International Journal of Innovative Computing, Information and Control
Volume **8**, Number **5(A)**, May **2012** ISSN 1349-4198
pp. 3103–3113

LOW COMPLEXITY ALGORITHM FOR INTER-LAYER PREDICTION OF H.264/SVC

Tian Song, Kentaro Takei and Takashi Shimamoto

Graduate School of Advanced Technology and Science
The University of Tokushima
Minami-Jyosanjima 2-1, Tokushima City 770-8506, Japan
{ tiansong; kentaro_t; simamoto }@ee.tokushima-u.ac.jp

Received March 2011; revised August 2011

Abstract. *Inter-layer prediction of H.264/SVC can help to improve the coding efficiency with increased computation complexity. This paper presents a low complexity algorithm for inter-layer prediction. The proposed algorithm focuses on the reduction of the candidate modes by making use of the correlations of the encoding cost between the base layer and enhancement layers. For macroblocks with the collocated macroblock coded by INTRA or INTER type two algorithms are proposed, respectively. Both algorithms efficiently decrease the redundant candidate modes by estimating from the base layer coding information. The experiment results show that the proposed algorithm can significantly reduce redundant computation complexity with almost no coding efficiency loss.*
Keywords: H.264, Scalable video coding, Hierarchical B-picture, Search range

1. **Introduction.** In recent years, encoding standard which can achieve the scalability has increasing requirement due to the diversification of the network environment and the applications. Enhancing to the successful H.264/AVC, a scalable extension is standardized as H.264/SVC [1] in 2007. A reference software is also developed by the joint video team (JVT) for SVC [2, 3]. The objective of H.264/SVC is to enable the generation of a unique bitstream that can adapt to various bit-rate, transmission channel and display capabilities. In H.264/SVC, three scalabilities: spatial scalability, temporal scalability and quality scalability are finally recommended in the final draft [4, 5, 6]. However, due to its high implementation complexity, the complexity reduction becomes an important research issue. Some previous works are proposed to reduce the complexity of temporal scalability by adaptively reducing the redundant encoding modes or efficiently constructing the GOP (Group of Pictures) [7, 8, 9].

The resolution diversity of current display devices motivates the improvement for spatial scalability. The spatial scalability is realized by introducing multiple display resolutions within a single bit-stream. Therefore, the information of the input sequences and the selected modes in the base layer can be used to estimate the optimal mode in the enhancement layers, which is called inter-layer coding [3]. In the inter-layer prediction, three new modes have been introduced using the motion vectors, residuals, and intra information from the base layer to select the best coding mode in the enhancement layers. Using these new modes, inter-layer predictions can not only achieve scalable features but also improve the coding efficiency.

However, the inter-layer modes have to perform multiple times rate-distortion optimization (RDO) process, by which very high computational complexity is induced. In particular, the residual prediction mode has to perform twice of the RDO process which doubles the computational complexity of the normal RDO process of H.264/AVC. A previous work introduced an efficient architecture to reduce the implementation complexity by changing the processing order [10]. However, the computational complexity reduction is not considered from the viewpoint of candidate modes reduction. Other works achieved a very high complexity reduction rate by decreasing some of the candidate modes for H.264/SVC [11, 12]. However, the improvement of these previous works is on the basis of the sacrifice of the video quality for enhancement layer.

For many typical implementations, increasing requirement for video quality is the most important issue. It is no doubt that the higher enhancement layer it is, the higher video quality is required. On the other hand, low power consumption is also an important issue for many handheld devices especially for those real-time video codec applications. Therefore, the video quality as well as the low complexity implementation is important. Our previous work gave an efficient mode reduction method to decrease the encoding complexity [13]. However, the complexity reduction rate is not efficient enough. In this work, the correlation between base and enhancement layer is deeply investigated and the INTRA and INTER coded macroblocks are analyzed individually.

2. **Inter-Layer Prediction.** In order to employ base layer coding information for spatial enhancement layer coding, additional macroblock modes have been introduced in spatial enhancement layers. Three techniques to realize inter-layer prediction have been recommended in H.264/SVC, namely inter-layer motion prediction, inter-layer intra prediction and inter-layer residual prediction. Therefore, in addition to the traditional coding modes of H.264/AVC, several new modes are introduced in H.264/SVC. Figure 1 shows the total modes of H.264/SVC.

