Fast Prediction Algorithm of Adaptive GOP Structure for SVC

Yi-Hau Chen, Chia-Hua Lin, Ching-Yeh Chen, and Liang-Gee Chen

Graduate Institute of Electronics Engineering and Department of Electrical Engineering,
National Taiwan University, Taipei, Taiwan
No. 1, Sec. 4, Roosevelt Road, Taipei 106, Taiwan

ABSTRACT

Adaptive group-of-picture (GOP) structure is an important encoding tool in multi-level motion-compensated temporal filtering coding scheme. Compared to conventional fixed-GOP scheme, it can dynamically adapt the GOP size to enhance the coding performance based on each sequence's characteristics. But the existing adaptive GOP structure (AGS) algorithm proposed in JSVM requires huge computation complexity. In this paper, a fast AGS prediction algorithm is proposed. At first, based on the relationship among coding performance, GOP size and corresponding intra block ratio, a sub-GOP size prediction model for different decomposition levels is developed based on the encoded intra block ratio. Then, a prediction scheme is proposed to implement AGS by the sub-GOP size prediction model. It can predict the following sub-GOP size by current sub-GOP's information instead of searching all possible sub-GOP composition. The experimental results show that the proposed algorithm with linear threshold has almost equivalent coding performance as AGS in JSVM but only one-fourth computation complexity for 4-level interframe coding scheme is required.

Keywords: adaptive GOP structure, SVC, JSVM, MCTF, Hierarchical B-picture, video coding.

1. INTRODUCTION

In recent years, interframe wavelet coding becomes a good alternative for scalable video coding, which concept is to perform wavelet transform in the temporal direction. But the coding performance is unacceptable without motion compensation (MC) until Ohm introduces block-based interframe scheme using Haar filter [1]. Then the lifting-based wavelet interframe scheme with the long tap filters, such as 5/3 filter, is proposed [2] to make the coding performances comparable to existing motion compensation prediction video standards, like MPEG-4 and H.264/AVC. For more details, please refer to [3] and [4]. Currently, the scalable extension of H.264/AVC with MCTF is adopted as Joint Scalable Video Model (JSVM), which is the verification model of the scalable video coding (SVC) standard developed by MPEG.

The JSVM adopts three interframe structure, like 5/3 MCTF, 1/3 MCTF, and Hierarchical B-picture (HB). Similar to conventional DWT filter, multi-level decomposition is supported in these interframe schemes to enhance coding efficiency. The group-of-picture (GOP) size is 2^N, where N is the number of decomposition levels. For example, a 4-level HB scheme contains 16 frames in each GOP as shown in Fig. 1. In the earlier work [5], SVC encodes the whole video sequence with one dedicated GOP size as 16 or 32 frames. In this scheme, if the encoded sequence has fast motion, the coding efficiency may be degraded in higher decomposition levels frames. In these frames, the distance between current and reference frames becomes too far in the time axis to predict well in the temporal direction. Hence, the coding efficiency is degraded.

Chen and Woods first introduce the concept of adaptive GOP size in [6]. Then, Park et al. construct the adaptive GOP structure (AGS) to improve the coding efficiency by changing the fixed-GOP MCTF structure to suitable adaptive sub-GOP sizes [7]. It can increase the coding gain about 0.62dB in comparison with JSVM 1.0 [8]. However, the decision algorithm of AGS is to compare the mean-square-error (MSE) of each possible sub-GOP size. That is, all sub-GOP frames should be encoded first, and it introduces huge computation complexity compared to the fixed-GOP structure. In general, for a 16-frame GOP, the computation complexity of AGS is 4 times than the fixed-GOP structure.

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, 65080U, © 2007 SPIE-IS&T · 0277-786X/07/$18

This paper proposes a fast AGS prediction scheme to reduce the huge computation complexity while maintaining the coding efficiency. Since the motivation of AGS in [7] is to utilize the temporal properties of video sequence, we first investigate the information which could be deeply influenced by the adopted GOP size. Based on the analysis, we propose a sub-GOP size prediction model by the encoded intra block ratio and its decomposition level. Furthermore, a fast prediction scheme for most suitable sub-GOP size is proposed in cooperation with the sub-GOP size prediction model. To reduce the huge computation complexity, the proposed scheme refers to the temporal characteristics of previous encoded sub-GOP instead of encoding all possible sub-GOP composition as in [7] and [8]. The experimental result shows that the proposed scheme can appropriately predict the sub-GOP composition and provide coding efficiency as good as AGS in JSVM. The comparison result of computation complexity, which is based on motion estimation (ME) times, also shows that the proposed algorithm's complexity is the same as the original fixed-GOP structure.

