International Journal of Soft Computing 6 (4): 102-110, 2011
ISSN: 1816-9503

Fast Mode Decision Algorithm for Inter-Frame Coding in H.264/AVC Extended Scalable Video Coding

Byung-Gyu Kim, Gwang-Soo Hong and Ki-Wook Rim
Department of Computer Science and Engineering, Sun Moon University,
Republic of Korea

Abstract: Scalable Video Coding (SVC) is the finalized standard as an extension of H.264/AVC to provide stable multimedia services to various service environment and end-systems. Usually an exhaustive block mode search has been employed to determine the best coding mode for each Macroblock (MB). This gives a high coding efficiency, however it causes very high computational complexity in encoding system. To reduce this complexity of the exhaustive mode search, researchers propose a fast mode determination algorithm for inter-frame coding based on correlative information between base layer and enhancement layers for providing spatial scalability with Medium Grain Signal to noise ratio (MGS) layer. In the proposed algorithm, researchers use Rate-Distortion (RD) costs of BLS (Base Layer Skip) and 16×16 modes and mode information of a corresponding block of the base layer to finish inter-mode search procedure early in the neighboring enhancement layer. Also, an adaptive thresholding scheme is designed. Researchers verify that the overall encoding time can be reduced up to 66.5% based on comparative analysis of experimental results with JSVM 9.12 reference software.

Key words: Scalable video coding, spatial scalability, H.264/AVC video, inter mode, inter-layer, adaptive thresholding

INTRODUCTION

In a networking environment such as the internet, end systems can have a wide variety of requirements of the available bandwidth and computing power as kinds of devices. The communication channels comprising today's network infrastructure span a broad bandwidth range. Compressed code-streams created for one particular class of resource limitation may not be satisfactory, efficient or even useful for servicing users with different resource capacities. To handle this heterogeneity issue a highly efficient scalable video coding technique with multiple layers is required to provide stable multimedia services to various end-systems. As increasing requirement of this kind, Joint Video Team (JVT) has finalized Scalable Video Coding (SVC) as an international standard (Schwarz *et al*., 2007).

The SVC is currently being developed as an extension of H.264/Advanced Video Coding (H.264/AVC) (Wiegand *et al*., 2003). Compared to the previous video coding standards, SVC is intended to encode the signal once but enable decoding from partial streams depending on the specific rate and resolution required by a certain application. In the current status of the SVC standard a coded bitstream is composed of a base layer and several enhancement layers (Schwarz *et al*., 2007). The base layer contains conventionally a reduced resolution or a reduced quality version of each coded frame for mobile devices such as portable phones, PDAs and smart phones having low computing power.

But this base layer can provide high quality and resolution video as applications. The upper enhancement layers are used to provide a higher quality service for PSNR, frame-rate and image resolution. For generation of the desired frame-rate, temporal decomposition of the structure of hierarchical B pictures has been widely used as shown in Fig. 1 (Schwarz *et al*., 2007) because of its simplicity and effectiveness.

In SVC, variable block size for intra and inter mode prediction maximizes the coding efficiency based on Rate-distortion Optimization (RDO) because of an extension of H.264/AVC (Wiegand *et al*., 2003). The block sizes are MODE SKIP, 16×16, 16×8, 8×16, 8×8, 8×4, 4×8 and 4×4 for inter-frame coding. Intra mode prediction (Nine modes for a 4×4 luma block and 8×8 luma block and four modes for 16×16 luma and I PCM blocks) follows inter mode prediction to determine the best residual image (Wiegand *et al*., 2003).

