Improvements of Multiple FGS Layers Coding for Low-Delay Applications in SVC

Yanyan Zheng*[1], Xiangyang Ji[1], Feng Wu[2], Debin Zhao[1], Wen Gao[1]

[1]Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

[2]Microsoft Research Asia, Beijing, China

ABSTRACT

For low-delay applications in Scalable Video Coding, there are two alternative coding strategies based on AR-FGS for compressing multiple FGS layers. However, their coding efficiencies both suffer largely from the inherent drifting errors. In this paper, a more efficient multiple FGS layers coding structure is presented, which can provide higher coding performance within wide bitrate range and stronger error resilience. This is achieved by incorporating the partial-reconstructed enhancement layer references, instead of the complete-reconstructed ones, into the motion-compensated prediction loop of the FGS layers. Thereupon the prediction drift can be effectively decreased, especially for the middle bitrate points. Further, through selecting different-quality enhancement layer references generated with cycle-based reconstruction mechanism, more flexible video quality can be supported to be more suitable for varied practical application requirements.

Keywords: FGS, SVC, leaky, low delay, error resilience, partial reconstruction

1. INTRODUCTION

The transmission of multimedia data over wireless channels or Internet has high probability to encounter bandwidth fluctuation and packet loss/error. To address such kind of problems, the bit-plane coding based MPEG-4 FGS [1] is first utilized, whose key feature of error resiliency is attained at the expense of lower coding performance compared to that of the non-scalable coding scheme, because no SNR enhancement layer information is introduced into the temporal prediction loop of either the base layer or the enhancement layer. Therefore, in order to improve the compression efficiency for the emerging standard JVT SVC [2], temporal prediction is optional to be employed when coding FGS layer; and a more efficient FGS entropy coding method called cyclical block coding [3] has replaced the original bit-plane coding in SVC.

For real time video communication applications, such as video conference, live broadcasting and instant video chat, the primary requirement for such kind of applications is to minimize the end-to-end delay which is crucial to ensure good interaction among the participants. Consequently the video sequences are encoded as either I-slice or P-slice without hieratical B coding structure which will incur a long delay [6]. And the benefit to introduce FGS coding in low-delay applications is that it can effectively support video transmission in changing channel. Nevertheless, the FGS coding efficiency of conventional close-loop P frame is unsatisfactory. Therefore, an efficient FGS coding scheme with temporal prediction called AR-FGS (fine granularity scalability with adaptive reference) [4] is adopted in SVC, which can generate more accurate prediction signal for FGS layer in key frames, since it is formed adaptively from the base layer and enhancement layer references based on the information coded in the base layer.

In order to further satisfy the diversified application demands, multiple FGS layers coding are often necessary to flexibly offer wider range of video quality according to the allowed bandwidth. Currently, there are two alternative structures [5] for coding multiple FGS layers in current JSVM. However, their common deficiency is the resulted low coding efficiency incurred by severe prediction drift in case of partial decoding. In this paper, based on the exiting work, we propose a more efficient multiple FGS coding strategy for practical low-delay applications. In the proposed coding structure, in order to better control error propagation, the second FGS layer utilize the partial-reconstructed pictures as references in its temporal prediction loop, instead of the complete-reconstructed frames which have the highest quality.

Visual Communications and Image Processing 2007, edited by Chang Wen Chen, Dan Schonfeld, Jiebo Luo,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6508, 65082Y, © 2007 SPIE-IS&T · 0277-786X/07/$18

And a small part of the second FGS layer data is incorporated into the first FGS layer prediction with appropriate leaky factor to enhance the coding performance of the high bitrate points. The advantage of this coding strategy is that, except around the highest bit-rate point, the coding performance can be greatly improved especially for the middle and higher bitrate range, at which a well performed FGS coding with better video quality is more appreciated, since FGS layers are usually truncated when large amount of multimedia meet limited bandwidth in real applications.