This paper is organized as follows. In Section II, the AGS is introduced and then the property of AGS is discussed. Then, the sub-GOP size prediction model is proposed in Section III. In Section IV, the fast AGS prediction scheme is proposed, and the experimental results are shown in Section V. Section VI concludes this paper.

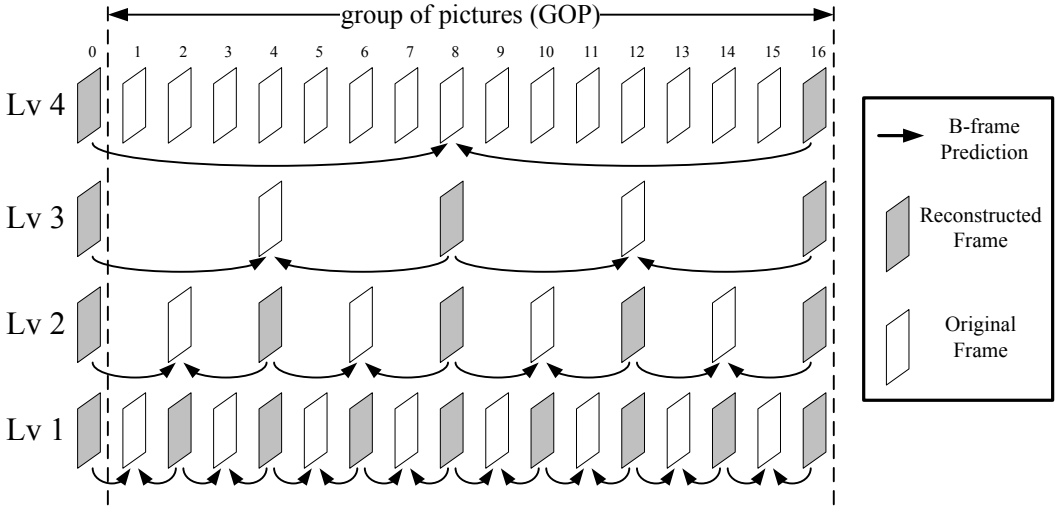


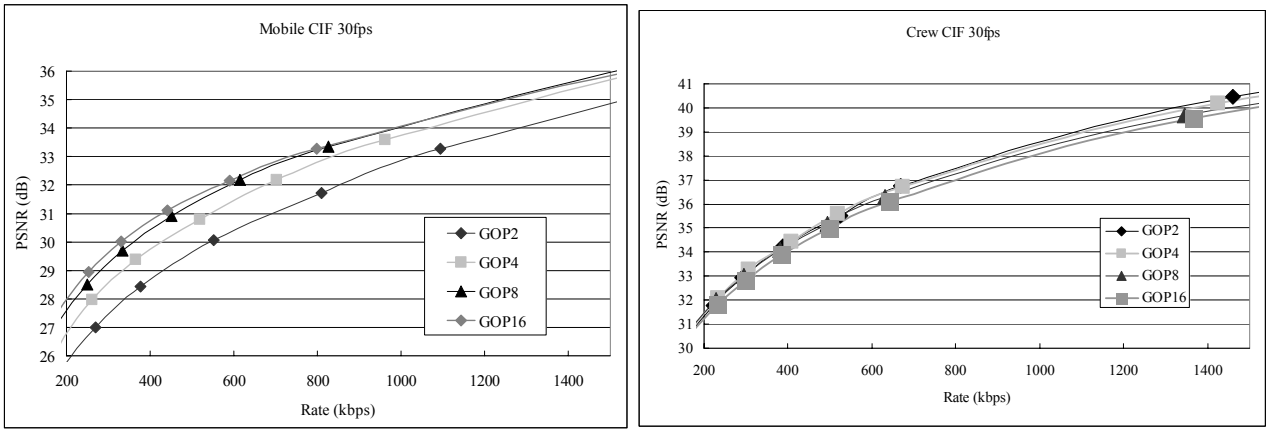
Fig. 1 The Coding structure of 4-level Hierarchical B-picture. There are 16 frames in this GOP.

2. ADAPTIVE GOP STRUCTURE IN JSVM

As described in Sec. 1, the coding efficiency of fixed-GOP structure may be degraded in the higher decomposition levels' frames while the encoded sequence is of fast motion. That is, the coding efficiency will be influenced by the applied GOP size. For example, Figure 2 shows the coding performances of the sequence "Mobile" and "Crew" under different fixed GOP sizes, respectively, and the Hierarchical B-picture in JSVM 3.0 is deployed. In Fig. 2(a), since the sequence "Mobile" is of small object motion and stable background content, a larger GOP size can leads to better coding efficiency. However, the sequence "Crew" has a quite different coding result. Because the sequence's content changes frequently, the temporal prediction coding efficiency of higher decomposition level's frame is not good enough. The above phenomenon shows that the coding efficiency is not proportional to the adopted GOP size. Moreover, the video characteristic in one sequence may be inconsistent among all frames so that the applied GOP size should be adapted according to current video's property.