Corresponding Author: Byung-Gyu Kim, Department of Computer Science and Engineering, Sun Moon University, Republic of Korea

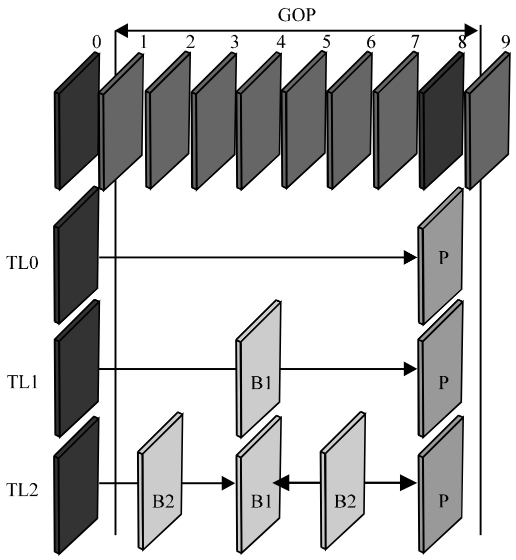

Fig. 1: The structure of hierarchical B picture for temporal decomposition when the size of a GOP is eight in SVC

Especially, INTRA BL (Intra Base Layer) has been added to improve the coding efficiency in the enhancement layers, resulting in an increase in the computational complexity. Additionally, there is a Base Layer Skip (BLS) mode that can be initiated by a base mode flag in the enhancement layer to directly interpolate all the motion vectors and residual values from the base layer without any further search. Motion-compensated prediction and intra-prediction are used for spatial scalability in single-layer coding in each spatial layer. However, in order to improve the coding efficiency for simulcasting of different spatial resolutions, additional inter-layer prediction mechanisms are incorporated (Schwarz *et al.*, 2007). The three additional inter layer prediction concepts (Schwarz *et al.*, 2007) that have been added in SVC are:

- Prediction of macroblock modes and associated motion parameters
- Prediction of the residual signal
- Prediction of the intra texture

Similar to H.264/AVC for an MB, SVC calculates the RD cost of every possible mode and selects the one with minimum RD cost as the best mode as the following (Wiegand *et al.*, 2003):

$$J(s, c, \text{MODE} \mid \text{QP}) = \text{SSD}(s,c, \text{MODEQP}) + \lambda_{\text{MODE}} \cdot R(s,c,\text{MODE} \mid \text{QP}) \quad (1)$$

Where:

QP = Quantization parameter
λ_{MODE} = Lagrangian multiplier
SSD = Sum of the squared differences between the original block and its reconstruction

R (s, c, MODE|QP) is the number of bits associated with the mode currently selected for the MB. Although, the inter-layer prediction mechanisms have been developed, calculation of the RD cost in SVC needs to execute both the forward and backward processes of integer transform, quantization, inverse quantization, inverse integer transform and entropy coding and this causes a high computational complexity to the encoder. Therefore, it is necessary to design a method which can reduce this complexity with minimal loss of image quality. Recently many kinds of fast mode decision schemes have been reported with the Rate-distortion (R-D) optimization (Wu *et al.*, 2004; Kim, 2008) for H.264/AVC encoding system. Wu *et al.* (2004) make use of the spatial homogeneity of a video object's textures and the temporal stationary characteristics inherent in video sequences. However, the method suffers from a drawback to obtain the edge image for textual information and a difference image for the temporal stationarity characteristics. Zhu *et al.* (2004) (Kuo *et al.*, 2006) proposed another approach for a fast inter mode decision that uses a pre-encoding, down-sampled image space. After obtaining candidate block modes, a refinement search is performed to find the best mode in the original image space. To detect an early SKIP block, Crecos and Yang (2006) also used the prediction and thresholding scheme. Yu *et al.* (2008) suggested a method using a hierarchical structure comprising three levels. Each level targets different types of macroblocks according to the complexity of the search process.

Also, an efficient algorithm based on macroblock tracking scheme has been reported by Kim (2008) and Hilmi *et al.* (2010). In this algorithm, researcher used the most correlated MB of the reference picture to design an early termination rule. Even though, the described algorithms are efficient in reducing the computational complexity with negligible quality degradation in H.264/AVC encoder but not applicable to the enhancement layers of an SVC encoding system. Moreover, researchers can utilize an inter-layer information between successive layers in SVC encoding system. There are very few studies although, it plays a very important role to reduce the overall complexity of SVC. Li *et al.* (2006a, b) proposed a fast mode decision algorithm for spatial scalability in SVC. In this scheme, they used the mode distribution relationship between base layer and enhancement layers (Li *et al*., 2006a). In (Li *et al*., 2006b), a fast mode decision algorithm has been proposed for inter-frame coding supporting spatial, coarse grain signal-to-noise ratio and temporal scalability by Li. It makes use of the mode-distribution correlation between the base layer and enhancement layers. After the exhaustive search technique is performed at the base layer, the candidate modes for enhancement layers can be reduced to a small number based on the correlation. A layer-adaptive mode decision algorithm and a motion search scheme have been suggested for CGS and temporal scalability by Lin *et al*. (2007). To reduce the mode search, they skipped modes with limited contributions to the coding efficiency based on statistical analysis.