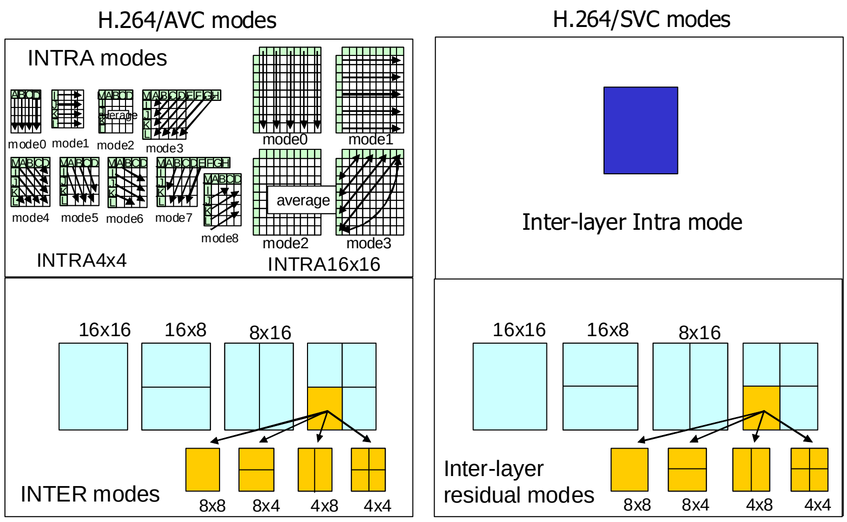

FIGURE 1. Coding modes of H.264/SVC

2.1. **Inter-layer prediction modes.** The same as H.264/AVC, generally INTRA and INTER types of coding modes are used. INTER type modes utilize the correlation in temporal direction and INTRA type modes utilize the correlation in spatial direction to decrease the redundant components. H.264/SVC employs a rate-distortion optimization (RDO) procedure which is very computationally intensive. This RDO process evaluates all possible modes to select the best one with the smallest rate-distortion cost (RDC). The evaluation of RDC can be described by the following equation.

$$J(s, c, Mode|QP, \lambda_{mode}) = SSD(s, c, Mode|QP) + \lambda_{mode} \cdot R(s, c, Mode|QP) \quad (1)$$

where QP is the macroblock quantization parameter, λ_{mode} is the lagrangian multiplier, SSD is the sum of square differences between the original block luminance signal denoted by s and its reconstructed signal denoted by c, and $R(s, c, Mode|QP)$ is the number of generated bits of a certain encoding mode. It includes the bits for the macroblock header, motion vector, and the DCT coefficients for the given block.

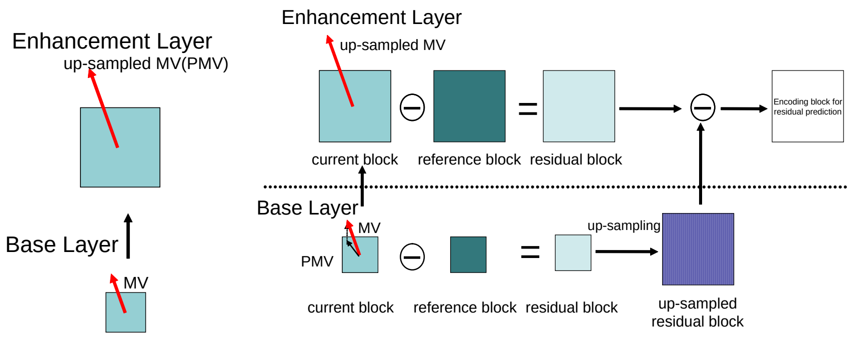

FIGURE 2. Inter-layer motion estimation; inter-layer INTRA prediction

As Figure 2 shows, Inter-layer residual prediction is one of the coding modes which are introduced by the inter-layer prediction to not only realize spatial scalability but also improve the coding efficiency. Because the motion vectors in the base layer and enhancement layer tend to have similar motion vectors, the up-sampled residual block also tends to have similar residuals with the corresponding block. When the residual prediction is used, the residuals of the corresponding 8×8 block in the base layer are block-wise up-sampled using a bilinear filter and used as prediction data for the residuals of the enhancement layer macroblock.

On the other hand, inter-layer intra prediction mode uses up-sampled intra coded macroblock as a prediction mode in enhancement layer coding.