The AGS adopted in JSVM 2.0 [9] regulates its possible sub-GOP structure and describe the sub-GOP mode decision procedure. Since multi-level interframe schemes, such as 5/3 MCTF, 1/3 MCTF, and HB, are all dyadic structures, all the sub-GOP size should be power of two frames. When the original fixed-GOP size is 16 frames, the AGS can achieve the best coding efficiency by using one of the applicable AGS as shown in Fig. 3.

Figure 4 depicts an algorithm of the selection mechanism for the AGS adopted as JSVM 2.0. Take the 16-frame GOP as example, the encoding procedure is performed by varying the sub-GOP size as 16-frame, 8-frame, 4-frame, and 2-frame. The MSE (Mean Square Error) of each sub-GOP will be calculated in one full-sized GOP structure. The "Sub-GOP mode decision" procedure selects the sub-GOP size with minimum MSE from 4 possible sub-GOP sizes to construct the GOP. For more detailed mode decision procedure, please refer to [9]. Then the GOP is re-encoded by the selected sub-GOP mode to generate the bitstream. This procedure is repeated for each full-sized GOP section of the video sequence. Since the pre-encoding procedures are required by AGS in JSVM 2.0, the computation complexity of AGS can be about 4 times that of original fixed-GOP scheme when the GOP is 16 frames. Such huge computation is the most serious problem to realize AGS. In this paper, our goal is to provide a prediction algorithm so that the computation complexity of pre-encoding procedures can be reduced.



(a) (b)

Fig. 2. The coding performances of different GOP sizes for the sequences Mobile (a) and Crew (b). The Hierarchical B-picture scheme in JSVM 3.0 is applied.

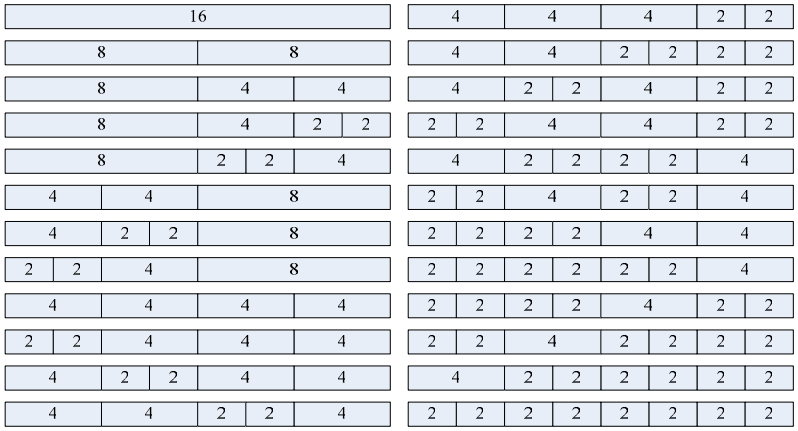


Fig. 3 All the applicable sub-GOP mode in an adaptive GOP structure when the original fixed-GOP size is 16 frames.

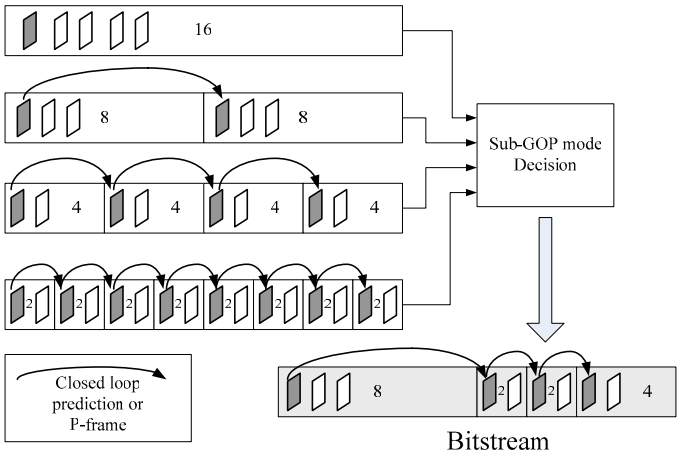
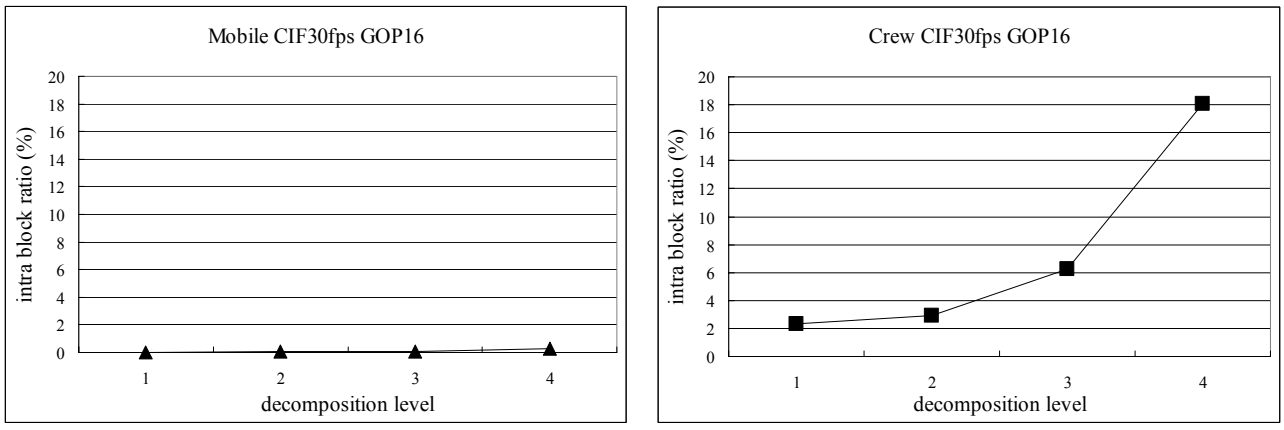


