EFFICIENT AND UNIVERSAL SCALABLE VIDEO CODING

Feng Wu[1]*, Shipeng Li*[1]*, Rong Yan*[2]*, Xiaoyan Sun*[3]*, Ya-Qin Zhang*[1]
[1] Microsoft Research Asia, Beijing, 100080, China
[2] Beijing Institute of Technology, Beijing 100081, China
[3] Harbin Institute of Technology, Harbin, 150001

ABSTRACT

This paper proposes a unified efficient and universal scalable video coding framework that supports different scalabilities, such as fine granularity quality, temporal, spatial and complexity scalabilities. The proposed framework is established upon the recent studies in fine granularity scalable (FGS) video coding. It contains two key points. Firstly, in order to improve the coding efficiency of the proposed framework, more than one motion compensation loop is used. Since high quality references are introduced into the enhancement layer coding, the proposed framework can efficiently compress different-resolution video at different layers for the purpose of the complexity and spatial scalability. Secondly, the drifting reduction techniques are studied in this paper. This helps the proposed framework to maintain good performance at lower enhancement bit rates. By defining coding modes, a macroblock level control mechanism is developed to achieve a better trade-off between low drifting errors and high coding efficiency.

1. INTRODUCTION

With the rapid developments in computers and networks, more and more users expect to enjoy multimedia services through various PC or non-PC devices over the Internet or wireless networks. However, this kind of ubiquitous multimedia services post great challenges to the conventional video coding techniques. Since the current Internet is a heterogeneous network, the connection speed between servers and clients may vary in a wide range. This requires video coding techniques provide different video bit rates according to the available channel bandwidth. In general, many non-PC devices only have low-resolution screen and limited computational power. This further requires that video coding techniques provide efficient video representation with different resolutions and decoding complexities.

A straightforward solution to meet the above requirements would be to compress the same video sequence into many bitstreams for every possible bit rate, resolution and device complexity. The video server chooses an appropriate bitstream to transmit to an individual user according to the actual connection speed and device capability. Obviously, this would be a great waste of system resources. On the other hand, even if we could manage to store all these bitstreams in the server, in a dynamically changing non-QoS guaranteed network, we still could not provide the best available video quality with a single bitstream. Therefore, an active research topic in video coding field is how to efficiently compress video sequences with different scalabilities, such as rate, quality, temporal, spatial and complexity scalabilities.

Various scalable video coding techniques have been developed rapidly in the past decade. Spatial and temporal scalable coding techniques that provide video at different resolutions and frame rates were accepted in some main video coding standards, such as MPEG-2, MPEG-4 and H.263++ [1]~[3]. In addition, fine granularity scalable (FGS) video coding techniques have been extensively studied in recent years. MPEG-4 standard already accepted it in the streaming video profile (SVP) [4]. Since both the base layer and the enhancement layer are always predicted from the base layer, the enhancement bitstream can be arbitrarily truncated in any frame without error propagation. MPEG-4 FGS provides a remarkable capability in readily and precisely adapting to channel bandwidth variations. However, this also makes MPEG-4 FGS suffer from severe degradations in coding efficiency. Furthermore, MPEG-4 FGS has difficulty to compress different-resolution video at different layers; otherwise the coding efficiency at the enhancement layer would be further degraded.

Progressive Fine Granularity Scalable (PFGS) coding scheme [5][6] is a significant improvement over FGS by introducing two prediction loops with different quality references. This paper proposes an efficient and universal scalable video coding framework developed based on the PFGS investigations. Firstly, more than one enhancement layers based on a common base layer are used to implement fine granularity quality, temporal, spatial and complexity scalabilities within the same framework. In order to achieve high coding efficiency for various scalabilities, multiple prediction loops with different quality references are employed in the proposed framework. By utilizing high quality reference in the spatial enhancement layer coding, the proposed framework not only fulfills efficient spatial scalability but also supports complexity scalability. The lower complexity bound is the low-resolution base layer decoding, which can be sufficiently low for many applications. Different resolution enhancement layers provide users different decoding complexities. Furthermore, the complexity of every enhancement layer can also be adjusted according to the number of bits decoded.