2. REVIEW OF EXISTING MULTIPLE FGS LAYERS CODING STRUCTURES

2.1 Multi-loop FGS coding structure

One straightforward method of coding multiple FGS layers is to perform motion compensation (MC) at each FGS layer, utilizing the current FGS layer and its subordinate FGS or base layer reference frames to form its temporal prediction, as illustrated in Fig.1. Although the compression efficiency of this structure approaches optimal, the coding complexity is too high to be suitable for practical applications, since it needs multi-loop MC both at encoder and decoder. (For clearness, two FGS layers coding is taken for instance to analyze the coding methods.)

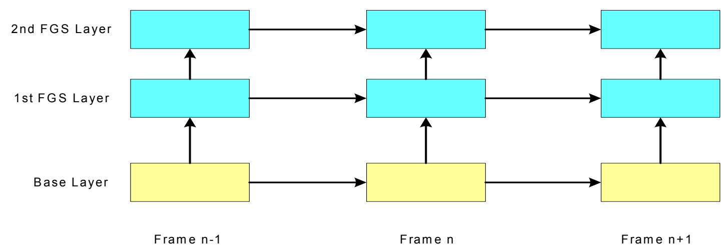


Fig.1. Multi-loop FGS coding structure.

2.2 Two-loop FGS coding structure

Considering decoding complexity, two multiple FGS layers coding structures, as shown in Fig.2 and Fig.3, are adopted in SVC. For two-loop FGS coding structure in Fig.2, there is only one motion compensation for multiple FGS layers which occurred when coding the first FGS layer, whose prediction signal is calculated using the highest FGS layer and base layer reference frames. Despite its low complexity of two-loop encoder and decoder, the coding performance of this two-loop structure is far from satisfactory, due to severe prediction drift in case of partial decoding.

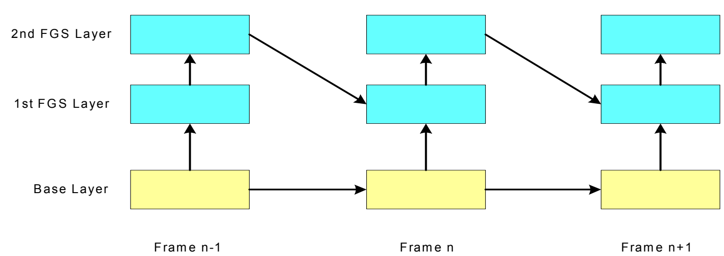


Fig.2. Two-loop FGS coding structure.

2.3 Decoder oriented two-loop FGS coding structure

The alternative method as illustrated in Fig.3 is more efficient than that in Fig.2 when coding multiple FGS layers. In this so-called *decoder oriented two-loop structure*, the temporal predictions for each FGS layer are generated using the current FGS layer and base layer reference frames. Thereupon, motion compensations are required for all FGS layers and base layer at the encoder side; while at the decoder side only the current highest FGS layer that can be received as well as the base layer needs MC for correctly decoding. Compared to the coding structure in Fig.2, the advantage of this structure is that the prediction drift is confined within each FGS layer, which can help prevent error propagation and thus yield better coding performance for the bit-rate points near the end of first FGS layer, but at the expense of lower compression efficiency for the higher bit-rate points of the second FGS layer.

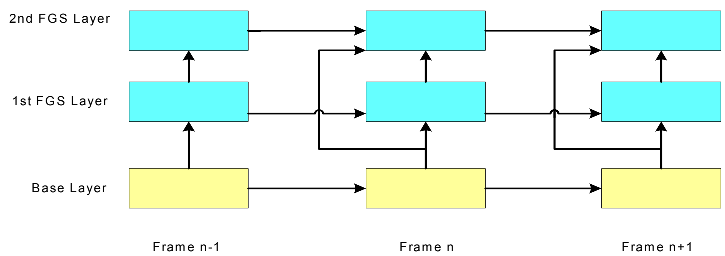


Fig.3. Decoder oriented two-loop FGS coding structure (**anchor)**.

To sum up, first, although multi-loop FGS coding structure in Fig.1 has the best coding efficiency, its decoding process needs at least three motion compensation loop and thus result in much higher decoding complexity, which can not be endured as part of standard. Secondly, the prediction drifts of the coding structure in Fig.2 are too large to promise eligible coding performance; despite it has the lowest encoding and decoding complexity. Finally, the *decoder oriented two-loop FGS coding structure* in Fig.3 is recommended to be used in SVC as a tradeoff between coding efficiency and complexity. Nevertheless, the inherent problem of prediction drifts still exits and limits the further improvement of its coding performance over wide bitrate range.