Fig. 4 The flowchart of AGS algorithm in JSVM on GOP size of 16 frames where the dark frames refer P-frames.

3. PROPOSED SUB-GOP SIZE PREDICTION MODEL

The target of our proposed fast AGS algorithm is to avoid repeatedly encoding via predicting the size of the next sub-GOP. The information applied to predict should actually reflect the coding efficiency of the prediction algorithm. In general, a sequence with larger motion is not suitable for a large GOP size because the limitation of search range induces inefficient motion vectors and higher bit-rate costs at higher decomposition level which means longer time intervals. But in lower decomposition levels, this sequence can still provide good coding performance since the temporal distance between reference frame and current frame is not far. It reveals that, for the same sequence, the temporal prediction results, such as MSE, inter/intra mode selection, and temporal prediction direction, may be quite different in different decomposition levels. Based on these prediction results, we can determine the most suitable GOP size to encode the sequence.

Among the temporal prediction results, the MSE and inter/intra mode selection can be regarded as the two most relevant information to the coding efficiency. However, it is quite hard to set a general MSE threshold to determine the best GOP size since the complexity of different sequences' texture could be quite different and influence the distribution of MSE. For the inter/intra mode selection result, the mode decision prefers intra block when the inter block costs more bit-rate. It implies that the temporal prediction efficiency becomes lower when more intra blocks are selected in current temporal decomposition level.



(a) (b)

Fig. 5 The intra block numbers of Mobile (a) and Crew (b) of GOP size = 32 under JSVM 3.0 with 5/3 MCTF coding scheme. Higher decomposition level refers to the layer of lower frame rate.

Figure 5 shows the intra block ratio of two sequences, Mobile and Crew. As shown in Fig. 2, these two sequences have different impacts by the number of decomposition levels. At low decomposition level, both sequences have quite low intra block ratio. At higher decomposition level, the sequence "Mobile" still maintains a low intra block ratio and its coding efficiency is improved, as the number of decomposition level increases. On the other hand, the intra block ratio of the sequence "Crew" increases fast and the coding performance is degraded apparently. Besides these two sequences, we can also find the similar characteristics for other sequences. Therefore, the proposed sub-GOP size prediction model adopts intra block ratio as criteria to make decision.

In general, a static threshold for intra block ratio is enough to decide the best sub-GOP size. However, since the frames in higher decomposition levels are more important, it is tolerable to use more intra blocks to enhance the quality of reconstructed frames. Therefore, a linear threshold of intra block ratio for sub-GOP size prediction model is proposed. For lower decomposition levels, a stricter threshold is given; for higher decomposition levels, a looser threshold is given.

4. PROPOSED FAST AGS PREDICTION SCHEME

As describe in Sec. 2, the AGS in JSVM decides the sub-GOP mode in aid of complex pre-encoding procedures. Our proposed scheme is to predict the next sub-GOP size to maintain the same computation complexity as the fixed-GOP structure. The proposed fast AGS prediction algorithm applies above sub-GOP size prediction model to predict next sub-GOP size. The concept of proposed scheme can be shown as Fig. 6. The "sub-GOP Predictor" predicts the next sub-GOP size based on the information of previous sub-GOP. Besides, for the sub-GOPs in the same GOP, they will follow the dyadic structure to constraint the choices of next sub-GOP size.