To speed-up the motion search, they reused the reference frame indices of the base layer and determined the initial search points using the motion vector at the base layer. This scheme is not suitable for spatial scalability. Researchers propose a fast mode determination algorithm for inter-frame coding supporting spatial scalability based on correlative information between base layer and enhancement layers. In the proposed algorithm, researchers use Rate-distortion (RD) costs of BL SKIP and 16×16 modes and mode information of a corresponding block of the base layer to finish the inter-mode search process early in the neighboring enhancement layer. Researchers introduce a thresholding scheme for this purpose, too.

PROPOSED FAST MODE DECISION ALGORITHM

Observations on correlation between neighboring layers (Base and enhancement layers)

Spatial scalability: In scalable video coding as explained above the encoded video bit stream is in layers which each layer is corresponding to a particular resolution or a particular QP or both along with this temporal scalability. Although, the layers are contained different video bit streams but being encoded is inherently the same procedure. This is the first reason for correlation between layers although, the extent of correlation depends on QP, detail of texture of the used picture frame and different resolution. In case of spatial scalability under ideal condition and same QP, the modes may be highly in correlation with each other according to the scale of resolution between base and enhancement layers. For example if the base layer is QCIF and enhancement layer is of CIF resolution then the base layer mode of macro block is 8×8 should be 16×16 in the enhancement layer. Similarly, 4×4 in base layer should be 8×8 in the enhancement layer. Even though, QP values are different in neighboring layers, researchers observed that this correlation is still very high.

Table 1: Statistical data for correlation between successive layers of spatial scalability with MGS SNR layers (Unit: %)

Sequences	Foreman	Bus	Football
Skip (MB_Direct)			
Skip (MB_Direct)	76.84	68.90	81.13
MB16×16	13.94	22.62	14.91
MB16×8	1.77	3.72	1.47
MB8×16	2.60	3.52	1.77
MB8×8	0.95	2.87	0.70
MB16×16			
SKIP(MB_Direct)	63.45	46.25	52.71
MB16×16	22.77	32.16	26.41
MB16×8	5.17	6.77	8.87
MB8×16	5.37	6.69	6.31
MB8×8	3.21	8.11	4.61
MB16×8			
SKIP(MB_Direct)	36.84	27.57	19.22
MB16×16	30.49	33.98	30.79
MB16×8	14.88	16.55	24.15
MB8×16	7.50	8.38	15.07
MB8×8	10.27	13.49	10.70
MB8×16			
SKIP(MB_Direct)	44.50	26.86	14.96
MB16×16	27.02	34.04	30.90
MB16×8	7.01	10.94	13.72
MB8×16	16.54	15.54	28.75
MB8×8	4.91	12.56	12.14
MB8×8			
SKIP(MB_Direct)	28.48	18.60	8.75
MB16×16	26.84	21.49	21.78
MB16×8	16.93	17.38	20.25
MB8×16	12.11	14.81	24.21
MB8×8	15.62	27.69	24.85