Secondly, with the introducing of part of the enhancement bitstream into the motion compensation loop, the drifting errors are inevitable when part of the reference enhancement bitstream is unavailable to the client due to packet losses or bandwidth drop. Therefore, different drifting reduction techniques are discussed in this paper. A new macroblock level control mechanism is proposed. By defining various coding modes at the enhancement layer, the different techniques can be combined together to achieve a better trade-off between high coding efficiency and low drifting errors.

This paper is organized as follows. Section 2 briefly describes the proposed efficient and universal scalable video coding framework. The new control mechanism is discussed in Section

3. Experimental results are given in Section 4. Finally, Section 5 concludes this paper.

2. THE EFFICIENT AND UNIVERSAL SCALABLE VIDEO CODING FRAMEWORK

Figure 1 illustrates the proposed efficient and universal scalable video coding framework. The input video with two resolutions is compressed in the proposed framework. Narrow rectangles denote the low-resolution video, and wide rectangles denote the high-resolution video. There are four different enhancement layers sharing a common base layer. The bottom layer is the base layer. It is usually generated as the lowest quality, lowest resolution, least smoothness and least complexity. The quality enhancement layer compresses the same resolution video as that at the base layer. It will improve the decoded quality of the base layer. The temporal enhancement layer improves the base layer frame rate and makes the decoded video look smooth. The rest two enhancement layers improve the video quality and frame rate at high-resolution. These two enhancement layers are optional in the proposed framework and appear only if the video with two different resolutions is encoded.

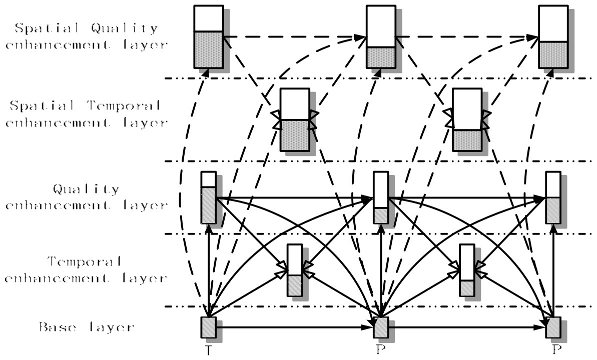
Spatial Quality enhancement layer
Spatial Temporal enhancement layer
Quality enhancement layer
Temporal enhancement layer
Base layer
I
P
P

Figure 1: The proposed efficient and universal scalable video coding framework.

Except that the base layer is encoded with the conventional DCT transform plus VLC technique, all of the enhancement layers are encoded with DCT transform plus the bit-plane coding technique. In other words, every enhancement layer bitstream can be arbitrarily truncated in the proposed framework. In Figure 1, each rectangle denotes the whole frame bitstream at one enhancement layer. The shadow region is the actual transmitted part, whereas the blank region is the truncated part. Hence the proposed framework provides the most flexible bit rate scalability.

The key problem in the proposed framework is still how to deal with motion compensation. If, as in MPEG-4 FGS, all layers are always predicted from the lowest quality base layer, the coding efficiency of the enhancement layers will be a big issue for most applications. Recent investigations show that part of the encoded information at the enhancement layer can be introduced into the prediction loop at the enhancement layer [7][8]. This will significantly improve the coding efficiency of the enhancement layer at moderate and high bit rates. Furthermore, error propagation can be effectively controlled by drifting reduction technique at lower enhancement bit rates [9]. On the other hand, some techniques were also proposed to use the information at the enhancement layer for the base layer prediction [10]~[12]. The two categories of techniques are compared in [13]. Obviously, utilizing the enhancement information for the base layer coding will be helpful to improve the coding efficiency at moderate and high bit rates. But, how to control the drifting errors at the base layer is still under studies.