3. PROPOSED MULTIPLE FGS LAYERS CODING STRUCTURE

3.1 Cycle based two-loop FGS coding structure

In this paper, we propose a more efficient and flexible multiple FGS layers coding strategy, called *cycle based two-loop structure*, as illustrated in Fig.4. Different from *decoder oriented two-loop structure*, when coding the second FGS layer, temporal prediction signal is adaptively formed using the base layer and the partial-reconstructed (dark green) reference pictures of the second FGS layer, instead of the complete-reconstructed (weak green) ones that have the highest quality.

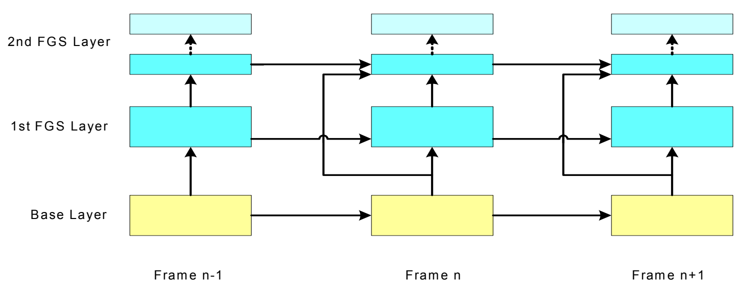


Fig.4. Cycle based two-loop FGS coding structure (**proposed)**.

Specially speaking, at the encoder side, each FGS layer has to perform motion compensation to generate its prediction signals. For the first FGS layer, the references are formed with the base layer and the first FGS layer pictures; while for the second FGS layer, it is the partial-reconstructed second FGS layer and the base layer that produce the actual references.

Accordingly, at the decoder side, if there are no bits from the second FGS layer arrived at the decoder, to decode the first FGS layer, two motion compensation loops have to be performed respectively at the base layer and the current first FGS layer to generate the reconstructed pictures both for displaying and references of the next frame; while if there are bits of the second FGS layer received by the decoder, and these bits all belong to the partial-reconstructed picture (weak green) used as reference for the next frame at encoder, only the base layer and the second FGS layer need to perform motion compensation in the same way as the previous case. Nevertheless, if there are bits of the second FGS layer that have not contributed to the generation of the partial-reconstructed references at encoder, then one additional IDCT and IQ would be needed to produce the partial-reconstructed pictures without these bits for the next frame used as references, and the complete-reconstruted pictures generated with these bits are utilized for displaying.

To note that, the purpose of motion compensation is to produce the prediction signals for either the base layer or the current highest FGS layer, so there will always just two MC loops at decoder side. But we can get different-quality reconstructions of the second FGS layer with or without those bits in the weak green part of Fig.4. The partition of the second FGS layer into the dark green and weak green is achieved by introducing one additional cycle-aligned reconstruction operation. The quality of the partial-reconstruted picture (weak green) can be adjusted by the cycle index [3], below which all bits are used to generate it. Therefore, in this paper, we name the proposed approach *cycle based two-loop FGS coding structure.* An illustrated example of the quality of the first and second FGS layer, as well as the cycle-based reconstructed pictures are displayed in Fig.5, in which the red line marked 'Cycle' has nearly the middle quality of the two ones marked '1st FGS' and '2nd FGS'.

