FAST MODE DECISION FOR COARSE GRAIN SNR SCALABLE VIDEO CODING

He Li[1], Z. G. Li[2], Changyun Wen[1,]*

[1]School of EEE, Nanyang Technological University,
Singapore 639798
[2]Media Division, Institute for Infocomm Research,
Singapore 119613

ABSTRACT

Scalable Video Coding (SVC) is an on-going standard and the current working draft (WD) is an extension of H.264/AVC. In the WD, exhaustive search technique is employed to select the best coding mode for each macroblock (MB). This technique achieves highest possible coding efficiency, but it results in higher computational complexity. To overcome this, we propose a novel fast mode decision scheme for coarse grain SNR scalability (CGS) in SVC. In this scheme, the mode distribution relationships between the base layer and enhancement layers are employed to reduce the candidate mode set at enhancement layers. The experimental results show that the proposed scheme provides significant reduction in computational complexity with negligible coding loss.

1. INTRODUCTION

Scalable video coding as proposed in [1] is an extension of H.264/MPEG-4 Advanced Video Coding (H.264/AVC). The basic design idea of the scalable H.264/AVC extension is to extend the hybrid video coding approach of H.264/AVC towards motion-compensated temporal filtering (MCTF). This open loop structure of a temporal subband representation offers the possibility to efficiently incorporate SNR, spatial and temporal scalability [2]. CGS scalability provides video at different quality levels. It contains a base layer and several enhancement layers, which are coded at the same spatial and temporal resolution.

Current SVC scheme shows significant achievements in terms of coding efficiency, robustness to a variety of network channels over contemporary video coding standards. In this coding system, block matching motion estimation is used to reduce the temporal redundancy between frames. The available MB modes in SVC include two intra modes such as INTRA_4×4 and INTRA_16×16, MODE_SKIP and seven inter modes [2]. For enhancement layers in CGS, another mode BL_pred is added to indicate that motion and prediction information including the partitioning of the corresponding MB of the base layer is used [3]. Similar to H.264/AVC, the motion estimation and mode decision process in SVC is performed by minimizing of the Lagrangian formulation of function J, where J is given by:

$$J(MODE \mid QP, \lambda_{SSD}) = D(MODE \mid QP) + \lambda_{SSD} R(MOED \mid QP)$$

where D is the distortion between the original MB and the reconstructed MB located in the reference frames. R denotes the bit cost for encoding the motion vectors, MB header and all the residual information. For each possible MB partition, the prediction method together with the associated reference indices r_0 and r_1 and motion vectors $\{mv_0\}$ and/or $\{mv_1\}$ is determined by

$$mv_{0/1}(r_{0/1}) = \arg\min_{m_{0/1}} \{D_{SAD} + \lambda_{SAD}(R(r_{0/1}) + R(mv_{0/1}))\}$$

Based on the given quantization parameter (QP), the Lagrangian multiplier. λ_{SSD} and λ_{SAD} can be derived by

$$\lambda_{SSD} = 0.85 \times 2^{QP/3-4}, \lambda_{SAD} = 0.92 \times 2^{QP/6-2}$$

Theoretically, large values of λ_{SSD} and λ_{SAD} work well at a low bit rate range while small values of λ_{SSD} and λ_{SAD} work well at a high bit rate range. The partition at high bit rate is thus finer than that at low bit rate. During mode decision process, all modes are examined using the Lagrangian function. As a result, SVC achieves optimal coding performance at the expense of tremendously increased computational complexity.

Motivated by the observations that correlation exists between the mode distribution at the base layer and that at enhancement layers, we propose an effective fast mode decision for CGS scalable coding at enhancement layers. Simulation results illustrate that our algorithm can achieve up to nearly 50% of encoding time saving with negligible PSNR loss and bit rate increase for all the layers.

The rest of this paper is organized as follows. In Section 2, statistical characteristics of MB modes distribution among SNR layers are studied. In Section 3, we propose a fast mode decision algorithm for MBs at the enhancement layer. Simulation results are presented in Section 4. Finally, Section 5 concludes the paper.

* Corresponding Author.