3. **Proposed Algorithm.** The motivation of this work is trying to reduce the implementation complexity of enhancement layer by cutting down the redundant candidate modes of H.264/SVC. By efficiently using the encoding information of base layer we propose two low complexity algorithms for enhancement layer encoding. One algorithm is proposed for the co-located macroblock of which is INTRA type coded in base layer and the other one is for INTER type coded, respectively.

3.1. **Macroblock with co-located block is INTRA type coded in base layer.** Two methods are proposed for the complexity reduction of the macroblock with collocated block is INTRA type coded in base layer.

3.1.1. *INTRA mode selection method.* The coefficients of one block in base layer commonly have very similar frequency components in the enhancement layer because the sequence of base layer is generated from down-sampled enhancement layer. In this work, we make use of this feature to cut down the candidate INTRA coding modes of enhancement layer.

We use the RDC of the selected INTRA mode ($RDC_{basebest}$) as a reference value. When $RDC_{basebest}$ is a 4×4 pixels mode, it is considered that there are much high frequency components. Therefore, we can reduce the 16×16 pixels modes from candidate modes.

$$RDC_{basebest} = RDC_{base16\times16} \tag{2}$$

In the proposed algorithm, when Equation (2) is satisfied, it is considered that there are less high frequency components. Therefore, all 4×4 pixel modes are not included in the candidate modes.

$$RDC_{basebest} = RDC_{base4\times4} \tag{3}$$

On the other hand, when Equation (3) is satisfied, it is considered that there are more high frequency components. Therefore, all 16×16 pixels modes are not included in the candidate modes and only 4×4 INTRA modes and IntraBL mode are used as candidate mode.

3.1.2. *INTER mode skip method.* Generally, because there are very similar frequency components in the base layer and enhancement layer, if the base layer is encoded by INTRA type mode the enhancement layer tends to be encoded by INTRA mode. However, sometimes INTER type modes are selected in the case of there is less correlations between base layer and enhancement layer. In this work, after the best INTRA mode $RDC_{Enh\_INTRAbest}$ is selected, it is compared with the RDC of skip mode which have the lowest computation complexity.

$$RDC_{Enh\_INTRAbest} < RDC_{Enh\_skip} \tag{4}$$

When Equation (4) is satisfied, the RDC of INTRA modes had lower cost and the RDO will be terminated without performing the INTER mode selection.

On the other hand, if Equation (5) is satisfied, the RDC of INTER mode may be lower. Therefore, the INTER modes are included in the candidate mode.

$$RDC_{Enh\_INTRAbest} \geq RDC_{Enh\_skip} \tag{5}$$

3.1.3. *Flow chart for INTRA mode decision.* Figure 3 shows the flow chart for INTRA mode decision algorithm.

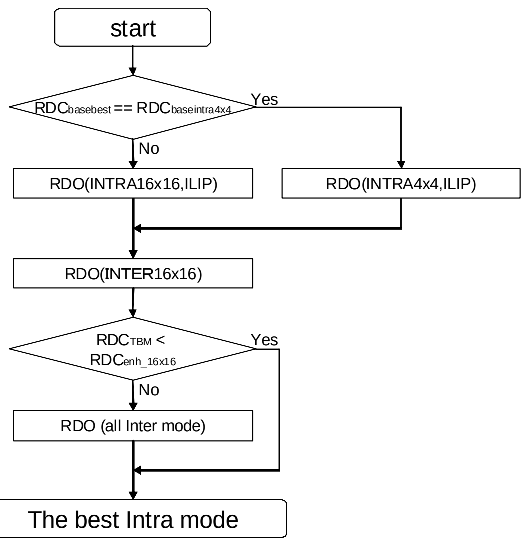

FIGURE 3. For macroblock with co-located block coded by INTRA type

As Figure 3 shows, in the first step $RDC_{basebest}$ is used to get rid of one candidate type mode between 4×4 and 16×16 pixels block. Then, the RDO for SKIP mode is performed and the RDC of SKIP mode is compared with $RDC_{Enh\_INTRAbest}$. If $RDC_{Enh\_INTRAbest}$ is smaller than that of SKIP mode, all the other INTER type modes will be excluded from the candidate modes.

