NEW INTER-LAYER INTRA PREDICTION FOR SPATIAL SCALABLE VIDEO CODING

Chang-Ming Lee, Yu-Ciao Yang, Jui-Chiu Chiang

Department of Electrical Engineering
National Chung Cheng University
Chia-Yi, Taiwan
Email: cmlee@ccu.edu.tw, ycyang@samlab.ee.ccu.edu.tw, rachel@ccu.edu.tw

ABSTRACT

Transmitting the digital audio/visual data over the internet becomes more and more popular lately. Due to the heterogeneous networks and different hardware devices, the users have different requirements for the same multimedia request. Thus, scalable video coding (SVC) was proposed and *became an extension of H.264/AVC video coding standard. To make the bitstream embedded while ensuring good coding efficiency, SVC uses inter-layer prediction to remove the redundancy between the enhancement layer and the base layer, in addition to intra prediction and inter prediction. After analyzing the coding efficiency brought by each kind of inter-layer prediction, the importance of inter-layer intra prediction is recognized. In this paper, we propose a new inter-layer intra prediction to improve the coding efficiency for spatial scalability. Experimental results show that an average 4.92% bitrate reduction is achieved by the proposed inter-layer intra prediction..*

1. INTRODUCTION

With the advance in video streaming in the network technology and also the increased demand of wireless transmission, many techniques are devoted to dealing with various transmission conditions. For different channel environment, as well as distinct transmission quality, it is an imperative issue to develop an efficient compression technique for multiple users who try to access the same video content through different communication links. As a result, the functionality of scalability is taken into account in the standard of MPEG 2[1]. Furthermore, the standardization of scalable video coding (SVC) was finalized lately [2] and become an extension of H.264/AVC [3] by offering the scalability in terms of spatial, temporal and quality scalability.

Figure 1 shows the coding architecture of SVC standard with two-layer spatial scalability and quality scalability. A low resolution input video can be optionally generated from a high resolution video by spatial downsampling and followed by H.264/AVC encoding to form the base layer. Then, a quality-refined version of the low resolution video can be obtained by combining the base layer with the enhancement layer where coarse grain scalability (CGS) or medium grain scalability (MGS) can be adopted. Similar to the H.264/AVC encoding procedure, for every MB of the current frame, only

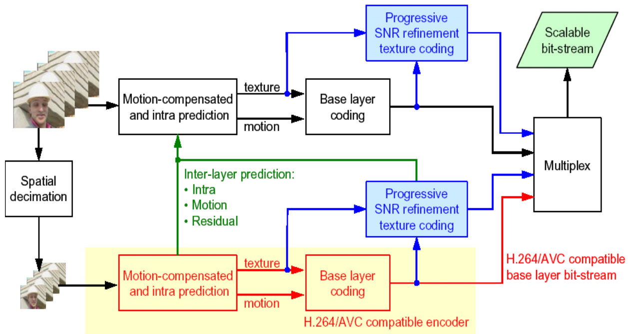

Figure 1. The SVC coding architecture with two spatial layers [2]

the residual with respect to its prediction will be encoded in scalable video coding.

The realization of spatial scalability in SVC is mainly relied on inter-layer prediction from the base layer to the enhancement layers, in addition to intra prediction and inter prediction. Accordingly, more redundancy can be removed and a better coding efficiency is yielded, compared to simulcast where each spatial layer is encoded by H.264/AVC independently. Three kind of inter-layer predictions are offered in spatial scalability, including inter-layer intra prediction (IBL mode), inter-layer motion prediction and inter-layer residual prediction by exploring the high correlation in textures, motions and residuals between the base layer and the enhancement layer, respectively.

Although an almost compatible coding efficiency compared to single layer coding can be achieved for SVC by inter-layer prediction, the computation requirement for SVC encoder is quite high. Hence, many fast mode decision algorithms specified for SVC were proposed recently [4-6].