2. CGS SCALABILITY IN SVC

CGS in SVC is achieved by encoding successive quality layers. At first, the texture information is encoded in an AVC compatible base layer to provide a minimum quality at a given quantization level. At enhancement layers, CGS is achieved by decreasing the quantization step size and encoding successive refinements of the transform coefficients. For motion information, MCTF is applied in each layer independently and a large degree of inter-layer prediction is incorporated [5][6]. Intra and inter MBs can be predicted using the corresponding signals of previous layers. Moreover, the motion description of each layer can be used for a prediction of the motion description for following enhancement layers.

MCTF plays an essential role in SVC. In MCTF, a pair of temporal low and high frequency frames are generated for each two consecutive input frames by Haar or 5/3 filtering. An example for four-stage MCTF structure is depicted in Figure 1 by using 5/3 filtering. After MCTF, the low-pass frame at the highest temporal level together with all the high-pass frames are entropy coded.

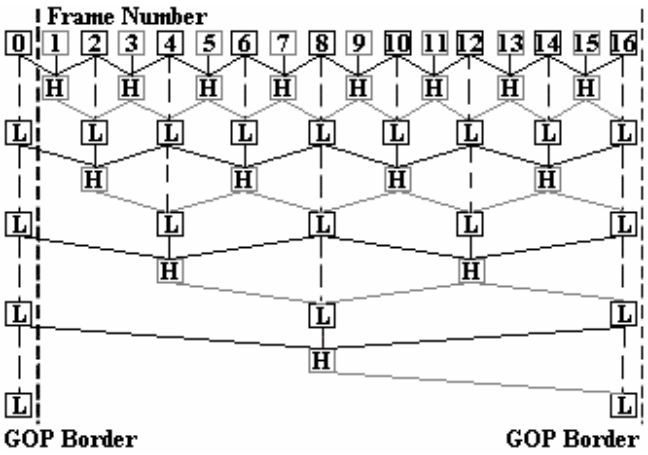

Figure 1 An Example for Four-stage MCTF

During our exhaustive experiments on various video sequences, the following relationships can be found:

- If the MB mode at the base layer is INTRA_4×4 or INTRA_16×16, it is probable that the corresponding MB at its enhancement layer is intra coded. The intra coded MB at the base layer indicates that one cannot find a good match in the reference frames. Similarly, at enhancement layers, it is also difficult for the MB to find a good match in the reference frames.
- For CGS, enhancement layers refine the transform coefficients generated at the base layer. Therefore, the partition at enhancement layers for each MB should be finer than that at the base layer.
- After MCTF, the high-pass frames at the higher temporal level have more texture and motion information compared with that at lower temporal level. Take Figure 1 as an example, the high-pass frames at position 8, 4 and 12 have more information than that at position 2, 6, 10, 14, 1, 3, 5, 7, 9, 11, 13 and 15. Therefore, it is probable that the MBs in high-pass frame at position 8, 4 and 12 are intra coded, especially for the video sequence with high motion and fine detail.

3. PROPOSED FAST MODE DECISION SCHEME

Based on the considerations above, three techniques are used in our fast mode decision algorithm.

3.1 Selective intra coding

This proposed algorithm is used to distinguish the MBs encoded with INTRA mode in B frames at enhancement layers. If the MB mode at the base layer is INTRA_4×4 or INTRA_16×16, then the candidate mode set for enhancement layer is reduced to BL_pred and INTRA_4×4. Since enhancement layers have refined information of the base layer, we use INTRA_4×4 to encode the MBs instead of INTRA_16×16.

After MCTF, high-pass frames at high temporal level have more motion and texture information than that at low temporal level. As a result, the MBs in these high-pass frames have more chance to be intra coded. In order to decide an MB mode in the frames that at high temporal level, we divide the modes into two classes: Class_inter and Class_intra. We regard MODE_8×8 and INTRA_4×4 as the representative block sizes of these two classes. The rate distortion cost for MODE_8×8 (RDcost8) and INTRA_4×4 (RDcost4) are estimated. If RDcost4 is less then RDcost8, we assume that probability of intra coding is very high. Then the best mode is INTRA_4×4. On the other hand, if RDcost8 is less than RDcost4, we can regard that the best mode would belong to Class_inter. Then the following techniques are used in the mode decision process for the MBs which belong to Class_inter.