In order to determine the correlation, researchers used different quantization parameter and statistical analysis was performed. Results are shown in Table 1. For this analytical data, researchers considered five inter modes such as SKIP, P16×16, P16×8, P8×16 and P8×8 except intra mode. With SKIP (MB_Direct) mode for the corresponding MB in the base layer, the probability that the current MB will be SKIP or P16×16 mode is almost 90%. The probability that the MB will be SKIP, P16×16, P16×8 and P8×16 is >95%. There is a high probability for the MB to be SKIP, P16×16, P16×8 and P8×16 if the mode of the co-located MB is SKIP in lower layer. When the co-located MB is P16×16 mode, the probability to be SKIP and P16×16 is less than the SKIP co-located MB. This probability of about 80% is still valuable information for evaluation of a fast algorithm. If the mode of the co-located MB is P16×8 or P8×16 in lower layer, the probability to be each mode goes to under almost 60%. From this viewpoint, there is a high probability for the MB to be SKIP and P16×16 modes if the mode of the co-located MB or corresponding MB is SKIP or P16×16 in the neighboring layer. This means that the extent of correlation between different layers can be estimated.

RD cost of Base Layer Skip (BLS) and 16×16 modes: There exists a mode called base layer skip mode in which all the information that has been obtained for the corresponding macro block in base layer when is signalled by a separate syntax element (Residual prediction flag). This mode is interpolated and then used to encode the current macro block in enhancement layer. The rate distortion cost corresponding to this particular mode is of particular interest. In this particular all the residual, motion vectors and even base layer mode are used to encode the present macro block. Motion vectors are interpolated in case of spatial scalability and directly used in case of CGS scalability similar is the case with residual where as base layer mode is used as the mode for current macro block. The BL SKIP mode amounts to minimum cost in terms of encoding hence, information of the rate distortion cost for this particular mode can give a good estimate of SAD (Sum of absolute differences).

The inference is that the rate distortion cost can be used to estimate the compatibility of that par ticular mode for the current macro block. For example if the base layer mode is 16×16 which is at 1/4 pixel motion estimation in enhancement layer it becomes an estimate of 1/2 pixel estimation of 16×16 mode or in case of CGS scalability 1/4 pixel estimation at different quantization parameter. In other words the rate distortion cost of base layer skip mode gives us an estimate by which researchers can determine the base layer mode for a given macro block. In order to estimate the mode based on RD cost of BL SKIP mode, researchers try to estimate the average difference between RD cost of base layer skip mode and best possible RD cost of that particular mode.

The average is taken and updated at the start of every new frame. As the average is added to the best RD cost of preceding macro block to form an adaptive thresholding scheme based on mode of base layer, researcher make a decision of whether estimating or by passing a particular mode in the enhancement layer.

Fast mode decision algorithm: Similar to H.264/AVC, there are seven macro block modes which have been discussed in the previous but as researchers observed from the statistics the distribution of modes with respect to number is not even. For example generally for all sequences for any practical quantization parameter the number of macro blocks which are classified as 16×16 and SKIP modes are maximum averaging generally of 80% of the total macro blocks when combined together. As researchers can see if researchers could limit the mode estimation for these macro blocks i.e., if we can exclude other modes which are less likely for a given macro block then lot of time could be saved in this process.

Researchers make the difference between RD costs of BL SKIP mode and 16×16 mode for the detection of complex or detailed areas of frames. So that researchers can apply the mode information from base layer along with this data to eliminate the candidate modes for mode estimation in enhancement layer. Researcher observe that the absolute difference between RD costs of BL SKIP mode and 16×16 mode tends to be on the higher side in case of macroblocks which belong to more detailed areas of the given frame. As the difference value increases, the probability of occurrence of smaller block types like 16×8, 8×16, 8×8 increases in that order. The proposed algorithm for achieving this by using the combination of mode correlation between base and enhancement layer and the adaptive thresholds discussed in the above section as follows:

- In initialization, for the first inter-frame in enhancement layer, researchers set a large value of two threshold values (Th1 = K* AvgDiff and Th2 = 3K* AvgDiff in Fig. 2 to perform full mode search and researchers estimate the average difference value between RD costs of the BL SKIP and 16×16 mode for updating the average difference (AvgDiff) in this frame. Go to the next step. The AvgDiff is defined for the next frame n+1 as:

$$\mathrm{AvgDiff}[n+1] = \left(\begin{array}{c} \mathrm{AvgDiff}[n-I]+ \\ \dfrac{\sum_{i=0}^{M-1} \left| \mathrm{RDC}^{I}_{\mathrm{BLSKIP}}[n] - \mathrm{RDC}^{I}_{16\times16}[n] \right|}{M} \end{array} \right) \times \frac{1}{2} \qquad (2)$$

