LAYER-ADAPTIVE MODE DECISION AND MOTION SEARCH FOR SCALABLE VIDEO CODING WITH COMBINED COARSE GRANULAR SCALABILITY (CGS) AND TEMPORAL SCALABILITY

[1]Hung-Chih Lin, [2]Wen-Hsiao Peng, [1]Hsueh-Ming Hang and [3]Wen-Jen Ho*

[1]Department of Electronics Engineering, National Chiao-Tung University, Taiwan
[2]Department of Computer Science, National Chiao-Tung University, Taiwan
[3]Networks and Multimedia Institute, Institute for Information Industry, Taiwan

ABSTRACT

In this paper, we propose a layer-adaptive mode decision algorithm and a motion search scheme for the scalable video coding (SVC) with combined coarse granular scalability (CGS) and temporal scalability. To speed up the encoder while minimizing the loss in coding efficiency, our layer-adaptive mode decision recursively refers to the prediction modes and quantization parameter of the reference/base layer to minimize the number of modes tested at the enhancement layer. Moreover, our motion search scheme adaptively reuses the reference frame indices of the base layer and determines the initial search point using the motion vector at the base layer or the motion vector predictor at the enhancement layer. As compared with JSVM 8, the proposed algorithms provide up to 75% overall time saving and more than 85% time reduction for encoding enhancement layers with negligible loss in coding efficiency.

Index Terms— Scalable Video Coding, Fast Mode Decision, Encoder Optimization

1. INTRODUCTION

To support clients with diverse capabilities, ISO/IEC MPEG and ITU-T form a Joint Video Team (JVT) to develop a scalable video coding (SVC) technology [1] that uses single bitstream to provide multiple spatial, temporal, and quality (SNR) resolutions. As a scalable extension of H.264/AVC, it provides a H.264/AVC-compatible base layer and multiple enhancement layers, which can be truncated and extracted on-the-fly to obtain a preferred spatio-temporal and quality resolution. The scalability is achieved by using the layered coding structure in combination with adaptive inter-layer prediction, which causes extremely high complexity.

In [2-5], fast mode decision algorithms are proposed to consider the computational correlations between different layers for accelerating the encoder. In [5], the intra prediction at the CGS enhancement layer is forced to be either Intra4x4 or IntraBL. In [2], the mode decision at the CGS layer is designed in such a way that the macroblocks at the enhancement layer will have finer partition than their counterparts at the base layer. Similar principles are further refined to additionally reuse the motion information and the rate-distortion cost of the base layer for spatial scalability [3]. In addition, in [4] the motion activity and the temporal level of the base layer are employed as the context for the mode selection in different temporal layers. However, all the schemes are designed with only 1 CGS/Spatial enhancement layer and 1 reference frame.

In this paper, we propose a layer-adaptive mode decision algorithm and a motion search scheme for the scalable video coding (SVC) with combined CGS and temporal scalability. The proposed schemes are tailored for multiple enhancement layers and reference frames. Also, the computational redundancy is minimized in addition to the mode reduction. Specifically, depending on the prediction modes and the quantization parameter of the reference/base layer, a look-up table is used to minimize the number of modes to be tested for each enhancement layer. Moreover, considering the similarity between the motion information of different coding layers, the reference frame indices of the base layer are adaptively reused. And, according to the macroblock partition, the initial point for each partition is selected by the motion vector at the base layer or the motion vector predictor at the enhancement layer. As compared with JSVM 8 [6], the proposed schemes provide up to 75% overall time saving and more than 85% time saving for encoding enhancement layers with negligible loss in coding efficiency.

The rest of this paper is organized as follows: Section 2 shows the correlations between coding layers. Based on the analyses, Section 3 presents the proposed layer-adaptive mode decision and motion search algorithms. In Section 4, the proposed schemes are compared with JSVM 8 [6]. Lastly, Section 5 summarizes the concluding remarks.

* Thanks to Institute for Information Industry, Taiwan for supporting this work.

