sum of the absolute values of the DCT residues predicted from the reconstruct SNR base layer is lower than that predicted from the reconstructed SNR enhancement layer, the reference for the temporal enhancement MB is the SNR base layer (LP mode); otherwise, the reference for the temporal enhancement MB is the reconstructed SNR enhancement layer (HP mode).

It is clear that, in the PFGST scheme, every temporal enhancement MB can flexibly and easily choose their temporal references according to the above criterion. The mode information of the temporal enhancement MB is encoded in the temporal enhancement bit-stream using the same VLC table that mentioned in the previous section, that is, the binary “1” represents the mode 1 and the binary “01” represents the mode 2.

4. EXPERIMENTAL RESULTS

Extensive experiments have been performed to verify the performance of the proposed MB-based PFGST scheme. To compare with our proposed PFGST scheme, an FGST scheme described in [6] has also been analyzed and implemented. Both SNR and temporal scalabilities are supported in the experiments. For the two coding schemes, the same experiment conditions described in the following paragraphs are used.

Four MPEG-4 test sequences are used in the experiments: two sequences with CIF format (Foreman and Coastguard) and two sequences with QCIF format (Coastguard and Foreman).

In the base layer, only the first frame is encoded as I-frame, and the others are encoded as P-frame. All B-frames are encoded as temporal enhancement frames. The encoding frame rate is 10Hz and 30Hz for the base layer and the enhancement layer, respectively. Both the SNR enhancement layer and the temporal enhancement layer are written into one single bit-stream. The bit rate of the enhancement layers is not constrained in encoding process. Since the enhancement layer provides an embedded bit-stream, the bit-stream can be truncated at any position to fit in the channel bandwidth. The truncating process can be independent of the encoder.

For the CIF sequences, the range of motion vectors is set to within ±31.5 pixels. The bit rate of the base layer is 128kbits/s with TM5 rate controller. The enhancement layer bit-streams are truncated at bit rates of 512kbits/s, 768kbits/s, … , 2048kbits/s with an interval of 128kbits/s while maintaining a fixed frame rate of 30 frames/s.

For the QCIF sequences, the range of motion vectors is set to within ±15.5 pixels. The bit rate of the base layer is 32kbits/s with TM5 rate controller. The enhancement layer bit-streams are truncated at bit rates of 80kbits/, 128kbits, … , 368kbits/s with an interval of 48kbits/s while maintaining a fixed frame rate of 30 frames/s.

For the MB-based PFGST scheme, some other major experiment conditions are shown in Table 1.

The curves of the average PSNR at each bit rate are given in Figure 4. Compared with the FGST scheme, the MB-base PFGST scheme can achieve significant coding efficiency gain up to 2.8dB in PSNR.

5. CONCLUSIONS

This paper presents a flexible and efficient MB-based PFGS coding scheme that can provide both fine granularity SNR and temporal scalabilities for Internet video streaming. Firstly, the MB-based PFGS is introduced. With the three INTER modes and the corresponding mode-selection scheme, it can make a good trade-off between low drifting errors and high coding efficiency while preserving the excellent property of fine granularity SNR scalability. Building upon the MB-based PFGS scheme, the MB-based PFGST is presented in which the temporal scalability is further provided in addition to the SNR scalability. By using the two prediction modes in the temporal enhancement macroblock coding, the proposed PFGST scheme shows a great improvement over the FGST in MPEG-4. Experimental results show that, compared with the FGST scheme, the MB-based PFGST scheme can obtain the coding efficiency gain up to 2.8dB while maintaining the excellent property of fine granularity SNR and temporal scalabilities as well.

Table 1 The experiment conditions for MB-based PFGST.

Sequence (Resolution)	Bits to generate the high quality reference	Acceptable loss factor
Foreman (CIF)	≥20000	2.8
Coastguard (CIF)	≥20000	2.8
Foreman (QCIF)	≥4000	2.3
Coastguard (QCIF)	≥3000	2.3

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] W. Li, “Streaming video profile in MPEG-4”, IEEE trans. Circuit and systems for video technology, special issue on streaming video (accepted).

[3] W. Li, “Fine granularity scalability in MPEG-4 for streaming video”, ISCAS 2000, Geneva, Switzerland, May 28-31, 2000.