3.2. **Macroblock with co-located block is INTER type coded in base layer.** Four methods are proposed for macroblock with co-located block is INTER type coded in base layer to reduce the redundant coding modes.

3.2.1. *Complexity reduction by changing the coding order for residual prediction.* The details of macroblock level processing for residual prediction are described in Figure 4 using a two layers residual prediction model. In the processing of RDO for inter-layer residual prediction mode, the residual block is up-sampled. Then, the residual block (C) which is the difference between the up-sampled block (B) and the residual block (A) in the enhancement layer are generated as the encoding macroblock. On the other hand, the RDO process for normal modes uses residual block (A) directly as the encoding macroblock. Comparing these two coding process, the computation of encoding block (A) for normal coding mode can be described as

$$Coeff(A) = Coeff(E) - Coeff(D) \tag{6}$$

and the computation of encoding block (C) for residual prediction coding can be described as

$$Coeff(C) = (Coeff(E) - Coeff(D)) - Coeff(B) \tag{7}$$

which can be also be presented as

$$Coeff(C) = (Coeff(E) - Coeff(B)) - Coeff(D) \tag{8}$$

where the Coeff represents the coefficients of macroblock. We define a new block G as

$$Coeff(G) = (Coeff(E) - Coeff(B)) \tag{9}$$

where the fixed value block G can be easily pre-calculated.

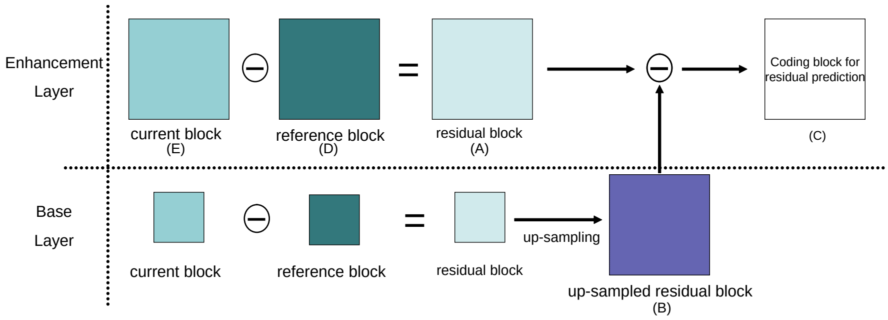

FIGURE 4. Residual prediction

It is obvious that the encoding for residual prediction can be treated as another RDO process for the block G. Together with the normal RDO process, the computation complexity double increased. Therefore, the computation complexity reduction for inter-layer residual prediction becomes the bottleneck of efficient implementation of H.264/SVC. In the next section, the proposed mode decision algorithm for H.264/SVC is presented.

Several times RDO process for traditional candidate modes in H.264/AVC and some additional new modes in H.264/SVC need to be performed. In an ordinary routine for the mode selection of H.264/SVC, the RDO for traditional modes is performed first. Then, the RDO for all modes with different block size have to be performed with residual prediction included. However, twice RDO processing for all block sizes is a redundant process which induces very high computation complexity. On the basis of many simulation results we find that the best block size selected by normal modes has deep correlation with the finally selected best mode. Utilizing this feature, we proposed a fast mode selection algorithm which can cut down the candidate modes for inter-layer residual estimation.

3.2.2. *SKIP mode decision.* When the macroblock is encoded by INTER type in base layer, there are high probability for the co-located macroblock in the enhancement layer to be encoded by INTER type. Moreover, because the resolution of enhancement layer are higher than the base layer, the predicted motion vector (PMV) tends to have higher precision to select the search center and the SKIP mode have more chances to be selected.

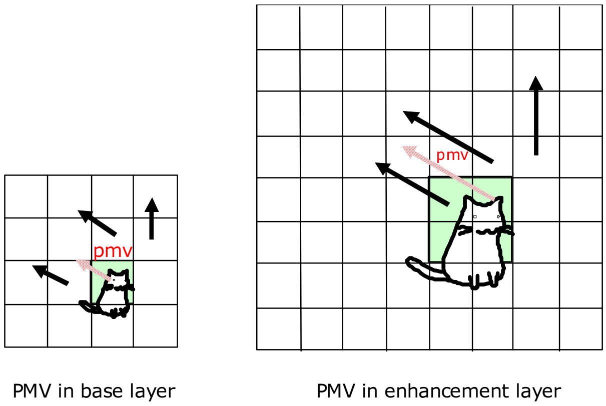

FIGURE 5. The generation of PMV

As Figure 5 shows, the PMV is generated by the adjacent MVs. Because the resolution of enhancement layer are higher than the base layer and the MV in the base layer is used as the PMV in the inter-layer motion prediction, high precision PMV can be generated in the enhancement layer.