Besides, to improve the coding efficiency of inter-layer prediction, several works present new ideas to modify the inter-layer prediciton. In [7], a new inter-layer intra prediction is propoed. Instead of using dierctly the interpolated image upsampled from the reconstructed low resolution image, motion estimation is performed on the interpolated image to compensate some possible drift due to interpolation. In [8], a 2D Wiener filter is used to replace the 4-taps filters used in the SVC standard to minimise the prediction error when inter-layer intra prediction is employed. However, both methods in [7][8] consider only informaiton from the current low-resolution image. Moreover, the upsampling process usually introduces blur artefact and limits the accuracy of prediction. Hence, in this paper, we propose a new inter-layer intra prediction algorithm for spatial scalability by introducing suitable high frequency element into the interpolated frame to enhance the similarity between the MB in the enhancement layer and the prediciton candiate derived from the base layer.

The rest of this paper is organized as follows. Section 2 presents some observations of spatial scalability realization. The proposed techniques and related experiments are detailed in Section 3and 4, respectively. Finally, Section 5 draws some conclusions.

2. OBSERVATIONS ON SPATIAL SCALABILITY

Table 1 illustrates the time distribution for coding two spatial layers using JSVM. It shows that the total encoding time for the high resolution layer is almost 10 times of that for the low resolution layer. The increased complexity is due to the inter-layer motion prediction and inter-layer residual prediction, where supplementary motion estimation and motion compensation are needed.

Table 1. Time distribution of encoding two spatial layers

Sequence	Foreman
Frame	100
QP	28
Inter layer prediction	on
Search range	16
Reference frame	1
Search Mode	full search
GOP	16

Time distribution of encoding two spatial layers

QCIF	9%
CIF	91%

Time distribution of encoding the high resolution layer

ME/MC	94.2%
Deblock filtering	0.1%
Upsampling for inter-layer intra prediction	0.1%
Transformation	2.3%
Quantization	0.7%
Entropy coding	1.6%
Else	1.0%

To understand the coding efficiency, as well as the complexity of inter-layer prediction, we disable one kind of inter-layer prediction each time while keeping the other two inter-layer predictions unchanged. Table 2 illustrates the results in terms of bitrate and PSNR variations with respect to the condition where all inter-layer predictions are enabled. The test sequence is Foreman with two spatial layers (QCIF, CIF) @30Hz and group of picture (GOP) size considered is 1, 8 and 16. It reveals that the bitrate increases obviously when inter-layer intra prediction is disabled, especially for a large QP value. Similar result is observed when the residual from the base layer is not referenced, especially at moderate bitrate, e.g. QP is 28. Besides, there is no significant impact if inter-layer motion prediction is disabled. These results demonstrate the importance of inter-layer intra prediction and inter-layer residual prediction when targeting at a better coding performance.

In addition, Table 3 details the statistical distributions of three kind inter-layer predictions employed for the MB in the enhancement layer. In particular, the adoption of inter-layer intra prediction is very high when GOP is 1, in particular for low bitrate scenario; it implies that the prediction from the base layer is more accurate than that from intra prediction. Besides, the inter-layer residual prediction is likely to be adopted for the moderate bitrate scenario. However, the adoption of inter-layer motion prediction is not very high and it indicates that the rate-distortion performance by predicting the motion vector of MBs in the enhancement layer from the spatial neighbouring blocks is better than that from corresponding blocks in the base layer.