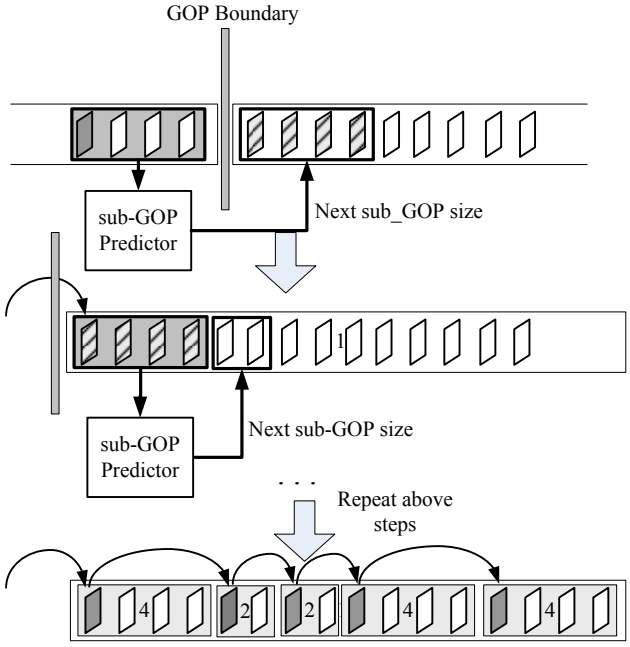


Fig. 6. The coding order of proposed fast AGS prediction scheme. Each frame is only encoded once.

The detailed steps of our proposed prediction scheme are as follows:

1. Set a fixed maximum GOP size (power of 2) and the threshold of intra block ratio for sub-GOP predictor.
2. Load the maximum GOP size frames into input buffer.
3. Encode the frames of the predicted sub-GOP size from the input buffer.
4. Calculate the average intra block ratio in each layer and find the most suitable sub-GOP size. The decision method can be represented as:

$$\begin{cases} R_n < I_n, & SGOP_{next} = 2^{n+1} \\ R_n < I_n \text{ and } R_{n-1} < I_{n-1} & SGOP_{next} = 2^n \\ R_{n-1} < I_{n-1} \text{ and } R_{n-2} < I_{n-2} & SGOP_{next} = 2^{n-1} \\ \vdots & \end{cases},$$

where n is the number of decomposition level, R_n is the average intra block ratio in level n, I_n is the threshold in level n, $SGOP_{next}$ is the predicted next sub-GOP size. Besides, in order to support the dyadic structure in AGS, some constraints are applied to refine the predicted results.

5. Repeat the step 3 to 5 while there are un-encoded frames in the input buffer.
6. If there are no un-encoded frames in input buffer, return to step 2.

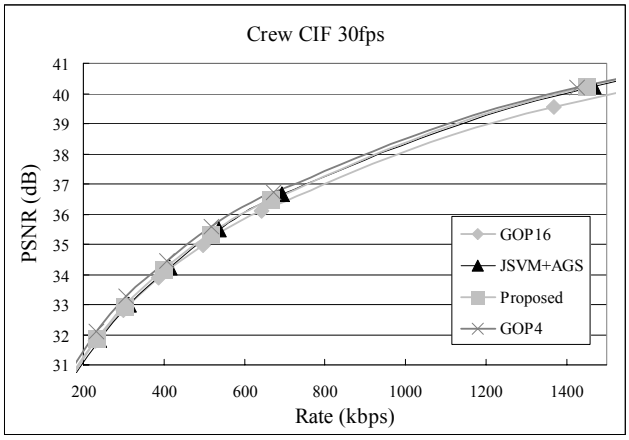
5. EXPERIMENTAL RESULTS

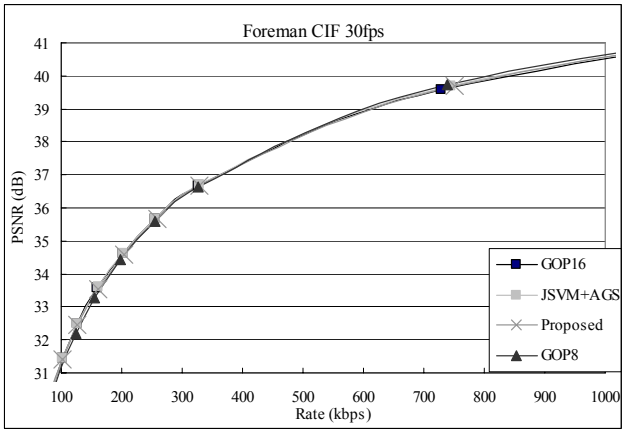
The settings of simulation environment are eight MPEG standard CIF 30 fps sequences with 289 frames. 4-level Hierarchical B-picture structure in JSVM 3.0 is adopted. The searching range is [-96, 96) for all decomposition levels and the fast search algorithm in JSVM is applied. The Lagrangian mode decision is applied for rate-distortion optimization. Based on our statistic results, we set the linear threshold of intra block ratio from 1% to 4% for all test sequences. Figure 7 shows the coding performance comparison of these eight sequences. We compare our proposed fast prediction scheme with two benchmarks: JSVM 3.0 with fixed GOP size of 16 frames [10] and JSVM 3.0 with AGS [8]. Both schemes are simulated by the same coding parameters as described in above.