Where:
M = Total number of macroblocks
I = Index of macroblock in the current frame n

- If mode of base layer is 16×16 or 16×8 or 8×16 and Th1 is less than the difference RD cost value between BL SKIP and 16×16 modes for the current MB, researchers add 16×8 and 8×16 as candidate mode for computing RD cost and go to step 3. Otherwise, go to step 3 without further mode search
- If mode of base layer is 16×8 or 8×16 or 8×8 and Th2 is less than the difference RD cost value between BL SKIP and 16×16 modes for the current MB, researchers add 8×8 sub-block types for further search and go to step 4. Otherwise, go to step 4 directly
- Decision for the best mode is that when researchers select a mode that has a minimum RD cost among the checked modes and update the average difference value between RD costs if the best mode is BLSKIP or 16×16 block type. Then, go to step 2 for processing the next MB repeatedly until the last MB is examined

Figure 2 shows the overall flow of the suggested algorithm. In the study, average difference is multiplied by the factor 3 when comparing to the mode 8×8, since 8×8 mode only occurs in case of large deviations from the mean value.

Value of the constant K can be variable. If researchers require more time saving, we can set larger value of K and vice versa. In this way, a trade-off can be achieved between time saving factor and required quality. The main contributions of the proposed algorithm can be summarized as follows:

- Based on correlative information, the proposed algorithm employs the low detailed mode types for the detection of complex or more detailed areas of frames. Especially, the RD cost values of BLS (Base Layer Skip) and 16×16 modes are used to decide whether the current MB is more complex or more detailed one

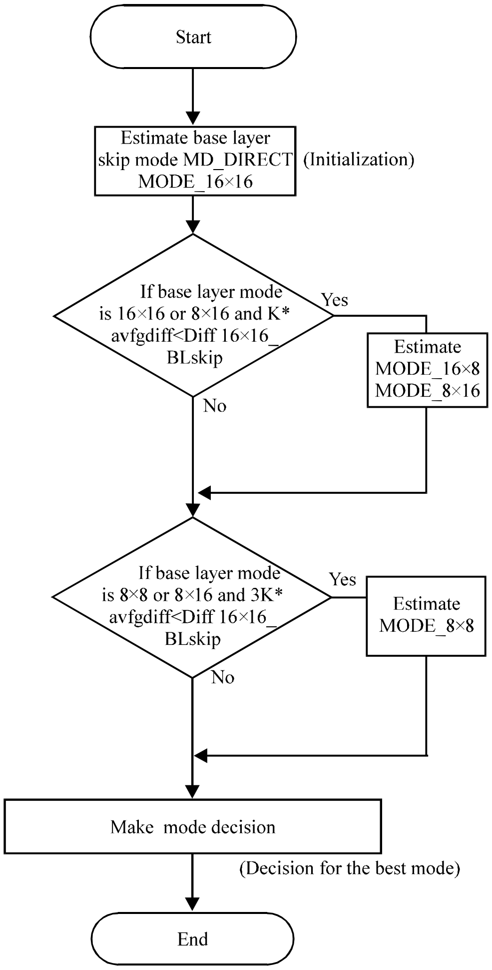

Fig. 2: Overall flow of the proposed algorithm

- Researchers have designed two-step thresholding scheme as complexity increases. The first threshold value (Th1) is used to check whether the current MB needs to examine rectangular block types (16×8 and 16×8). The second one (Th2) is for checking on whether it needs more detailed search (8×8 sub-mode types)
- With the designed threshold values Th1 and Th2, researchers can control trade-off between time saving factor and required quality