Table 2. Coding efficiency with/without inter layer predictions

(a) GOP=1

GOP=1	**No inter-layer intra pred.**	
CIF	Δbitrate(%)	ΔPSNR(dB)
QP=16	11.22%	0.3420
QP=28	15.06%	0.1279
QP=40	28.65%	0.7032

(b) GOP=8

	No inter-layer intra pred.		**No inter-layer residual pred.**		**No inter-layer motion pred.**	
QP	Δbitrate(%)	ΔPSNR (dB)	Δbitrate(%)	ΔPSNR (dB)	Δbitrate(%)	ΔPSNR (dB)
16	2.16%	0.0345	1.22%	-0.0051	0.31%	-0.0059
28	5.62%	0.0543	2.65%	0.0034	0.54%	-0.0088
40	14.17%	0.5869	0.45%	-0.0074	-0.50%	-0.0245

(c) GOP=16

	No inter-layer intra pred.		**No inter-layer residual pred.**		**No inter-layer motion pred.**	
QP	Δbitrate(%)	ΔPSNR (dB)	Δbitrate(%)	ΔPSNR (dB)	Δbitrate(%)	PSNR (dB)
16	1.31%	0.0158	1.37%	-0.0135	0.36%	-0.0046
28	3.46%	0.0159	3.07%	0.0073	0.77%	0.0008
40	9.24%	0.4865	0.57%	-0.0022	-0.89%	-0.0327

Table 3. Percentage of inter layer prediction employed in the enhancement layer

	GOP=1			**GOP=4**		
QP	**motion**	**intra**	**residual**	**motion**	**intra**	**residual**
16		89.39%		6.78%	22.55%	29.79%
28		93.98%		6.77%	23.69%	34.28%
40		97.80%		3.34%	24.46%	24.48%
	GOP=8			**GOP=16**		

QP	motion	intra	residual	motion	intra	residual
16	7.77%	11.91%	33.57%	8.28%	6.60%	35.57%
28	7.93%	12.39%	40.34%	8.34%	6.81%	43.14%
40	3.94%	12.74%	28.42%	4.29%	6.93%	30.61%

3. PROPOSED ALGORITHM

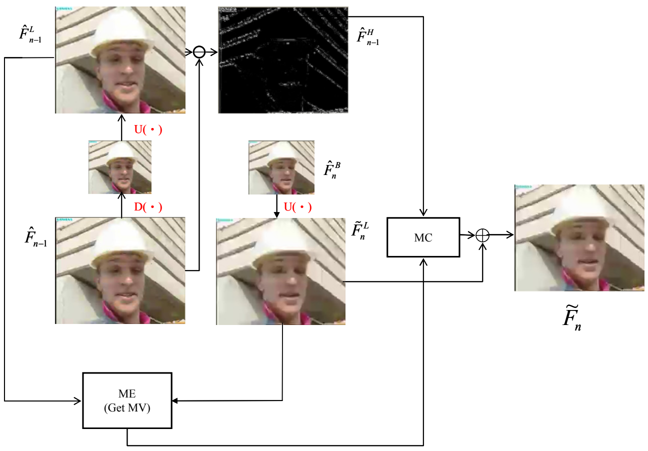

Figure 2. Block diagram of the proposed algorithm

From Table 2 and 3, we observe that the coding efficiency is significantly improved by the inter-layer prediction, especially the inter-layer intra prediction. Generally, the inter-layer intra prediction is carried out as below: the low resolution layer is first encoded, then the reconstructed MB in the low resolution layer will be upsampled using a 4-tap filter [9] and served as a prediction for the corresponding MB in the high resolution layer. Such a prediction can be treated as an additional intra-prediction mode with a block size of 16×16, and it is often called IBL mode. After rate-distortion computation, if the cost of IBL mode is smaller than intra prediction, as well as inter prediction, the MB in the high resolution layer will be predicted by the upsampled reconstructed block in the low resolution layer.

Undoubtedly, the interpolation by spatial filtering brings the blurring artefact and the coding efficiency is expected to be improved if high frequency component is introduced appropriately on the interpolated frame. Thus, we propose a new inter-layer intra prediction algorithm by exploiting the redundancy between the reconstructed current low-resolution frame and the previously reconstructed high-resolution frame.