[4] F. Wu, S. Li and Y.-Q. Zhang, “A framework for efficient progressive fine granularity scalable video coding”, IEEE trans. Circuit and systems for video technology, special issue on streaming video (accepted)

[5] F. Wu, S. Li and Y.-Q. Zhang, “DCT-prediction based progressive fine granularity scalable coding”, ICIP2000, Vancouver, British Columbia, Canada, Sep. 10-13, 2000.

[6] M. van der Schaar and H. Radha, “A novel MPEG-4 based hybrid temporal-SNR scalability for Internet video”, ICIP2000, Vancouver, British Columbia, Canada, Sep. 10-13, 2000.

[7] Text of ISO/IEC 14496-2 “MPEG-4 Video FGS v.40,” N3317, Proposed Draft Amendment (PDAM), Noordwijkerhout, the Netherlands; March 2000.

[8] Mihaela van der Schaar, Hayder Radha, Yingwei Chen, “An all FGS solution for hybrid Temporal-SNR scalability”, ISO/IEC JTC1/SC29/WG11, MPEG99/m5552, Maui meeting, November, 1999

[9] F. Wu, S. Li, X. Sun, R. Yan and Y,-Q. Zhang, “Macroblock-based progressive fine granularity scalable coding”, ISO/IEC JTC1/SC29/WG11, MPEG2001/N6779, Pisa meeting, January, 2001.

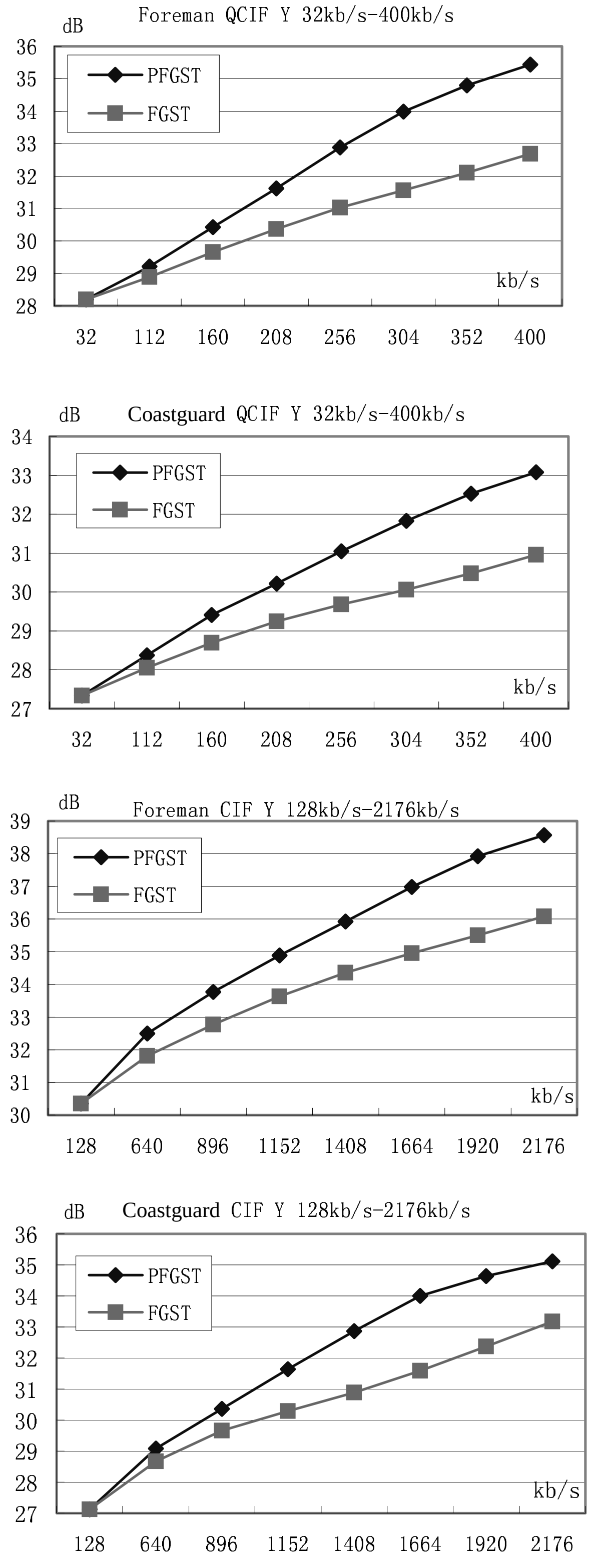

Figure 4: PSNR versus bit rate comparison between the MB-based PFGST and FGST.