Table 1. Conditional Probability of the Macroblock Modes at the Enhancement Layer

Enhancement layer	Inter prediction modes at base layer							
	$(Qp_B, Qp_E) = (39,29)$				$(Qp_B, Qp_E) = (27,17)$			
	16x16	**16x8**	**8x16**	**8x8**	**16x16**	**16x8**	**8x16**	**8x8**
16x16	0.44	0.25	0.28	0.04	0.19	0.18	0.15	0.10
16x8	0.19	0.25	0.12	0.04	0.16	0.18	0.07	0.05
8x16	0.20	0.12	0.29	0.04	0.17	0.08	0.17	0.05
8x8	0.17	0.38	0.31	0.88	0.48	0.56	0.61	0.80

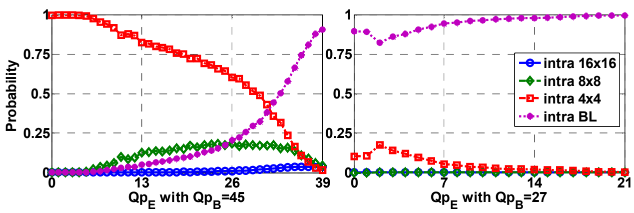

Fig. 1. Distribution of the intra prediction modes at the enhancement layer.

2. CORRELATIONS BETWEEN BASE AND ENHANCEMENT LAYERS

In this section, we analyze the correlations between the coding layers with a focus on the configuration of combined CGS and temporal scalability. The statistical analyses are conducted based on the coding of 1 base layer and 1 CGS layer in CIF resolution. For simplicity, the Qp_B, Qp_E, and Qp_0 thereafter stand for the quantization parameters of the base layer, the enhancement layer, and the reference layer from which an enhancement layer is predicted, respectively.

With similar video characteristic, the macroblock partition at the enhancement layer is strongly related to that at the base layer. To characterize the correlations, in Table 1 we show the conditional probability of the macroblock modes at the CGS layer when the base layer is coded at different quality with the associated Qp_E setting suggested by JVT/MPEG committee. From Table 1, the partition of the base layer can be a good reference for the prediction of the macroblock partition at the enhancement layer. For instance, if a macroblock at the base layer is coded with the 16x8/8x16 partition, it is unlikely that its counterpart at the enhancement layer will have an aspect ratio of 8x16/16x8. Moreover, when the base layer uses the 8x8 partition for better prediction, the enhancement layer will also benefit from using the same partition. Further, increasing the quality of the enhancement layer tends to increase the percentage of the 8x8 partition.

Similar to the macroblock partition, the distribution of the intra prediction mode is highly dependent on the quality of the base and enhancement layers. In Fig. 1, when the base layer is coded with better quality using a smaller quantization parameter, most of the intra predictions are made of the IntraBL, which refers to the intra-coded macroblocks at the base layer. On the other hand, while the

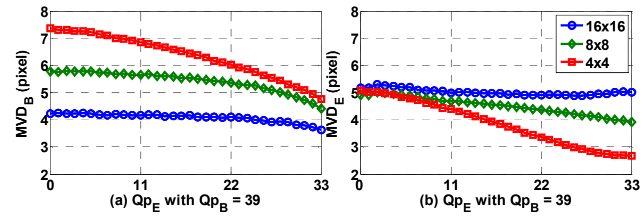

Fig. 2. The motion vector differences using (a) the motion vector of the base layer or (b) the motion vector predictor at the enhancement layer as reference.

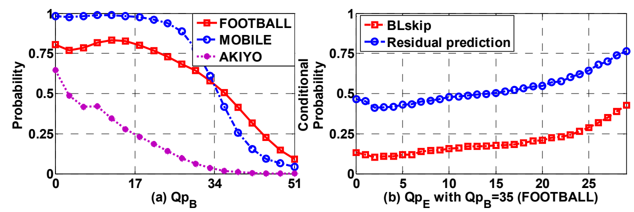

Fig. 3. (a) Probability of testing residual prediction. (b) Conditional probability of using residual prediction.

quality of the enhancement layer is gradually improved, the intra predictor is switched from the base layer to the enhancement layer. Particularly, for better prediction, the percentage of the Intra4x4 is increased more noticeably than the other two modes. Similar results are also reported in [5].