In this work, the RDC for SKIP mode (RDC_{Enh_skip}) and 16×16 mode ($RDC_{Enh\_16\times16}$) is used to evaluate the precision of PMV. When

$$RDC_{Enh\_skip} < RDC_{Enh\_16\times16} \tag{10}$$

is satisfied, it is considered that PMV is precisely selected and the RDO for INTER mode is terminated.

In this work, RDO for three traditional modes: Skip, Direct, 16×16 modes are performed at first. If the encoding cost of skip mode or the direct mode are smaller than that of the 16×16 mode the best mode of these three modes are defined as a temporary best mode (TBM). If the 16×16 mode has the smallest encoding cost, the RDO for all the other modes have to be performed.

If Equation (10) is not satisfied, the motion estimation will not be terminated.

3.2.3. *Inter-layer residual prediction (ILRP) skip.* ILRP mode is selected when there are high correlation between base layer and enhancement layer. If the correlation can be precisely predicted, redundant candidate modes can be efficiently reduced. In this work, ΔRDC of TBM is used which can be defined as

$$\Delta RDC = \frac{|RDC_{base} - RDC_{enh}|}{RDC_{base}} \tag{11}$$

When ΔRDC is smaller than a threshold (TH), the ILRP mode is skipped. TH is set to 0.05 in this work on the basis of many simulation results. Figure 6 shows two typical simulation results.

TH = 0.05 is efficient for almost sequences. However, when there are significant motions in the sequences even there is no much correlation the RDC may be small. For this reason, another condition is used. In this work, only when both

$$RDC_{TBM} < RDC_{ILRP\_skip} \tag{12}$$

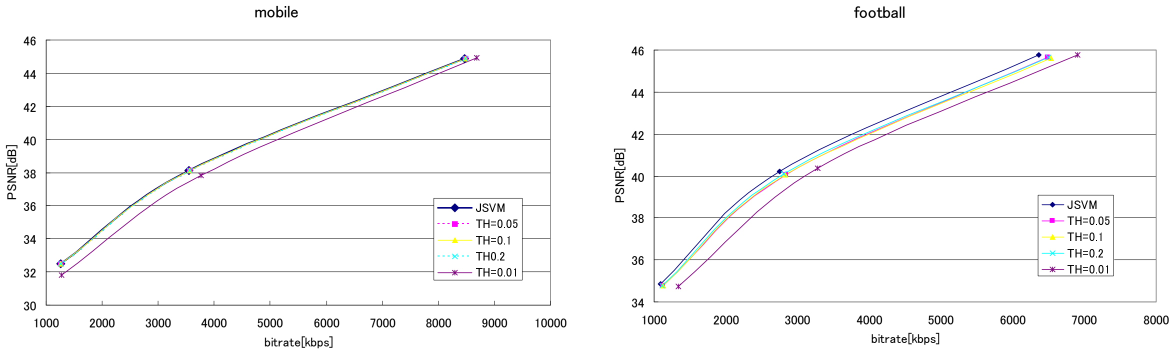

FIGURE 6. Threshold decision

and

$$\Delta RDC > TH \tag{13}$$

are satisfied ILRP mode is skipped.

3.2.4. *Candidate modes reduction for ILRP.* Due to the strong correlation between the base layer and enhancement layer, generally the same mode will be selected. Even sometimes different mode may be selected it is able to be predicted. For example, when the TBM is 16×16 mode there should have no more high frequency components in the macroblock. Therefore, it is hardly to be considered to select 4×4 mode. In this work, we further cut down the candidate modes for ILRP as shown in Table 1.