In the proposed framework, every layer including the base layer can select the prediction from the two different quality references. As shown by solid arrows with solid lines in Figure 1, the base layer and the quality enhancement layer use the reconstructed base layer and the reconstructed quality enhancement layer at a given bit rate as references. As shown by hollow arrows with solid lines, the temporal enhancement layer is bi-directional predicted from the base layer and the quality enhancement layer. The predictions for the two high-resolution enhancement layers are denoted by solid arrows with dashed lines and hollow arrows with dashed lines, respectively.

3. THE CONTROL MECHANISM AT MACROBLOCK LEVEL

The main difference among the efficient fine granularity scalable video coding frameworks developed in recent years lies in the usage of reference for prediction and reconstruction. In this paper, a control mechanism is proposed at macroblock level to integrate these schemes into a unified framework. Various coding modes are defined in the base layer and the enhancement layer coding. Each coding mode has its unique reference for prediction and reconstruction. In other words, each base or enhancement macroblock informs the client which reference is used through the coding mode.

3.1 The definition of coding mode

Figure 2 illustrates five different coding modes for the base layer and the quality enhancement layer coding. They are derived from the fine granularity scalable video coding schemes proposed in [4][7][8][10][11]. The rectangular boxes in the first row denote the base layer, and the rectangular boxes in other rows denote the bit planes of the quality enhancement layer. Gray rectangular boxes indicate those to be reconstructed as references. Solid arrows with solid lines are for temporal predictions, hollow arrows with solid lines are for reconstruction of high quality references, and solid arrows with dashed lines are for predictions in DCT domain.

In Mode 1, all layers are both predicted and reconstructed from the previous low quality reference. This mode is extensively used in all schemes except for [10]. Since the base layer bit rate is usually very low, it is reasonable to assume that the low quality reference is always available to the client. Therefore, there is no drifting error in this mode. The coding efficiency of this mode is low due to low quality temporal prediction.

In Mode 2, the base layer is predicted and reconstructed from the previous low quality reference, but the quality enhancement layer is predicted and reconstructed from the previous high quality reference. This mode is used in the schemes [7] and [8]. It can significantly improve the coding efficiency at moderate and high bit rates. There is no drifting error at the base layer. When the channel bandwidth is not high enough to transmit the high quality reference, this mode would cause drifting errors at the quality enhancement layer.

In Mode 3, all layers are predicted and reconstructed both from the previous high quality reference. This mode is used in the schemes [10] and [11]. If the high quality reference is available to the client, it can provide better coding efficiency. However, when the channel bandwidth somehow drops, the high quality reference cannot be completely transmitted to the client. The drifting errors would inevitably appear at both the base layer and the quality enhancement layer.

In Mode 4, the enhancement layer is predicted from the previous high quality reference, while reconstructed from the previous low quality reference at both encoder and decoder. This mode was proposed in [7] for the purpose of drifting reduction. Since the low quality reference is always consistent at both the encoder and the decoder, the drifting errors caused by Mode 2 can be greatly reduced with Mode 4. Therefore, Combination of Mode 2 and Mode 4 can achieve a trade-off between low drifting errors and high coding efficiency.

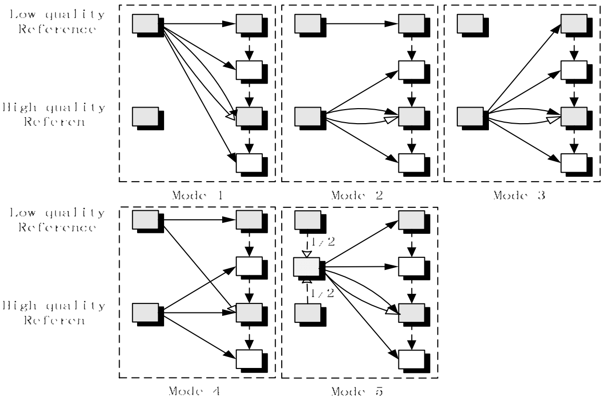

Figure 2: The various coding modes for the base layer and the quality enhancement layer coding.