The motion vectors of the base layer and the enhancement layer are also correlated to a certain extent. In Fig. 2 (a), the motion vector of the base layer provides a good prediction for that at the enhancement layer when the macroblock has a partition size greater than 8x8. On the other hand, Fig. 2 (b) shows that the motion vector predictor at the enhancement layer is a better choice for a submacroblock partition. By adaptively selecting one of the predictors as the initial search point, the search range at the enhancement layer can be reduced. Similarly, the statistical analyses also reveal that the macroblocks at the enhancement layer are likely to have the same reference pictures as their counterparts at the base layer. An exception is that when the base layer is coded at low bit rate and the macroblock selects a partition size of 16x16, the motion information of the base layer may not be reliable to be used at the enhancement layer.

For the inter-layer residual prediction, the testing is performed when the residues at the base layer are not quantized to zero. In Fig. 3 (a), the probability for testing the inter-layer residual prediction increases significantly when the quality of the base layer is improved by decreasing the quantization parameter. However, from the conditional probability in Fig. 3 (b), only half of the macroblocks tested for inter-layer residual prediction are actually coded with residual prediction. Moreover, depending on the quality of the enhancement layer, a large percentage of those macroblocks may be coded with BLSkip mode, which reuses

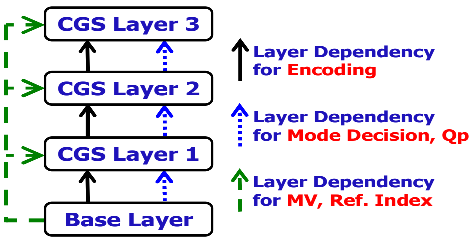

Fig. 4. Layer dependency for encoding and mode decision.

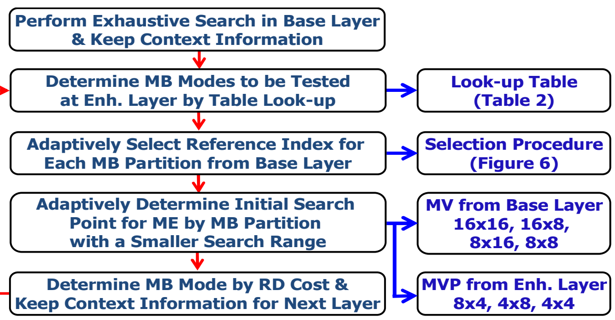

Fig. 5. Layer-adaptive mode decision.

the motion information of the base layer. As a result, significant reduction in coding time is expected by properly enabling the testing of inter-layer residual prediction.

3. PROPOSED MODE DECISION ALGORITHM

Based on the observations in Section 2, we develop a layer-adaptive mode decision algorithm and a motion search scheme for the combined CGS and temporal scalability. As shown in Fig. 4 and Fig. 5, to make use of the inter-layer correlations, the base layer is encoded with exhaustive search and all the motion information for each possible combination are kept. The coding modes to be tested at an enhancement layer are then looked up in Table 2 by referring to the quantization parameters and the coding modes of the reference layer. Furthermore, the algorithm in Fig. 6 determines the reference frames for motion search and depending on the partition size, the initial point is adaptively selected by the motion vector at the base layer or the motion vector predictor at the enhancement layer.

3.1. Layer-Adaptive Mode Decision

To achieve great saving in coding time while minimizing the loss in coding efficiency, the layer-adaptive mode selection in Table 2 is designed in such a way that the inter-layer correlations are respected. Statistically, the macroblock modes that have limited contribution to coding efficiency are skipped for testing. For instance, when the reference layer is coded using a quantization parameter in the range of (34~51), the inter-layer residual prediction is skipped due to the reference layer of poor quality. Moreover, as suggested by the analyses, when a macroblock is coded with