3.2 Reduced candidate mode set

Since enhancement layers have refined information of that of the base layer, we can reduce the mode set for certain MBs. If the MB mode at the base layer is MODE_8×8, then the candidate mode set is reduced to BL_pred and MODE_8×8. If the MB mode at the base layer is MODE_16×8 (or MODE_8×16), then the candidate mode set is reduced to BL_pred, MODE_8×8 and MODE_16×8 (or MODE_8×16).

3.3 Selective residual prediction at enhancement layers

This algorithm is used to examine whether the MBs at enhancement layers need residual prediction. For many MBs (such as the MBs at the background), the residual information after quantization is almost zero especially when the quantization parameter value is large at the base layer. In such case, we don't need to do the residual prediction at enhancement layers. For each MB, all the

luminance pixel information at the base layer will be added together, if they are less than certain threshold, only rate distortion optimization without residual prediction is performed at enhancement layers.

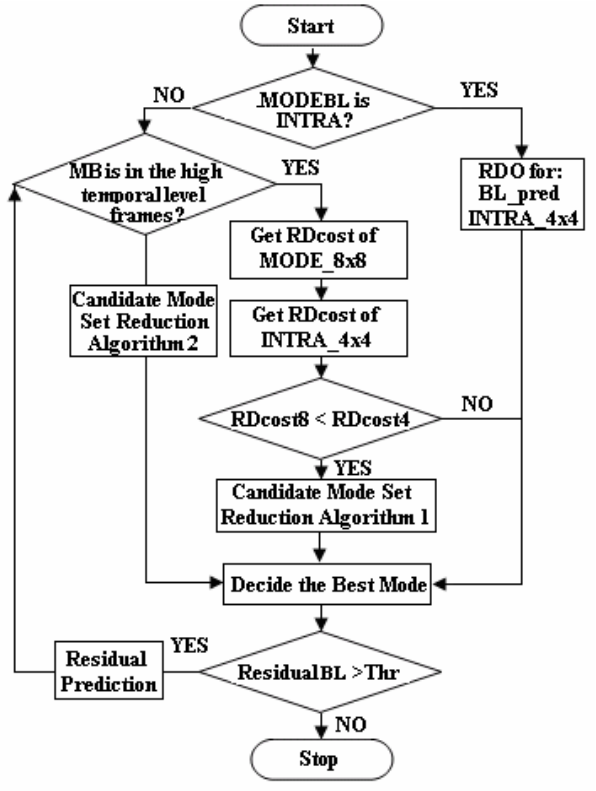

Figure 2 Flowchart of Proposed Fast Mode Decision

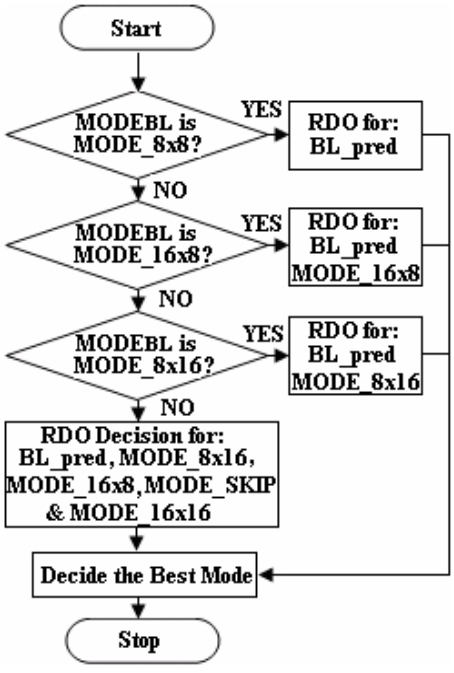

Figure 3 Candidate Mode Set Reduction Algorithm 1

A flowchart of the overall process of the proposed scheme is shown in Figure 2. $MODE_{BL}$ represents the best mode selected at the base layer for certain MB. RDcost8 and RDcost4 stand for the rate distortion cost for MODE_8×8 and INTRA_4×4 respectively. $Residual_{BL}$ represents the summation of quantized luminance texture information at the base layer for certain MB. If it is greater than certain threshold value, then motion estimation with residual prediction at enhancement layers are performed. Figure 3 shows the candidate mode set reduction algorithm 1. For mode set reduction algorithm 2, MODE_8×8 is added to each RDO mode decision process.