In Mode 5, a new reference is generated with the average of the low quality reference and the high quality reference as shown by the rectangular box with mesh shadow in Figure 2. All layers are predicted and reconstructed from the generated reference. This mode is proposed in [8] and [11]. Since the coded enhancement information has a contribution in the generated reference, this would also causes drifting errors at the base layer and the quality enhancement layer when the high quality reference is not available at the decoder.

More coding modes can be readily added in the proposed control mechanism to generalize into more scalable applications as long as they have the virtue in improving coding efficiency or reducing error propagation. For example, Mode 5 suggests a general mode with the parameter interpolation instead of the average. The interpolation factors can be encoded into the bitstream. In fact, modes 1, 3, 5 all become special cases of the generalized mode. MPEG-4 SVP also contains the fine-grain temporal scalable coding technique, namely, MPEG-4 FGST [14]. Since the temporal frames are predicted from the reconstructed base layer, MPEG-4 FGST suffers from severe degradation in coding efficiency too. Similar to the PFGS technique, an improved fine granularity temporal scalable coding was proposed in [15]. These two coding techniques are special cases for the temporal enhancement layer by defining some coding modes. The fine-grain spatial scalable video coding developed in [16] is a special case in the spatial-quality enhancement layer of the proposed framework.

3.2 Drifting reduction

The proposed framework utilizes the enhancement information to improve the coding efficiency of the scalable video coding. The cost is to possibly cause severe drifting errors when the high quality reference is not available. There are two categories of drifting errors in the proposed framework. The first one is the drifting errors occurring at the enhancement layer due to Mode 2, Mode 3 and Mode 5. The second one is the drifting errors occurring at the base layer due to Mode 3 and Mode 5.

The straightforward method to eliminate the drifting errors is the INTRA refresh. But, inserting too many INTRA macroblocks will affect the performance of the proposed framework due to the low efficiency of the INTRA coding. In scalable video coding, one assumption is that the low bit rate base layer can always be correctly and completely transmitted to the client. Therefore, one important method is to exploit the available base layer to reduce and even eliminate the drifting errors. If there is no drifting error at the base layer, both Mode 1 and Mode 4 can effectively reduce the drifting errors occurring at the enhancement layer.

Leak prediction is another good technique to reduce the drifting errors [18][19]. The basic principle is to gracefully decay the enhancement information introduced in the prediction loop so as to limit the error propagation and accumulation. This technique can control the drifting errors in the given range. Some scalable coding schemes with this technique show very promising results in achieving a better trade-off between low drifting errors and high coding efficiency [19][20].

The drifting errors at the base layer are hard to reduce and control. Besides the simple INTRA refresh, theoretically, the principle of Mode 4 and Leak prediction can be employed to reduce the drifting errors at the base layer. The performance and the negative effects on coding efficiency are being investigated

4. EXPERIMENTAL RESULTS

The proposed efficient and universal scalable video coding framework is verified in this section. The MPEG-4 test sequence Foreman with CIF format is used in the experiment. The base layer codec is the MPEG-4 advanced simple profile. Only the first frame is encoded as I frame, and the rest of frames are encoded as P frames. TM5 rate control method is used in the base layer coding. The range of motion vectors is limited to ±31.5 pixel with half pixel precision.

Figure 3 shows the performance of the quality enhancement layer. The base layer bit rate is 128kbps, and the coding frame rate is 10Hz. Only Mode 1, Mode 2 and Mode 4 are used in the quality enhancement layer. In other words, there is no drifting error at the base layer. The high quality is reconstructed at the bit plane where the bits generated are more than 20000. At 768kbps, the proposed quality enhancement layer outperforms MPEG-4 FGS up to 1.8dB. The coding efficiency gap between the proposed framework and the non-scalable coding is about 1.0dB.