TABLE 1. ILRP candidate modes

TBM	SVC candidate mode
16×16 mode	16×16, 16×8, 8×16
16×8 mode 8×16 mode	16×16, 16×8, 8×16, 8×8 8×4, 4×8, 4×4
8×8 mode 4×8 mode 8×4 mode 4×4 mode	16×8, 8×16, 8×8 8×4, 4×8, 4×4

3.2.5. *INTRA mode skip.* When the co-located macroblock in base layer is encoded by INTER mode with a very small RDC, it has more chance to be encoded by INTER type mode. The probability can also be predicted from the base layer by using the RDC of the best mode in base layer. When

$$RDC_{base\_best} > RDC_{enh\_INTERbest} \tag{14}$$

is satisfied the INTRA mode is considered can be cut down from the candidate modes for enhancement layer.

3.2.6. *Flow chart of the proposed algorithm.* All of the four approaches are carefully considered and a efficient complexity reduction algorithm is proposed for macroblock with co-located block is INTER type coded in base layer. The flow chart is shown in Figure 7.

First, the RDO for SKIP mode and 16×16 mode are performed. If the RDC of SKIP mode is smaller than that of 16×16, all other INTER type modes is cut down from candidate modes. In the case that the RDC of SKIP mode is bigger than that of 16×16

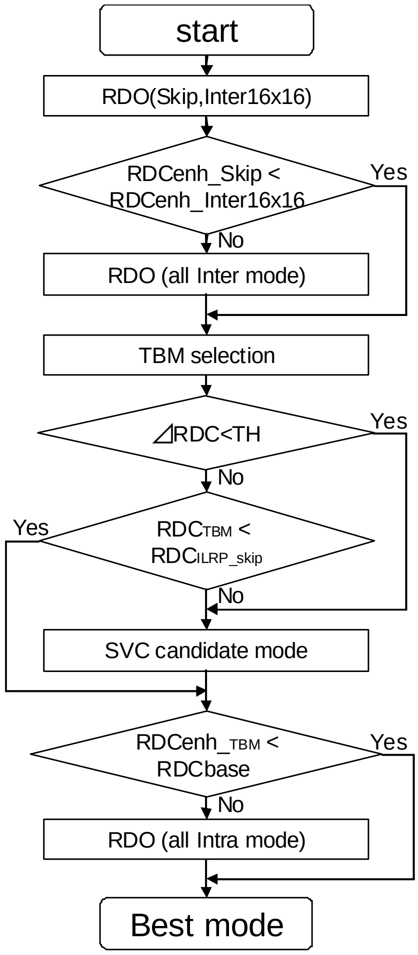

FIGURE 7. For macroblock with co-located block coded by INTER type

mode, the other INTER modes have to perform RDO. Then, the best mode is selected as TBM. Next, ΔRDC is calculated using the best mode of base layer and the TBM. If the ΔRDC is smaller than TH, a proposed SVC candidate mode which is listed in Table 1 is used. In the case when the RDC_{TBM} is smaller than RDC_{ILRP_SKIP}, all the candidate modes in Table 1 are cut down. In the end, the RDC_{base_best} and $RDC_{enh\_INTERbest}$ are compared with decide if or not it is necessary to perform RDO for INTRA modes.

4. **Simulation Results.** In order to evaluate the proposed algorithm, the proposed algorithms are implemented in the reference software JSVM [14]. The simulation conditions are concluded in Table 2. A QCIF-CIF two layers model is used without changing of the basic parameters of JSVM.

TABLE 2. Simulation conditions

Codec	JSVM 9
Frame Rate	30Hz
Frame Numbers	65
Base Layer	QCIF
Enhancement Layer	CIF
QP Setting	20 ∼ 40
GOP Size	16

Firstly, the complexity reduction rate of the proposed algorithm is compared with the previous work. The simulation results of the proposed algorithm are shown in Table 3. $\Delta Bitrate$, $\Delta Time$ and $\Delta PSNR$ in Table 3 shows the results in comparison with the JSVM.

TABLE 3. Simulation results (QP = 28)

sequence	$\Delta PSNR$ [dB]		$\Delta Bitrate$ [%]		$\Delta Time$ [%]	
	[12]	Proposed	[12]	Proposed	[12]	Proposed
mother	–0.06	0.00	0.13	0.09	–66.0	–67.4
foreman	–0.14	0.04	0.09	–1.26	–53.7	–55.2
container	–0.02	–0.01	0.21	0.03	–68.7	–71.1
tempete	–0.06	–0.03	0.71	0.62	–54.0	–57.7
silent	–0.04	0.00	0.48	0.04	–62.1	–67.8
mobile	N/A	–0.04	N/A	0.24	N/A	–50.4
football	N/A	–0.01	N/A	0.22	N/A	–45.4

As shown in Table 3, the proposed algorithm shows higher complexity reduction than the previous work. In particular, the proposed algorithm shows almost no PSNR loss which is efficient than previous one. The proposed algorithm is also compared with our previous one. The simulation results of the proposed algorithm are shown in Table 4.