4. SIMULATION RESULTS

The proposed fast mode decision scheme is embedded in JSVM 2.0 encoder [2]. The test platform used is Intel Pentium IV, 1.83GHz CPU, 256M RAM with Windows XP professional operating system. The test condition is shown in Table I. In our experiments, four standard test sequences including FOREMAN, FOOTBALL, CITY and HARBOUR have been tested. We only consider the two-layer case and the QP value setting for the base layer and the enhancement layer are shown in Table I. The GOP size is set to be 8. These sequences are all 72 frames long. They represent sequences with large and medium motion.

TABLE I. SIMULATION CONDITION

		All Tested Video Sequences
QP	*Base*	40
	Enhancement	10, 15 and 20
Total Number of Frames		72
Resolution		QCIF for both layers
Coding Option Used		MV search range is ±32 pels RDO is enabled. Reference frame number is 1. MV resolution is 1/4 pel.
Codec		JSVM 2.0 encoder

The testing parameters in our experiments include the average time saving, Y-PSNR and bit rate for both layers. We use Time Saving (TS) to indicate the average time saving in encoding process:

$$TS = \frac{T_{JSVM} - T_{proposed}}{T_{JSVM}} \times 100\%$$

where T_{JSVM} and $T_{proposed}$ are the encoding time of JSVM 2.0 encoder and its modified encoder according to the proposed fast mode decision method, respectively.

Table II to IV show the coding results of JSVM and our proposed scheme with different QP values at the enhancement layer. The results show that the proposed method is very effective in reducing the encoding time, especially for the sequence with high motion and fine detail.

The total encoding time is reduced up to about 49%. Figure 4 to 5 present the rate distortion curves for FOREMAN and FOOTBALL. From these figures, we can conclude that our scheme can achieve consistent time saving over a large bit rate range with negligible loss in PSNR and increments in bit rate.

TABLE II. SIMULATION RESULTS WITH QP VALUE AT ENHANCEMETN LAYER IS 10

Sequence	JSVM Scheme		Proposed Scheme		TS(%)
	Bit rate	*PSNR*	*Bit Rate*	*PSNR*	
Foreman	34.6900	30.1923	34.6900	30.1923	41.98%
	758.580	48.6579	761.427	48.6075	
Football	73.2667	27.8863	73.2667	27.8863	48.95%
	1413.94	48.6198	1428.54	48.6833	
City	30.4117	29.8263	30.4117	29.8263	41.24%
	799.818	48.2633	800.872	48.1715	
Harbour	46.2033	27.5756	46.2033	27.5756	44.07%
	1339.31	47.6632	1343.14	47.6070	

TABLE III. SIMULATION RESULTS WITH QP VALUE AT ENHANCEMENT LAYER IS 15

Sequence	JSVM Scheme		Proposed Scheme		TS(%)
	Bit rate	*PSNR*	*Bit Rate*	*PSNR*	
Foreman	34.6900	30.1923	34.6900	30.1923	41.61%
	452.288	45.5309	454.625	45.4699	
Football	73.2667	27.8863	73.2667	27.8863	48.39%
	981.375	44.8161	993.740	44.9089	
City	30.4117	29.8263	30.4117	29.8263	39.88%
	442.602	44.7726	442.580	44.6587	
Harbour	46.2033	27.5756	46.2033	27.5756	40.43%
	915.288	44.0445	918.713	43.9611	

TABLE IV. SIMULATION RESULT WITH QP VALUE AT ENHANCEMENT LAYER IS 20

Sequence	JSVM Scheme		Proposed Scheme		TS(%)
	Bit rate	*PSNR*	*Bit Rate*	*PSNR*	
Foreman	34.6900	30.1923	34.6900	30.1923	41.27%
	282.540	42.4896	284.172	42.4100	
Football	73.2667	27.8863	73.2667	27.8863	47.48%
	659.422	40.9418	665.060	40.9952	
City	30.4117	29.8263	30.4117	29.8263	38.74%
	251.893	41.6254	252.418	41.5105	
Harbour	46.2033	27.5756	46.2033	27.5756	42.64%
	572.958	40.1346	574.568	40.0383	

5. CONCLUSION

In this paper, a novel fast mode decision algorithm at enhancement layers for CGS scalable video coding is presented. With this scheme, the candidate mode set is reduced at enhancement layers. The results of the simulations in Section 4 demonstrate that the proposed algorithm can save up to 48.95% of the encoding time as compared to the original JSVM 2.0 encoder. Moreover, it introduces insignificant picture degradation and bit rate increase.