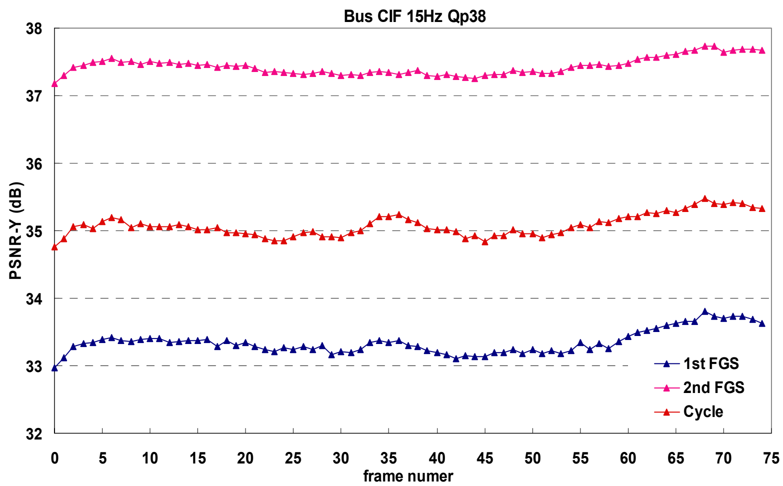


Fig.5. Comparison of the quality of the first, second FGS layer and the partial-reconstruted pictures.

3.2 Optimization of cycle based two-loop FGS coding structure

Theoretic analysis and experimental results both show that, there would be loss of coding efficiency near the high end of each FGS layer, if the enhancement layer references are made up of the partial-reconstructed pictures instead of the complete-reconstructed ones. It is because that, high-quality references generally produce accurate prediction signals, and result in residual pictures with low-power energy, thus lead to high compression ratio. Thereupon, in order to make up this coding loss, one effective technique is to incorporate a small part of the second FGS layer data into the motion compensation loop of the first FGS layer to construct its prediction signal more correctly with appropriate leaky factor γ, as shown in Fig.6:

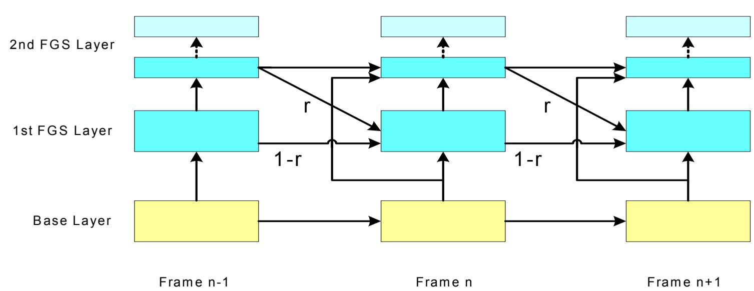


Fig.6. Optimized cycle based two-loop FGS coding structure (**proposed**).

Different from anchor method in Fig.3, the enhancement layer reference frame of the first FGS layer is calculated according to (1):

$$R_e^{n-1} = (1-\gamma)\cdot R_{F_1}^{n-1} + \gamma \cdot R_{P_2}^{n-1} \quad (1)$$

in which $R_{F_1}^{n-1}$ and $R_{P_2}^{n-1}$ respectively stands for the first FGS layer reference frame and the partial-reconstructed reference frame which belongs to the second FGS layer, and γ is their corresponding leaky factor, whose value should be set small to keep proper balance between the coding efficiency of the first and second FGS layer. After attaining the enhancement layer reference frame R_e^{n-1}, as done in anchor methods, the final adaptive reference frame R_A^n is computed by adding current base layer reconstruction frame X_b^n to the scaled motion-compensated difference between the enhancement layer reference R_e^{n-1} and the base layer reference R_b^{n-1} with leaky factor α, according to (2):

$$R_A^n = X_b^n + \alpha * MC(R_e^{n-1} - R_b^{n-1}) \quad (2)$$

The purpose of this technique is to try to take advantages of both anchor approaches in Fig.2 and 3. In fact, a good tradeoff between them, adjusted by the newly-introduced leaky parameter γ, can help to obviously enhance the coding efficiency of the second FGS layer, especially for the high bitrate points; though the coding performance is negatively influenced slightly for narrow bitrate range around the end of the first FGS layer. Besides, for encoder and decoder matching, a loop filter is performed at encoder when reconstructing the first FGS layer to offer smoother reconstructed references for better forming the leaky-based temporal prediction.