Figure 2 depicts the block diagram of the proposed algorithm for inter-layer intra prediction. The details are described as follows:

1. The previously reconstructed high-resolution frame $\hat{F}_{n-1}$ is downsampled and upsampled to generate $\hat{F}^{L}_{n-1}$, which can be seen as the low-frequency part of $\hat{F}_{n-1}$. Then the high frequency part of $\hat{F}_{n-1}$ is obtained by

$$\hat{F}^{H}_{n-1} = \hat{F}_{n-1} - \hat{F}^{L}_{n-1} \quad (1)$$

2. Then, the reconstructed current low-resolution frame $\hat{F}^{B}_{n}$ is upsampled to have $\tilde{F}^{L}_{n}$. For each 16x16 block in $\tilde{F}^{L}_{n}$, a motion estimation over $\hat{F}^{L}_{n-1}$ is performed and get a pair of motion vector accordingly.
3. Finally, we can introduce the high frequency information of $\hat{F}^{H}_{n-1}$ to $\tilde{F}^{L}_{n}$ by motion compensation with the motion vector available during the motion estimation between $\tilde{F}^{L}_{n}$ and $\hat{F}^{L}_{n-1}$. The new inter-layer intra prediction is expressed as below

$$\tilde{F}_{n} = \tilde{F}^{L}_{n} + MC(\hat{F}^{H}_{n-1}) \quad (2)$$

where $MC(\hat{F}^{H}_{n-1})$ denotes the operation of motion compensation. Then, in addition to the original IBL mode, the new inter-layer intra prediction is served as a candidate and the rate-distortion optimisation will choose a best prediction among all possible modes.

Note that, there is no need to transmit the motion vectors between $\hat{F}^{L}_{n-1}$ and $\tilde{F}^{L}_{n}$, which can be calculated at the decoder.

4. EXPERIMENTAL RESULTS

To evaluate the coding performance, we implement the proposed algorithms on JSVM 9.18 [9]. Four sequences with diverse motion characteristics are tested, including "Foreman", "Bus", "Football" and "Akiyo". Table 4 shows the associated encoding configuration for simulations, where the QP used for both layers are the same. Table 5 summarised the performance of the spatial scalable video coding scheme after realizing the proposed inter-layer intra prediction as another candidate mode, in addition to original supported prediction modes. The average bitrate reduction is 4.92%. Table 6 list the percentages of each mode employed in I-frames. It shows that the proposed inter-layer intra prediction is chosen for 17.44% to 25.84% MBs in the enhancement layer. Besides, we compare the proposed algorithm with an interpolation scheme presented in [10] where a hybrid DCT-Wiener-Based filter is designed. Table 7 lists the performance of the proposed algorithm and the method in [10] using the Bjontegaard metric [11]. It indicates the proposed algorithm outperforms the one in [10] due to the introduction of appropriate high-frequency component in the inter-layer intra prediction mode.

5. CONCLUSION

In this paper, we proposed a new inter-layer intra prediction by introducing appropriate high frequency component from the previously reconstructed frame into the interpolated frame. Experimental results indicate the proposed algorithm can reduce the bitrate up to 9% for spatial scalability realization while maintaining the image quality. To further improve the coding efficiency, a sophisticated method for motion estimation could be carried out to enhance the correctness of correspondence matching.

Table 4. Simulation configurations

	BL	EL
Resolution	QCIF	CIF
QP	16, 24, 32, 40	
Frame Rate	30	
Intra Period	1	
Entropy Coding	CABAC	
Inter-layer prediction	Adaptive	
Number of Frames	100	

Table 5. RD performance of the proposed algorithm

sequence	QP	ΔPSNR(dB)	ΔBitrate(%)
Foreman	40	-0.07	-3.78
	32	-0.04	-4.62
	24	-0.04	-4.20
	16	-0.04	-2.93
Bus	40	-0.07	-6.35
	32	-0.06	-5.92
	24	-0.05	-4.91
	16	-0.05	-3.69
Football	40	-0.09	-2.85
	32	-0.04	-3.02
	24	-0.03	-2.52
	16	-0.03	-1.79
Akiyo	40	-0.23	-7.01
	32	-0.29	-9.22
	24	-0.22	-9.00
	16	-0.13	-6.94
AVG		-0.09	-4.92