TABLE 4. Simulation results (QP = 32)

sequence	$\Delta PSNR$ [dB]		$\Delta Bitrate$ [%]		$\Delta Time$ [%]	
	[13]	Proposed	[13]	Proposed	[13]	Proposed
mother	–0.00	0.00	–0.05	0.09	–67.18	–67.4
foreman	–0.03	0.04	0.18	–1.26	–35.54	–55.2
mobile	–0.04	–0.04	0.23	0.24	–35.95	–50.4
football	–0.01	–0.01	0.22	0.22	–19.43	–45.4

As Table 4 shows, the proposed algorithm can achieve improved complexity reduction rate than the previous work.

The R-D curves of some test sequences are shown in Figure 8-Figure 10.

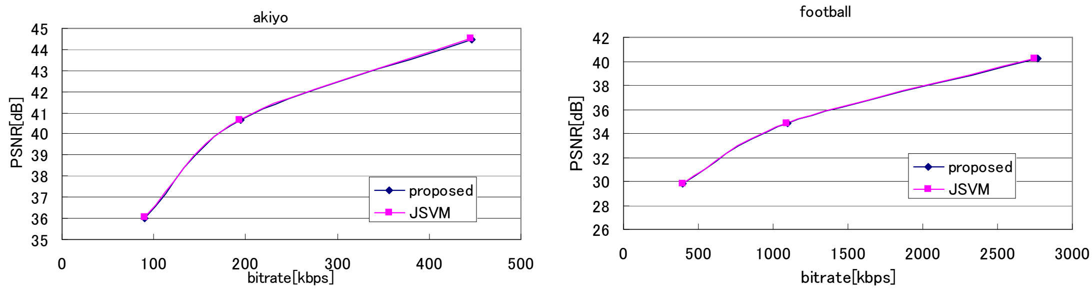

FIGURE 8. R-D curves for akiyo and football

From these R-D curves it is clear that the proposed algorithm induces almost no video quality loss. As for the PSNR and bitrate decrease, the proposed algorithm induced almost no video quality loss.