3.3 Error resilience of cycle based two-loop FGS coding structure

There are two leaky factors α and β of AR-FGS corporately used to adjust the tradeoff between the coding efficiency and error resilience, based on the information coded in the base layer [4]. Actually they perform much better than other approaches, but there still exits large error propagation in each FGS layer once the video stream can not be completely received by the decoder. As indicated in Fig.7, no matter which pair of leaky factors of AR-FGS is used, the rate-distortion curves always tend to show concave (low coding quality in PSNR compared to high bitrate), especially in the middle bitrate range of the second FGS layer. This is because the worst case usually corresponds to the situation that the encoder use the whole FGS layer information as references, but only half of them arrived to decoder, which scenario directly leads to the major loss of quality.

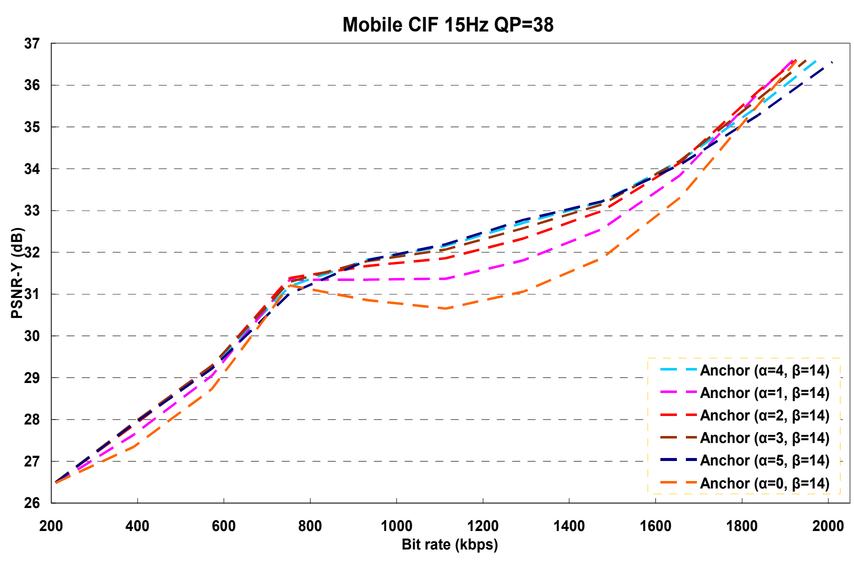


Fig.7. R-D curves with different leaky factors of decoder oriented two-loop structure.

Therefore, in order to avoid this worst case happening to the proposed structure, the cycle index, which is used as the boundary between the partial-reconstruted pictures and the complete-reconstruted ones, is generally selected on the rule of generating the partial-reconstruted references with approximately half amount of bits that spent in coding the second FGS layer, as indicated in Fig.5. Consequently, in this way, the introduced partial-reconstructed references can help further prevent the propagation of drifting errors of the second FGS layer, since the employed partial-reconstructed references can be more accurately reproduced in most cases of partial decoding from the view of decoder side and then the largest prediction error existing in the middle bitrate range can be emphatically mitigated.

Furthermore, besides better coding performance and stronger error resilience, another benefit of the proposed approach is that, the quality of the partial-reconstructed references can be adjusted by pointing different cycle index. Experimental results show that the bigger the cycle index is set, the closer the highest-point coding result of the proposed method is to that of anchor method; and accordingly a convex hull rather than a concave is lifted up to the higher bitrate points. Therefore, if the transmission bandwidth is limited and varied with time, then the corresponding cycle index can be set smaller to achieve better coding quality at low and middle bitrate range; while if the channel condition is good enough, the cycle index can be set bigger to obtain higher coding efficiency at the middle and high bitrate points. In all, the cycle-based reconstruction mechanism can offer flexible coding quality according to different practical requirements, which feature is particularly welcome by the low-delay applications.

3.4 Complexity of cycle based two-loop FGS coding structure

At encoder side, the additional operations against to *decoder orientated two-loop*, are inverse DCT transform and inverse quantization as well as a reference frame buffer storing the partial reconstructed reference which in fact can be technically saved in practical implementation. Particularly all these are needed only when coding each frame's second FGS layer. At decoder side, all the above will not be needed at all, except that current to be decoded frame has higher quality than its version as partial reconstructed references employed at encoder, which corresponds to the last case discussed in section 3.1 and has relatively low probability to occur in practical applications with limited bandwidth. Therefore, the proposed *cycle based FGS coding structure* has nearly the same encoding and decoding complexity as anchor method of *decoder oriented two-loop structure*, which is multi-loop encoder and two-loop decoder.