EXPERIMENTAL RESULTS

Various standard sequences were used to verify the performance of the proposed fast mode decision algorithm for inter-frame coding. The test condition is shown in Table 2 for encoding SVC video with two layers (Base layer and 1 enhancement layer). JSVM 9.12 reference software of the JVT (Joint Video Team) was used as a reference code for evaluation of the encoding performance. Researchers compiled this JSVM software on Visual C++ 6.0 with Win32 Release mode. All algorithms for comparison were run on an HW platform of a Pentium-4 PC with a 3.0 GHz quad-core CPU and 1 GByte of RAM. Researchers defined three measures for evaluating the encoding performance including average ΔPSNR, average ΔBitrate and an encoding-time saving factor, ΔT. The average ΔPSNR is the difference in (dB) between the average PSNR of the proposed method and the corresponding value of another method. The average PSNR is defined as:

$$\overline{\text{PSNR}} = \frac{4.\text{PSNR}_Y + \text{PSNR}_{Cb} + \text{PSNR}_{Cr}}{4} \quad (3)$$

where, PSNR_Y PSNR_{Cb} and PSNR_{Cr} are peak to noise ratios of the luminance and two chrominance components,

Table 2: Simulation condition

Parameters	All tested video sequences
QP setting	
Base	20, 30, 40
Enhancement	20, 30, 40
Resolution	
Base	QCIF
Enhancement	CIF
Frame rate	
Base	15 fps
Enhancement	15 fps
Coding option used	SNR layers: MGS Frames = 200, RD opt. enabled ME: Full search SR: 32, No. of reference = 1 MV resolution = 1/4 pel Coding structure: IBBB BP
Codec	JSVM 9.12

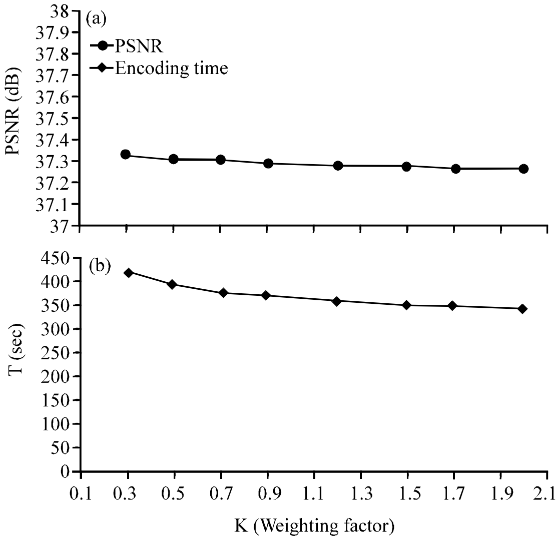

Fig. 3: Analysis graphs for the soccer sequence; a) K vs. PSNR and b) K vs. T (Encoding time)

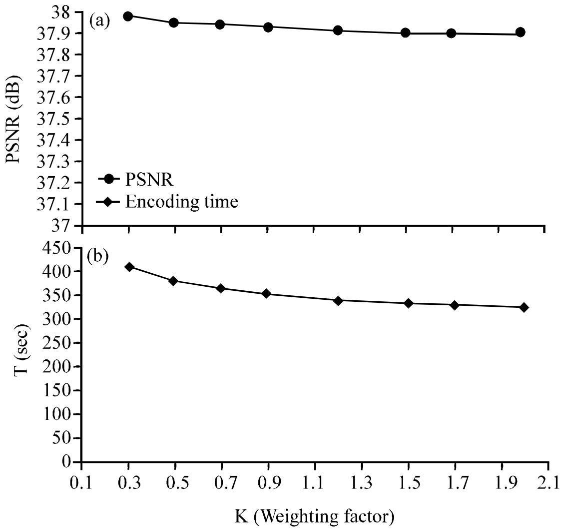

Fig. 4: Analysis graphs for the Foreman sequence; a) K vs. PSNR and b) K vs. T (Encoding time)

respectively. As performance improves, this criterion becomes larger. The average ΔBits is the bit-rate difference as a percentage between compared methods. Lastly, the encoding-time saving factor ΔT is defined for a complexity comparison as:

$$\Delta\text{Time} = \frac{\Delta\text{Time}_{\text{Proposed}} - \Delta\text{Time}_{\text{Original}}}{\Delta\text{Time}_{\text{Original}}} \times 100\% \qquad (4)$$

under the condition that the Full Mode Search (FMS) is optimum (Reference). As this value increases, the performance speed is increased. The required encoding time is measured by JSVM reference software and was defined as the total encoding time including IDR slices, motion vector searches, and transforms. It must be noted that positive values for the $\overline{\text{PSNR}}$ and ΔBitrate indicate increments and negative values indicate decrements.