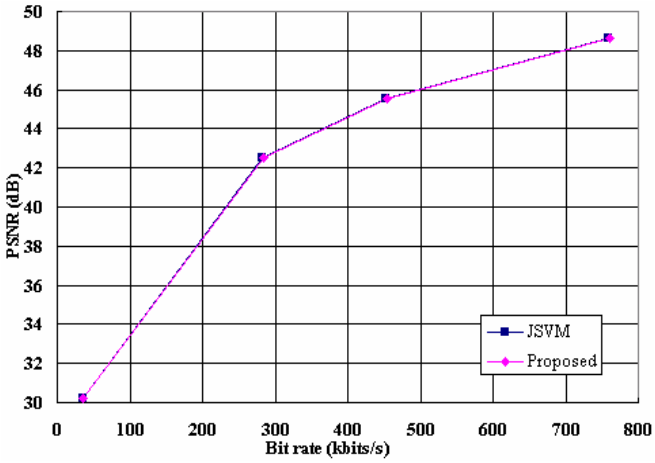

Figure 4 Rate Distortion Curve for FOREMAN

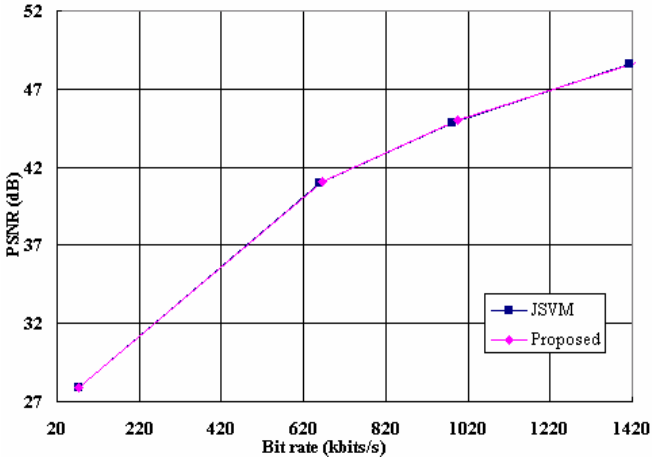

Figure 5 Rate Distortion Curve for FOOTBALL

6. REFERENCES

[1] H.Schwarz, T. Hinz, H. Kirchhoffer, D. Marpe, and T. Wiegand, "Technical Description of the HHI proposal for SVC CE1," ISO/IEC JTC1/WG11, Doc. M11244, Palma de Mallora, Spain, Oct. 2004.

[2] J. Reichel, H. Schwarz, and M. Wien, "Joint Scalable Video Model 2.0 Reference Encoding Algorithm Description," ISO/IEC JTC1/SC29/WG11, Doc. N7084, Buzan, Korea, April. 2005.

[3] J. Reichel, H. Schwarz, and M. Wien, "Scalable Video Coding – Working Draft 1," Joint Video Team (JVT), Doc. JVT0-N020, Hong Kong, CN, Jan. 2005

[4] H. Schwarz, D. Marpe, T. Schierl and T. Wiegand, "Combined Scalability Support for the Scalable Extension of H.264/AVC," IEEE ICME, Amsterdam, Netherlands, July. 2005.

[5] H. Schwarz, D. Marpe and T. Wiegand, "Inter-layer Prediction of Motion and Residual Data," ISO/IEC JTC 1/SC 29/WG 11/M11043, USA, July. 2004.

[6] Z. G. Li, X. K. Yang, K. P. Lim, X. Lin, S. Rahardja, and F. Pan, "Scalable Video Coding with Grid Motion Estimation and Compensation," US Provisional Patent Application (I2R Ref: P2004003-ETPL Ref: I2R/P//1899/PCT), Filed in 23 June, 2004.