4. EXPERIMENTS AND ANALYSIS

4.1 Experimental conditions

Experiments are performed based on the new release SVC reference software JSVM_5_10. Two different FGS coders are *decoder oriented two-loop structure* as anchor and *cycle based two-loop structure* as proposed (integrated into JSVM_5_10). In order to verify the coding efficiency of the proposed approach, simulations are carried out on various video sequences, which include CIF sequences: foreman, bus, mobile and football at 15Hz, and 4CIF sequences: city, harbour, crew, soccer at 30Hz [7]. And the test conditions are set referring to the requirements in [8] as following: the first frame is encoded as intra frame, and all the remaining frames are encoded as inter-P pictures; two FGS layers are appended on top of base layer (although the experimental results are tested for two FGS layers coding, the proposed approach is also suitable for the case with three FGS layers on top of base layer); initial QP for base layer is 38, and QP step for FGS layers is 6; AVC interpolation filter is used for MC in FGS layers. Particularly, all the leaky factors for anchor are strictly selected to achieve the best coding performance as specified in [8].

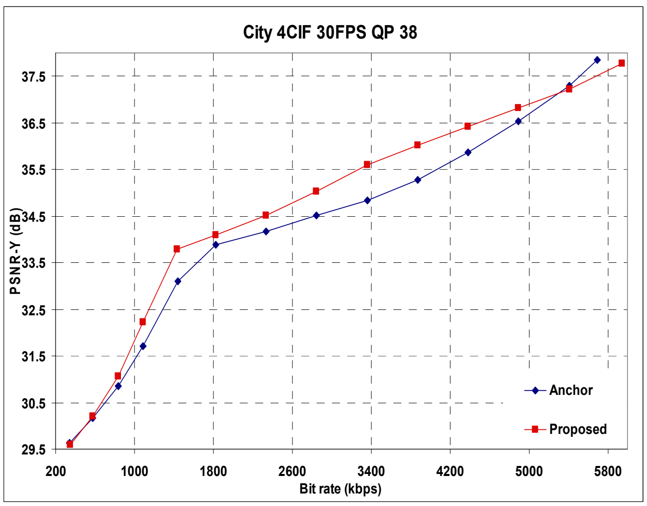
4.2 Analysis of experimental results

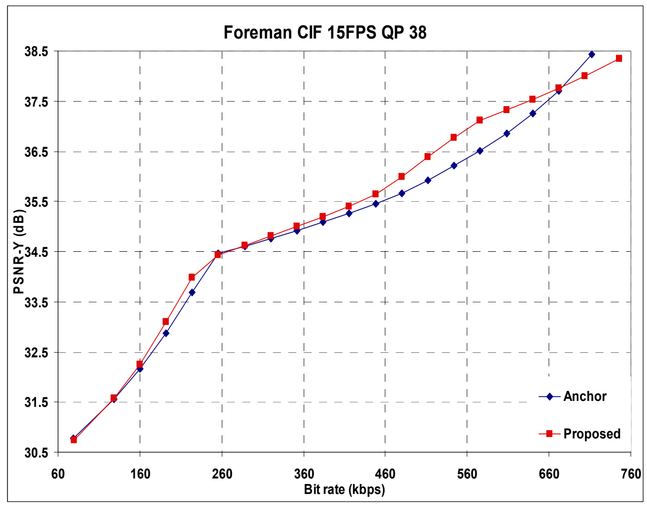
According to the experimental results shown in Fig.8, the proposed approach can significantly outperform anchor method at the second FGS layer (up to 0.8 dB); attributed to the utilization of partial-reconstructed references which can help greatly decrease the prediction drift and thus improve the coding performance around the middle bitrate range of the second FGS layer. Also the incorporated leaky prediction technique with factor γ plays a role to further enhance the compression efficiency of the higher bitrate points. Besides, thanks to the additional loop filter that can make the first