Researchers used Li's method for an objective comparison of the encoding performance (Li *et al.*, 2006a, b). This is well-known as simple and fast algorithm for providing spatial scalability in SVC. Researchers have tested the algorithm for determining a constant K with some sequences. Among the results, those of the Foreman and Soccer sequences have been shown in Fig. 3 and 4 as variation of K. As researchers mentioned before, there is a trade-off between time saving factor and required quality according to variation of the constant K in the proposed algorithm. In Fig. 3, the reduction of the encoding time becomes very small and stationary value when the degradation of quality is almost zero from K = 0.7 or 0.9. For the Foreman, researchers can see similar phenomenon as shown in Fig. 4. Based on this analysis, researchers set as K = 0.7 in the experimental analysis. Figure 5 and 6 shown the Rate-distortion (RD) curves for four sequences. The proposed mode decision algorithm exhibited an RDO performance similar to the JSVM original encoder with the full intra mode search. When Li's algorithm employed, a large loss of quality of approximately 0.1~0.24 (dB) for the Foreman and city sequences occurred comparing to the original encoder. This is undesirable from the viewpoints of image quality and network bandwidth. However, the proposed algorithm produced almost the same performance in overall bit rate. Moreover in very low bitrate area, the proposed algorithm yielded better RDO performance because of a lot of bit saving effect of the method.

For the Bus sequence, researchers can see there is a little loss of PSNR in the low bitrate range when Li's algorithm applied. In higher bitrate range, the proposed and Li's methods are very similar to the full mode search method. In the Harbour sequence, a similar performance was observed over the range of about 600 kbps.

For the Foreman sequence, the algorithm showed better performance than Li's algorithm in overall bitrate range. From these results although, researchers can observe that the algorithm produced similar performance to the full mode search method an improvement of up to 65% was obtained for the speed-up gain of the encoding system.

Table 3 shows the results for all algorithms using only one reference frame when QP = 20, 30 and 40 for base and enhancement layers. In Table 3, QP_B means a QP value of the base layer and QP_E means that