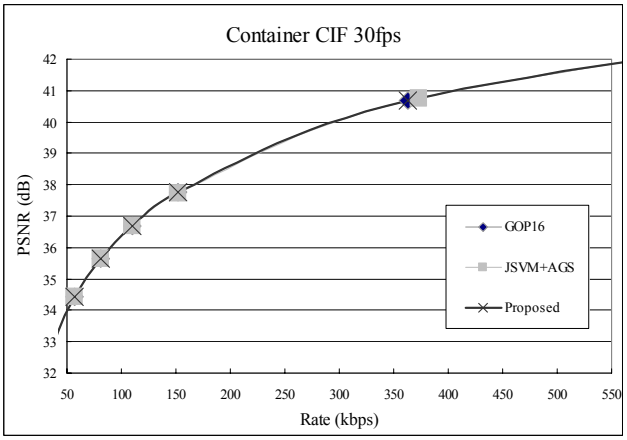
In Figure 7, our proposed fast AGS scheme shows equivalent coding performance to JSVM 3.0 with AGS. For the sequences which prefer larger GOP size, the proposed method can also choose larger sub-GOP mode and provide the same coding performance as the fixed 16-frame GOP. For the sequences with fast motion or dynamic content, such as "Crew", "Stefan", and "Dancer", the proposed method and AGS[8] both have better coding performance than that of fixed 16-frame GOP. Besides, Fig. 7 also lists the coding performance of smaller fixed GOP as another benchmark. And the proposed method can match up with the results of these smaller fixed GOP except the sequence "Dancer". The degradation between "GOP2" and our method could be derived from the sequence's very complex texture. It makes the coding gain of inter-coded block become overestimated and misleads our method to choose larger sub-GOP in some cases.

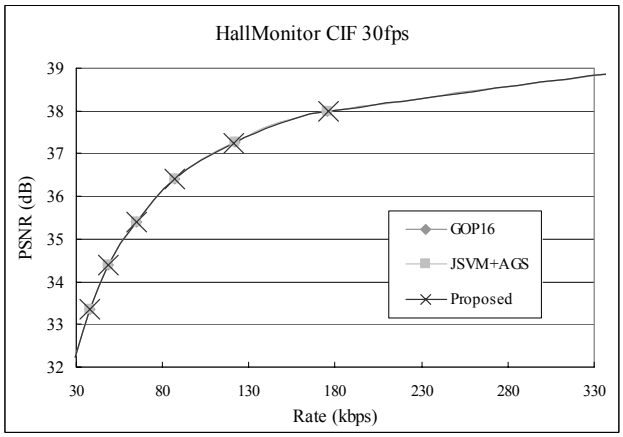
Table 1 summarizes the difference of coding performance between the proposed method and the two benchmarks from Fig. 7. For the computation complexity, since the motion estimation occupies most computation time in JSVM, we evaluate the computation complexity by the number of motion estimation. Note that we set the complexity of P-frame and B-frame as 1 and 2, respectively. Therefore, the smaller GOP structure has a smaller computation complexity than that of the larger GOP structure. From Table 1, our proposed method shows almost the same coding efficiency as the AGS while its computation complexity is only about one-fourth of AGS. Compared to fixed GOP structure, the proposed method can appropriately reduce the complexity by using smaller GOP while providing equivalent or better coding efficiency.

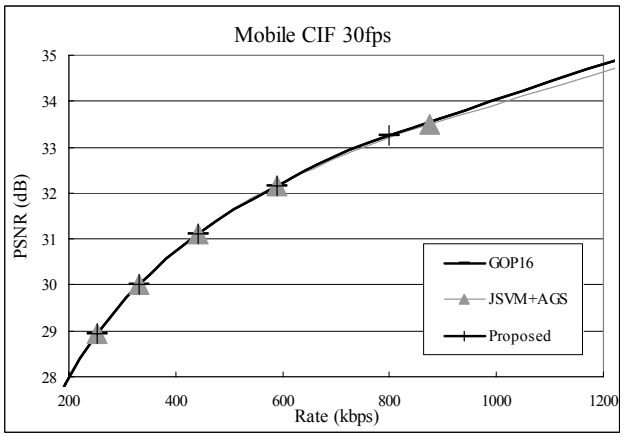
For the sequence of higher resolution, the proposed method still works well. As shown in Fig. 8, the proposed sub-GOP prediction model still helps the AGS scheme to select more efficient sub-GOP mode. Compared to fixed 16-frame GOP, it provides about 0.5dB coding gain at 2Mbps.