Table 6. Percentages of each mode employed in I-frames

(a) Foreman

Original				
	QP=16	QP=24	QP=32	QP=40
Intra16*16	1.57%	2.12%	6.03%	11.45%
Intra4*4	33.85%	27.45%	19.44%	12.82%
IBL	64.58%	70.43%	74.53%	75.73%
Proposed				
	QP=40	QP=32	QP=24	QP=40
Intra16*16	1.53 %	1.64 %	5.03%	10.79%
Intra4*4	26.42%	19.84%	14.43%	10.29%
IBL	54.61%	55.41%	54.70%	57.57%
New IBL	17.44%	23.10%	25.84%	21.35%

(b) Akiyo

Original				
	QP=16	QP=24	QP=32	QP=40
Intra16*16	2.33	14.68	20.26	14.25
Intra4*4	27.50	15.86	10.53	4.31
IBL	70.17	69.47	69.21	81.44
Proposed				
	QP=16	QP=24	QP=32	QP=40
Intra16*16	2.06%	13.78%	19.70%	13.91%
Intra4*4	17.98%	8.70%	5.79%	2.27%
IBL	52.32%	51.88%	50.87%	61.27%
New IBL	27.64%	25.65%	23.63%	22.56%

Table 7. Performance for the proposed IBL mode and the method in [10] with respect to the original SVC standard

(a) proposed

	Foreman	Bus	Football	Akiyo
BDBR(%)	-3.36	-4.58	-1.99	-5.13
BDPSNR(dB)	0.21	0.39	0.14	0.34

(b) method in [10]

	Foreman	Bus	Football	Akiyo
BDBR(%)	-0.42	-0.52	-0.44	-0.61
BDPSNR(dB)	0.03	0.05	0.03	0.04

REFERENCES

[1] ISO/IEC. IS 13818: Information technology - Generic coding of moving pictures and associated audio information: Systems, 1995. (MPEG-2)

[2] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 17, no.9, pp.1103-1120, Sept. 2007.

[3] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 13, pp.560-576, July 2003.

[4] He Li, Z. G. Li, "Fast Mode Decision Algorithm for Inter-Frame Coding in Fully Scalable Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, issue: 7, pp. 889-895, July 2006.

[5] W. Zhang, A. Men, P. Chen, "Adaptive Inter-layer intra Prediction in Scalable Video Coding," in *Proc.* of *IEEE International Symposium on Circuits and Systems*, 2009, pp. 876-879.

[6] J.-C. Chiang and C.-Y. Hu, "Fast Mode Decision Algorithm for Spatial Scalable Video Coding," *IET Electronics Letters*, vol. 46, no. 25, 9th December, 2010.

[7] W. Zhang, A. Men, P. Chen, "Adaptive inter-layer intra prediction in scalable video coding," in *Proc.* of *IEEE International Symposium on Circuits and Systems*, 2009, pp. 876-879.

[8] M. M. Islam, V. K. Asari, M. N. Islam, M.. A. Karim, "Super-Resolution Enhancement Technique for Low Resolution Video," *IEEE Transactions on Consumer Electronics,* vo.56, issue: 2, pp. 919-924, May 2010.

[9] "Joint Scalable Video Model (JSVM) 9.18 Software," *JVT of ITU-T VCEG and ISO/IEC MPEG.*

[10] Z. Wu, H. Yu, and C. W. Chen, "A New Hybrid DCT-Wiener-Based Interpolation Scheme for Video Intra Frame Up-Sampling," *IEEE Signal Processing Letters*, vol.17, no. 10, pp. 827-830, Oct. 2010.

[11] G. Bjontegaard, "Calculation of Average PSNR Differences between RD-curves," *13th VCEG-M33 Meeting*, Austin, TX, Apr., 2001.