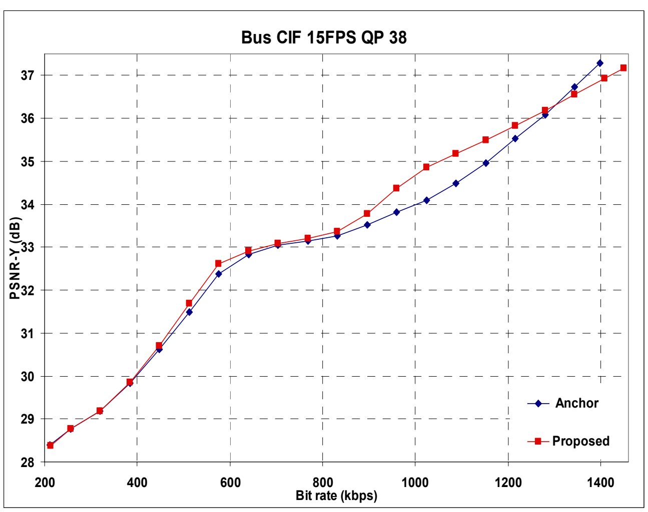
FGS layer references smoother, better coding results are achieved for the first FGS layer. In all, compared to anchor method, the proposed approach can greatly enhance coding quality nearly over wide bitrate range, as well as offering better error control with similar coding complexity.

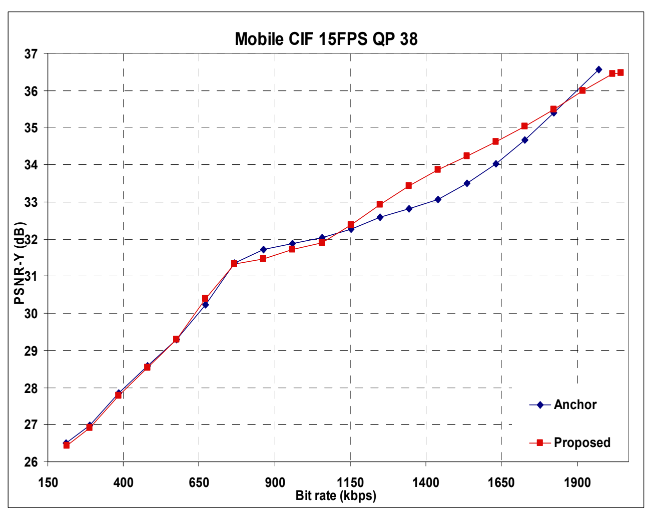
Furthermore, for the experimental results, we may arrive at a conclusion: compared to the proposed approach, the higher compression efficiency near the highest bitrate point for anchor method, to some extent, is achieved at the cost of lower coding performance at the middle bitrate points whose results show concave in R-D curves. On the contrary, the proposed structure can achieve much higher coding quality at the middle bitrate points whose results show convex in R-D curves, at the expense of slight loss around the high end of the second FGS layer, due to the relatively lower quality of prediction signal.

However, for real applications in the Internet or wireless environment, the limited bandwidth of changing channel generally can not always guarantee the overall transmission of large multimedia streaming. It means that, although better coding quality around the highest bitrate point can be obtained by using anchor method, this gain has high probability to be covered up or mitigated once the streaming is truncated. In practical applications, it should be more appreciated to get higher coding performance at the middle bitrate range at which the streaming is more often applied, and this is another advantage of the proposed structure over the anchor method.