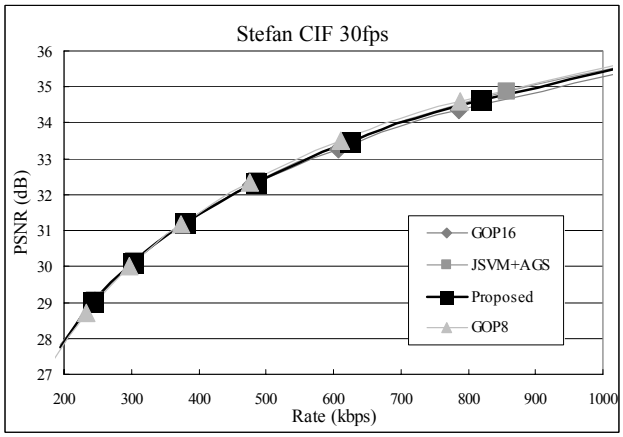

Crew CIF 30fps
PSNR (dB)
Rate (kbps)
GOP16
JSVM+AGS
Proposed
GOP4


Foreman CIF 30fps
PSNR (dB)
Rate (kbps)
GOP16
JSVM+AGS
Proposed
GOP8


Container CIF 30fps
PSNR (dB)
Rate (kbps)
GOP16
JSVM+AGS
Proposed


HallMonitor CIF 30fps
PSNR (dB)
Rate (kbps)
GOP16
JSVM+AGS
Proposed


Mobile CIF 30fps
PSNR (dB)
Rate (kbps)
GOP16
JSVM+AGS
Proposed


Stefan CIF 30fps
PSNR (dB)
Rate (kbps)
GOP16
JSVM+AGS
Proposed
GOP8

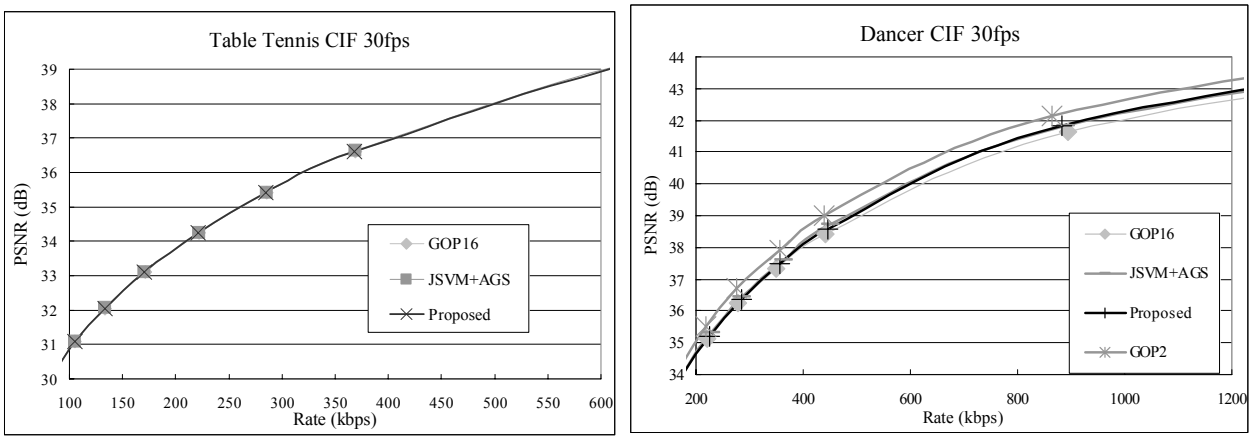


Fig. 7 Coding performance comparison between JSVM with fixed GOP size, JSVM with AGS, and the proposed fast prediction scheme. All schemes are simulated by JSVM 3.0 with Hierarchical B-picture.

Table 1. Summary of the coding performance and computation complexity comparison between JSVM, AGS, and proposed scheme for 8 CIF sequences. The maximum GOP size is 16 frames. The value of 9 Qp are 18, 24, 30, 32, 34, 36, 38, 40, and 42.