Table 2. Look-up Table for Layer-adaptive Mode Decision

Candidate modes in layer *N*	Best prediction modes in layer *M* with Qp_0											
	Region 1: Qp_0 in (0~33)						Region 2: Qp_0 in (34~51)					
	Inter prediction modes					Intra	Inter prediction modes					Intra
	Direct	16x16	16x8	8x16	8x8		Direct	16x16	16x8	8x16	8x8	
Direct	O	O	O	O	O	O	O	O	O	O	O	O
16x16		O					O	O	O	O		O
16x8		O	O				O	O	O			O
8x16		O		O			O	O		O		O
8x8			O	O	O			O	O	O	O	
BLskip	O	O	O	O	O		O	O	O	O	O	
$Direct_{RES}$	O	O	O	O	O	O						
$16x16_{RES}$		O										
$16x8_{RES}$		O	O									
$8x16_{RES}$		O		O								
$8x8_{RES}$			O	O	O							
$BLskip_{RES}$	O	O	O	O	O							
MODE_SR	O	O	O	O	O							
Intra4x4	O	O	O	O	O	O	O	O	O	O	O	O
IntraBL						O						O

8x8 prediction mode includes 8x8, 8x4, 4x8, and 4x4 partition modes. The suffix, RES, denotes a prediction mode with residual prediction.

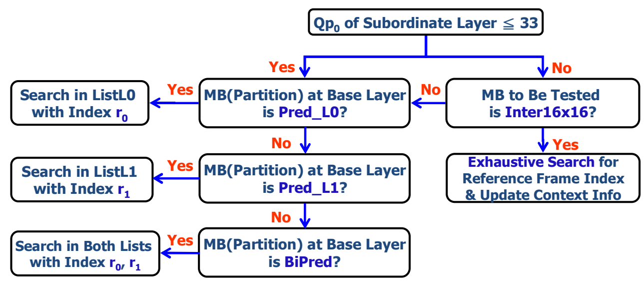

Fig. 6. Layer-adaptive selection of reference frame index.

16x8/8x16 partition, its counterpart at the enhancement layer will not be evaluated with the partition of 8x16/16x8. Also, for a macroblock coded with the 8x8 partition, its counterpart at the enhancement layer will also not be tested with any modes that have a partition size greater than 8x8. For further improvement, the intra prediction at the enhancement layer is enforced to be either Intra4x4 or IntraBL.

On the other hand, when the reference layer is coded with better quality, the major changes comparing with the previous case includes that (1) all the modes with inter-layer residual prediction are considered for testing, and (2) when a macroblock at the reference layer is coded with 16x8, 8x16, or 8x8, only the 8x8 prediction mode and the mode with the same partition are tested. For the later case, although Table 1 suggests that such design may not be optimal in terms of the mode distribution, the experimental results show that replacing the 16x16 partition with the 16x8 or 8x16 partition has ignorable influence to the coding efficiency, especially when the enhancement layer is coded with high quality.

3.2. Layer-Adaptive Motion Search

Analogy to the mode selection, the layer-adaptive motion search in Fig. 6 is designed to avoid exhaustive motion search at the enhancement layer by taking advantage of the motion information at the base layer. When a macroblock at the enhancement layer is to be tested with the prediction modes other than 16x16, the associated reference indices at the base layer are reused. However, exhaustive search is performed for the 16x16 prediction mode when the reference layer is coded with poor quality, which suggests that the motion information at the base layer is not reliable for being used at the enhancement layer.

4. EXPERIMENTS

For the experiments, the proposed algorithm is implemented with JSVM 8 [6] and tested using 10 video sequences in CIF/4CIF formats. The hierarchical bidirectional prediction is enabled for the temporal scalability while 3 CGS enhancement layers are created in addition to the base layer for the SNR scalability. Moreover, the video sequences used for statistical analyses are separated from those for testing. More encoder configurations are detailed in Table 3.

In Table 4, the proposed schemes are compared with JSVM 8 [6] in terms of the overall time saving (TS), the time saving for encoding the enhancement layers (TS_E), and the complexity ratio (CR) defined as the encoding time of the base layer over the encoding time of the 3 enhancement layers. As shown, the proposed scheme averagely provides up to 75% overall time saving and more than 85% time saving for encoding the enhancement layers. Moreover, the average increase of bit rate is confined within 1% and the PSNR loss is negligible. With minor changes in PSNR and bit rates, one cannot distinguish the differences between the proposed schemes and JSVM 8 [6] in terms of rate-distortion performance. In addition, the encoding of 3 enhancement layers takes similar or less time as for the encoding of 1 base layer. Similar results are also observed with different GOP structures and numbers of coding layers. Further, as compared with the prior work in [2], in which only 1 reference frame is used for temporal prediction, the proposed scheme shows similar or better performance with more reference frames.