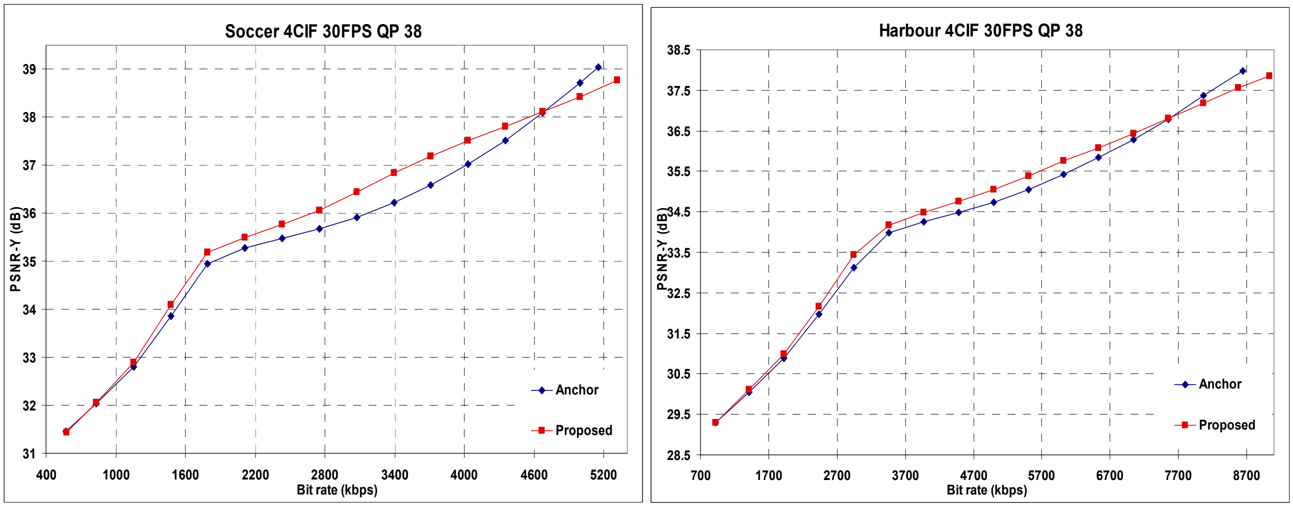


Fig.8. R-D curves of anchors and proposed coding structure.

5. CONCLUSIONS ANF FUTURE WORK

In this paper, we have presented an efficient multiple FGS layers coding structure, which can provide higher coding performance over wide bitrate range with similar coding complexity of multi-loop encoder and two-loop decoder, compared to the exiting method in SVC. Moreover, the proposed approach can not only better control error drifting, but also support more flexible coding quality according to the changing bandwidth based on the available cycle-based reconstruction mechanism, which is of great use for practical low-delay applications in wireless or Internet where large multimedia streaming usually occurs to truncation. An extension of this work may be incorporating the partial reconstruction concept into each FGS layer (including the first FGS layer) and optimizing the leaky predictions, in which way the error resilience can be further strengthen and the coding performance can be enhanced over all the bitrate points.

REFERENCES

1. W. Li, "Overview of fine granularity scalability in MPEG-4 video standard," IEEE Trans. CSVT, vol. 11, pp. 301-317, Mar.2001.
2. T. Wiegand, G. Sullivan, J. Reichel, H. Schwarz and M. Wien, "Joint Scalable Video Model JSVM-6: Joint Draft 6 with proposed changes," Joint Video Team, Doc. JVT-S202, Geneva, Switzerland, April, 2006.
3. J. Ridge, Y. Bao, M. Karczewicz, X. Wang, "Cyclical block coding for FGS," ISO/IEC JTC1/WG11, Doc. m11509, Hong Kong, CN, Jan. 2005.
4. Y. Bao and M. Karczewicz, "CE7 Report: FGS coding for low-delay applications," Joint Video Team, Doc. JVT-O039, Nice, France, Oct. 2005.
5. Y. Bao and M. Karczewicz, "CE7 Report, Multiple FGS layer coding for low-delay applications," Joint Video Team, Doc. JVT-R077, Bangkok, Thailand, Jan. 2006.
6. H. Schwarz, D. Marpe, and T. Wiegand, "Analysis of Hierarchical B Pictures and MCTF," Proc. ICME 2006, Toronto, Canada, July, 2006.
7. Y. Zheng and X. Ji, "Improvements to multiple FGS layers coding for low-delay applications," Joint Video Team, Doc. JVT-T075, Klagenfurt, Austria, July, 2006.
8. S. Jeong, S. Kamp, X. Wang and X. Ji, "CE5 : Improvement AR PR Slices," Joint Video Team, Doc. JVT-T305r1, Klagenfurt, Austria, July, 2006.