Figure 4 shows the performance of the temporal enhancement layer. The base layer bit rate is 128kbps, and the coding frame rate is 10Hz. There are two temporal frames between two I/P frames. The reconstructed quality of two references is the same as that in Figure 3. The proposed temporal enhancement layer outperforms MPEG-4 FGST up to 2.5dB.

The performance of the spatial-quality enhancement layer is shown in Figure 5. In this experiment, the low-resolution video (QCIF) is compressed at the base layer, whereas the high-resolution video (CIF) is compressed at the enhancement layer. In the MPEG-4 spatial scalable coding, the base layer bit rate is 32kbps, and the enhancement layer bit rate is 64Kbps. The spatial-quality enhancement layer can be arbitrarily truncated. The overall coding efficiency of the proposed spatial-temporal enhancement layer is better than that of the MPEG-4 spatial scalable video coding. Furthermore, the decoded video quality can be smoothly improved as channel bandwidth increases.

5. CONCLUSIONS

This paper proposed an efficient and universal scalable video coding framework. It can simultaneously support fine-granularity quality, temporal, spatial and complexity scalabilities. In the framework, a control mechanism is developed at macroblock levels. By defining various coding modes, the proposed framework unified the efficient fine granularity scalable video coding techniques developed in recent years. The initial experimental results show that the proposed framework can significantly improve the efficiency of video coding with various scalabilities. Furthermore, the latest studies show, if the proposed framework is integrated with H.26L, its coding efficiency can be further improved up to 4.0dB higher than that of MPEG-4 FGS [21].

6. REFERENCES

[1] MPEG video group, “Information technology - Generic coding of moving pictures and associated audio”, ISO/IEC 13818-2, International standard, 1995.

[2] MPEG video group, “Generic coding of audio-visual objects” ISO/IEC JTC1/SC29/WG11 N2502, Nov. 1998.

[3] ITU-T, “Recommendation H.263: Video coding for low bit rate communication”, version 2, March, 1993.

[4] W. Li, “Streaming video profile in MPEG-4”, IEEE trans. CSVT, Vol 11, no 3, 301-317, 2001.

[5] F. Wu, S. Li, Y.-Q. Zhang, “DCT-Prediction based progressive fine granularity scalable coding”, ICIP2000, vol 3, pp566-569, Vancouver, Sep, 2000.

[6] F. Wu, S. Li, Y.-Q. Zhang, “A framework for efficient progressive fine granularity scalable video coding”, IEEE trans. CSVT, Vol. 11, no 3, 332-344, 2001.

[7] X. Sun, F. Wu, S. Li, W. Gao, Y,-Q. Zhang, “Macroblock-based progressive fine granularity scalable video coding”, ICME2001, Japan, Aug 2001.

[8] W. S. Peng, Y. K. Chen, “Mode-adaptive Fine Granularity Scalability”, ICIP 2001, 993-996, Greece, Oct. 2001.

[9] F. Wu, S. Li, B. Zeng, Y.-Q. Zhang, “Drifting Reduction in Progressive Fine Granular Scalable Video Coding”, Picture Coding Symposium (PCS), Seoul, April 2001.

[10] R. Kalluri, M. Schaar, “Single-loop motion-compensated based fine-granular scalability (MC-FGS) with cross-checked results”, ISO/IEC JTC1/SC29/WG11, m6831, Pisa, Jan. 2001.

[11] A. Reibman, L. Bottou, A. Basso, “DCT-based scalable video coding with drift”, ICIP 2001, 989-992, Greece, Oct. 2001

[12] A. Reibman, L. Bottou, “Managing drift in DCT-based scalable video coding”, Data Compression Conference 2001, 351-360, Salt Lake City, USA, April 2001.