5. CONCLUSION

In this paper, we propose a layer-adaptive mode decision algorithm and a motion search scheme for the scalable video coding (SVC) with combined coarse granular scalability (CGS) and temporal scalability. The computational correlations between coding layers are explored for accelerating the encoding process. As compared with JSVM 8 [6], our scheme provides up to 75% overall time saving and more than 85% time saving for encoding enhancement layers with negligible changes in coding efficiency. The scalability is achieved with minimized extra complexity. Moreover, the proposed scheme can be further extended to support fully scalable video coding.

Table 3. Testing Conditions

Testing Sequences	CIF	AKIYO, FOOTBALL, FOREMAN, MOBILE, STEFAN
	4CIF	CITY, CREW, HARBOUR, ICE, SOCCER
(Qp_B,Qp_{E1},Qp_{E2},Qp_{E3})	(40,30,20,10) and (30,20,10,0)	
Encoder Configuration	MV Search Range: 32 pixels RDO: Enabled, GOP Size: 8 Number of Reference Frames: 3	

Table 4. Performance Comparisons with JSVM 8 [6]

Sequence	Qp_B	Qp_E	ΔPSNR (dB)	ΔBitrate (%)	TS(%)	TS_E(%)	CR1	CR2
FOREMAN	40	30	-0.02	0.49	73.4	87.5	5.21	0.65
		20	-0.04	3.15				
		10	0.01	1.65				
	30	20	-0.01	1.44	77.8	90.0	6.36	0.64
		10	0.01	0.71				
		0	0.01	0.25				
STEFAN	40	30	-0.02	0.62	71.9	84.8	5.55	0.84
		20	-0.01	1.11				
		10	0.00	0.61				
	30	20	0.00	0.30	75.4	86.6	6.77	0.91
		10	0.00	0.18				
		0	0.00	0.11				
HARBOUR	40	30	-0.01	0.30	74.9	87.1	6.14	0.79
		20	0.00	0.26				
		10	0.00	0.09				
	30	20	0.00	0.09	77.6	87.8	7.60	0.92
		10	0.00	0.03				
		0	0.00	0.00				
SOCCER	40	30	-0.02	0.70	74.3	87.4	5.67	0.71
		20	0.01	1.89				
		10	0.02	0.76				
	30	20	0.01	0.86	77.6	89.2	6.70	0.72
		10	0.01	0.32				
		0	0.00	0.09				

TS: time saving of overall encoding time
TS_E: time saving of overall encoding time only at enhancement layers
CR1: the complexity ratio of base layer to enhancement layers with JSVM 8
CR2: the complexity ratio of base layer to enhancement layers with proposed algorithm

6. REFERENCES

[1] T. Wiegand, G. Sullivan, J. Reichel, H. Schwarz, and M. Wien, "Joint Draft 8 of SVC Amendment," *ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-U201*, 2006.

[2] H. Li, Z.-G. Li, and C.Wen, "Fast Mode Decision for Coarse Grain SNR Scalable Video Coding," *IEEE ICASSP*, 2006.

[3] H. Li, Z.-G. Li, and C.Wen, "Fast Mode Decision for Spatial Scalable Video Coding," *IEEE ISCAS*, 2006.

[4] H. Li, Z.-G. Li, and C. Wen, "Fast Mode Decision Algorithm for Inter-Frame Coding in Fully Scalable Video Coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 7, pp. 889-895, 2006.

[5] L. Yang, Y. Chen, J. Zhai, and F. Zhang, "Low Complexity Intra Prediction for Enhancement Layer," *ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-Q084*, 2005.

[6] J. Reichel, H. Schwarz, and M. Wien, "Joint Scalable Video Model JSVM-8," *ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-U202*, 2006.