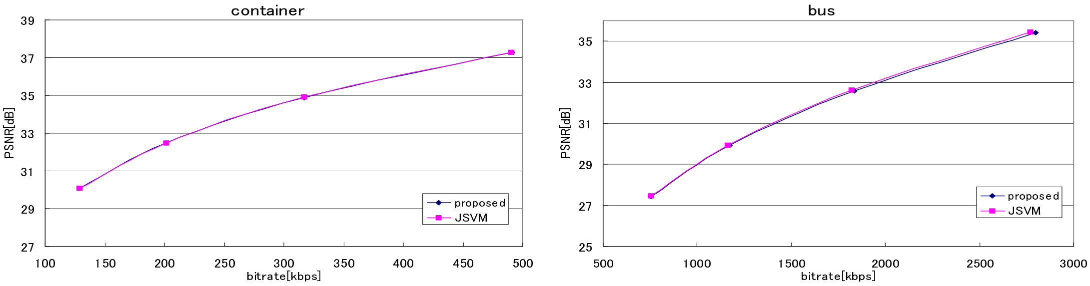

Figure 9. R-D curves for container and bus

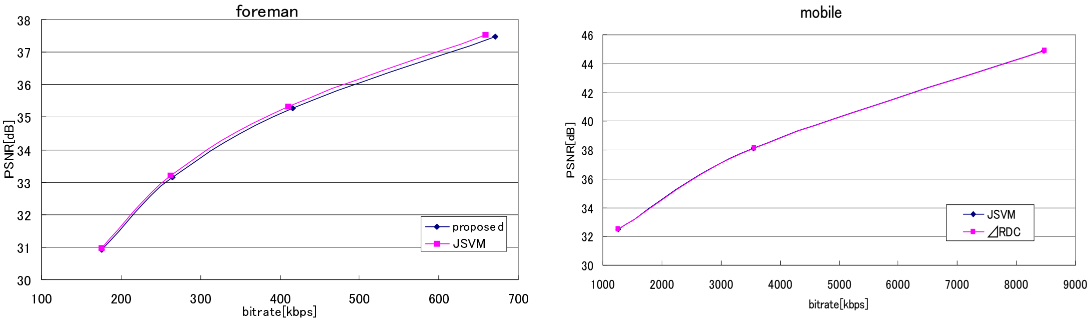

Figure 10. Foreman

5. **Conclusion.** In this paper, a low complexity algorithm for inter-layer prediction is proposed. Proposed algorithm focused on the reduction of the candidate modes by making use of the correlations of the encoding cost between the base layer and enhancement layers. For macroblock with co-located macroblock coded by INTRA or INTER type, two algorithm are proposed, respectively. Both algorithms efficiently decreased the redundant candidate modes by estimating from the base layer coding information. The experiment results show that proposed algorithm can significantly reduce redundant computation complexity with almost no coding efficiency loss.

REFERENCES

[1] J. Reichel, H. Schwarz and M. Wien, Scalable video coding-join draft 4, *ISO/IEC JTC1/SC29/WG11/JVT-Q201*, Nice, France, 2005.

[2] H. Schwarz, D. Marpe and T. Wiegand, Overview of the scalable H.264/MPEG-4 AVC extension, *Proc. of IEEE International Conference on Image Processing*, pp.161-164, 2006.

[3] H. Schwarz, D. Marpe and T. Wiegand, Overview of the scalable video coding extension of the H.264/AVC standard, *IEEE Trans. on Circuits and Systems for Video Technology*, vol.17, no.9, pp.1103-1120, 2007.

[4] H. Schwarz, D. Marpe and T. Wiegand, Joint draft ITU-T rec. H.264—ISO/IEC 14496-10/Amd.3 scalable video coding, *Doc. JVT-X201*, Geneva, 2007.

[5] C. A. Segall and G. J. Sullivan, Spatial scalability within the H.264/AVC scalable video coding extension, *IEEE Trans. on Circuits and Systems for Video Technology*, vol.17, no.9, pp.1121-1135, 2007.

[6] H. Schwarz, D. Marpe and T. Wiegand, Hierarchical B pictures, *Joint Video Team, Doc. JVT-P014*, Poznan, Poland, 2005.

[7] T. Song, S. Matsuoka, Y. Morigami and T. Shimamoto, Coding efficiency improvement with adaptive GOP selection for H.264/SVC, *International Journal of Innovative Computing, Information and Control*, vol.5, no.11(B), pp.4155-4165, 2009.

[8] T. Hamamoto, T. Katayama, T. Song and T. Shimamoto, Low complexity mode selection method for hierarchical B-picture of H.264/SVC, *ICIC Express Letters*, vol.3, no.4(A), pp.1179-1184, 2009.

[9] T. Hamamoto, T. Song, T. Katayama and T. Shimamoto, Complexity reduction algorithm for hierarchical B-picture of H.264/SVC, *International Journal of Innovative Computing, Information and Control*, vol.7, no.1, pp.445-457, 2011.
[10] H. Li, Z. G. Li and C. Y. Wen, Fast mode decision algorithm for inter-frame coding in fully scalable video coding, *IEEE Trans. on Circuit and System for Video Technology*, vol.16, no.7, 2006.
[11] S. Lim, J. Yang and B. Jeon, Fast coding mode decision for scalable video coding, *Proc. of ICACT*, pp.1897-1900, 2007.
[12] B. Lee, M. Kim, S. Hahm, C. Park and K. Park, A fast mode selection scheme in inter-layer prediction of H.264 scalable extension coding, *Proc. of IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, pp.1-5, 2008.
[13] Y. Morigami, T. Song, T. Katayama and T. Shimamoto, Low complexity algorithm for inter-layer residual prediction of H.264/AVC, *Proc. of International Conference on Image Processing*, 2009.
[14] J. Reichel, H. Schwarz and M. Wien, Joint scalable video model 9 (JSVM 9), *Joint Video Team, Doc. JVT-V202*, 2007.
[15] G. Bjontegaard, Calculation of average PSNR differences between RD-curves, *VCEG-M33*, Austin, 2001.