	PSNR(dB)			Proposed scheme PSNR Improvement(dB)			Computation complexity (Average on 9 Qp)(%)	
Sequences	JSVM	AGS	Proposed	Compare to JSVM	Compare to AGS	@Bitrate (kbps)	Compare to JSVM	Compare to AGS
Crew	36.90	37.15	37.15	0.25	0.00	784.00	89.53	24.56
Foreman	38.30	38.35	38.30	0.00	-0.05	512.00	97.55	26.76
Container	37.25	37.25	37.25	0.00	0.00	128.00	99.98	27.43
HallMonitor	38.08	38.08	38.08	0.00	0.00	192.00	98.77	27.09
Mobile	33.20	33.15	33.20	0.00	0.05	784.00	100.00	27.43
Stefan	34.34	34.53	34.48	0.14	-0.05	784.00	95.78	26.28
TableTennis	38.13	38.14	38.14	0.01	0.00	512.00	99.04	27.17
Dancer	41.10	41.34	41.34	0.24	0.00	784.00	98.67	27.07

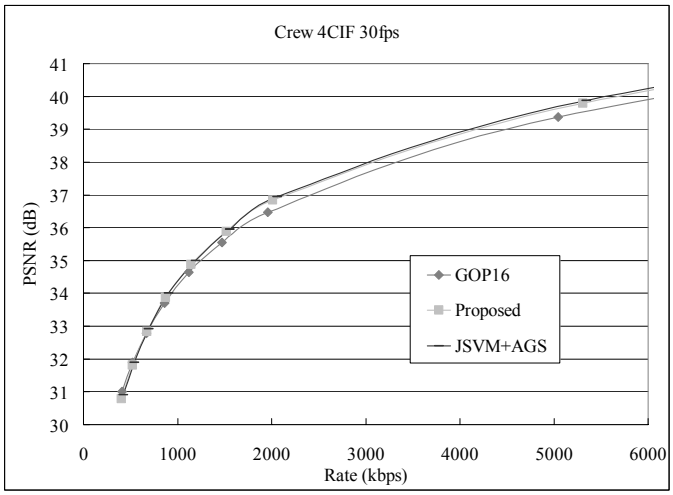


Fig. 8 Coding performance comparison of the sequence “Crew” with 4CIF 30fps.

6. CONCLUSION

In this paper, we propose a fast AGS prediction algorithm for SVC and other interframe coding schemes. A sub-GOP size prediction model is proposed based on the linear threshold of intra block ratio. And a fast prediction is proposed to avoid the complex pre-encoding procedures in JSVM. For most sequences, the coding performance of proposed scheme is the same as original AGS while it can reduce the computation complexity to about 26% compared to current AGS scheme. In our proposed model, only one spatial layer is considered to predict sub-GOP size. In our future work, we will extend the proposed prediction scheme to multi-spatial layers by taking the intra mode in inter-layer prediction into consideration.

REFERENCES

[1] J.-R. Ohm, "Three Dimensional Subband Coding with Motion Compensation," in IEEE Transaction on Image Processing, vol. 3, no. 5, pp. 559-571, Sept. 1994.

[2] A. Secker and D. Taubman, "Motion-compensated highly scalable video compression using an adaptive 3D wavelet transform based on lifting," in Proc. IEEE International Conference on Image Processing, 2001, pp. 1029–1032.

[3] D. Taubman, "Successive refinement of video: fundamental issues, past efforts and new directions, " in International Symposium on Visual Communications and Image Processing, 2003, pp. 791– 805.

[4] J.-R. Ohm, "Advances in scalable video coding," Proc. of IEEE, pp. 42–56, Jan. 2005.

[5] ISO/IEC JTC 1, "Scalable Video Model 3.0," ISO/IEC JTC1/SC29/WG11 Doc. N6716, Oct, 2004.

[6] P. Chen, and J. W. Woods, "Bidirectional MC-EZBC with Lifting Implementation," IEEE Trans. on Circuits and Systems for Video Technology, vol. 14, no. 10, pp. 1183-1194, Oct. 2004.

[7] Park G. H., Park M. W., Jeong S., Cha J., Kim K, and Hong J. "Adaptive GOP Structure for SVC," ISO/IEC JTC1/SC29/WG11 MPEG2005/M11563, Jan., 2005.

[8] Park G. H., Park M. W., Jeong S., Cha J., Kim K, and Hong J. "Improve SVC Coding Efficiency by Adaptive GOP Structure," ISO/IEC JTC1/SC29/WG11 JVT-O018, Apr., 2005.

[9] ISO/IEC JTC 1, "Joint Scalable Video Model (JSVM) 2.0 Reference Encoding Algorithm Description ," ISO/IEC JTC1/SC29/WG11 Doc. N7084, Apr, 2005.

[10] ISO/IEC JTC 1, "Joint Scalable Video Model 3.0," ISO/IEC JTC1/SC29/WG11 Doc. N7796, Jan, 2006.