[13] F. Wu, S. Li, X. Sun, R. Yan, Y,-Q. Zhang, “Comparisons between the one-loop and two-loop solutions for improving the coding efficiency of FGS”, The second workshop and exhibition on MPEG-4, 79-82, San Jose, June, 2001.

[14] M. Schaar, H. Radha, “A hybrid temporal-SNR fine-granular scalability for Internet video”, IEEE trans. CSVT, Vol 11, no 3, 318-331, 2001.

[15] X. Sun, F. Wu, S. Li, W. Gao, Y-Q. Zhang, “Macroblock-based temporal-SNR progressive fine granularity scalable video coding”, ICIP2001, 1025-1028, Greece, Oct. 2001.

[16] R. Yan, F. Wu, S. Li, R. Tao, Y. Wang, “Efficient Video Coding with Hybrid Spatial and Fine-Grain SNR Scalabilities”, SPIE Visual Communications and Image Processing, 2002. (to appear)

[17] A. Fuldseth, T. Ramstad, “Robust subband video coding with leak prediction”, DSP workshop, 57-60, Norway, 1996.

[18] M. Ghanbari and V. Seferidis, “Efficient H.261-based two-layer video codecs for ATM networks”, IEEE trans. CSVT, Vol 5, no 2, 171-175, 1995.

[19] S. Han, B. Girod, “SNR scalable coding with leaky prediction”, ITU-T Q.6/SG16, VCEG-N53, Santa Barbara, USA, 2001.

[20] H. Huang, C. Wang, T. Chiang, “A robust fine granularity scalability using trellis based predictive leak”, submit to ISCAS 2002.

[21] Y. He, R. Yan, F. Wu, S. Li, “H.26L-based fine granularity scalable video coding”, ISO/IEC JTC1/SC29/WG11, m7788, Pattaya, Dec. 2001.

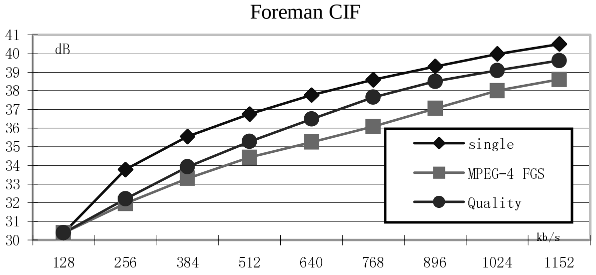

Figure 3: The PSNR verses bit rate at the quality enhancement layer.

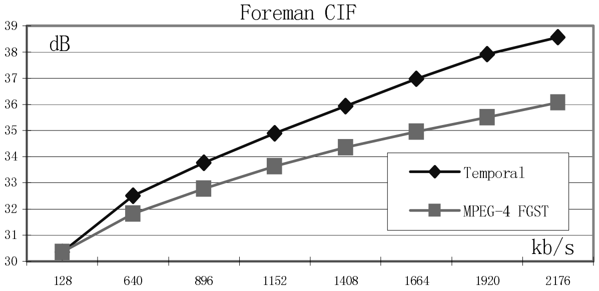

Figure 4: The PSNR verses bit rate at the temporal enhancement layer.

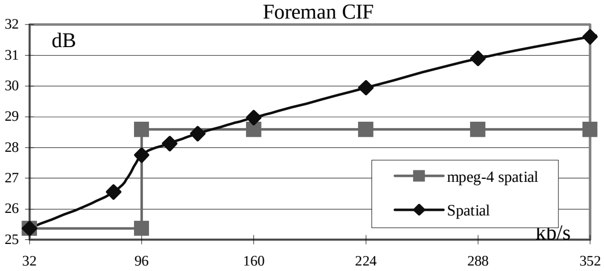

Figure 5: The PSNR verses bit rate at the spatial-quality enhancement layer.