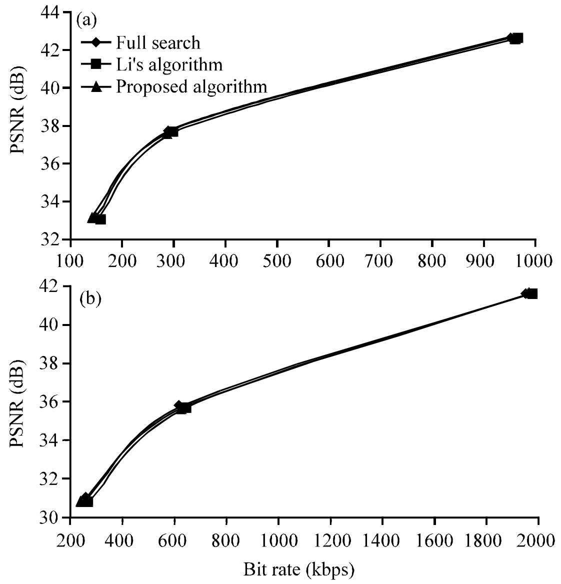

Fig. 5: Rate-distortion curves for; a) City and b) Bus sequences

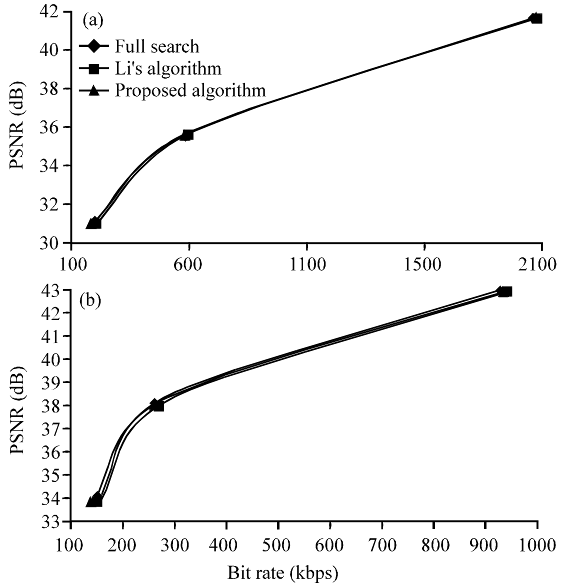

Fig. 6: Rate-distortion curves for; a) Harbour and b) Foreman sequences

Table 3: Performance comparison of the proposed algorithm for spatial scalability

			Performance parameters		
Contents	QP values	Test methods	ΔPSNR	ΔBits	ΔT
City	QP_B = 20	Li's method	-0.02	2.007	48.43
	QP_E = 20	Proposed method	-0.01	0.725	53.18
	QP_B = 20	Li's method	-0.07	3.989	54.10
	QP_E = 20	Proposed method	-0.07	0.306	66.23
	QP_B = 20	Li's method	-0.12	2.184	53.29
	QP_E = 20	Proposed method	-0.12	-8.732	65.80
Foreman	QP_B = 20	Li's method	-0.02	1.786	38.63
	QP_E = 20	Proposed method	0.00	0.536	44.93
	QP_B = 20	Li's method	-0.12	2.342	45.98
	QP_E = 20	Proposed method	-0.08	-0.425	59.63
	QP_B = 20	Li's method	-0.25	-0.631	49.46
	QP_E = 20	Proposed method	-0.22	-7.644	64.30
Bus	QP_B = 20	Li's method	-0.01	1.032	31.43
	QP_E = 20	Proposed method	0.00	0.528	39.40
	QP_B = 20	Li's method	-0.05	2.688	38.93
	QP_E = 20	Proposed method	-0.06	0.660	55.49
	QP_B = 20	Li's method	-0.13	3.976	47.05
	QP_E = 20	Proposed method	-0.19	-7.089	65.92
Football	QP_B = 20	Li's method	-0.01	0.326	30.06
	QP_E = 20	Proposed method	0.00	0.062	36.66
	QP_B = 20	Li's method	-0.06	0.883	36.96
	QP_E = 20	Proposed method	-0.01	0.078	48.71
	QP_B = 20	Li's method	-0.23	-1.579	51.66
	QP_E = 20	Proposed method	-0.11	-3.218	61.47
Harbour	QP_B = 20	Li's method	0.00	0.314	33.95
	QP_E = 20	Proposed method	0.00	0.155	36.07
	QP_B = 20	Li's method	-0.03	1.820	46.79
	QP_E = 20	Proposed method	-0.04	-0.092	56.08
	QP_B = 20	Li's method	-0.10	1.512	54.51
	QP_E = 20	Proposed method	-0.10	-8.734	66.51
Soccer	QP_B = 20	Li's method	-0.02	1.011	39.95
	QP_E = 20	Proposed method	0.00	0.330	43.09
	QP_B = 20	Li's method	-0.06	2.156	46.61
	QP_E = 20	Proposed method	-0.06	0.064	56.81
	QP_B = 20	Li's method	-0.18	0.997	51.47
	QP_E = 20	Proposed method	-0.18	-6.231	64.88

of neighboring enhancement layer, respectively. The proposed algorithm achieved a better speed-up factor with a minimal loss of image quality and a lot of bit decrement compared with result for Li's method. In sequences with stationary or slow motion such as the city, Li's method yielded a smaller speed-up factor by almost 10% than the proposed algorithm. There was a negligible loss of quality and a large bitrate saving fastor by about 8.73% at higher QP (QP_B = 40 and QP_E = 40). In all other sequences, researchers could observe a similar bitrate saving effect in higher QP value.

Researchers can see that the suggested algorithm is superior to Li's method for sequences with fast object's motion or a large global motion (Bus, football and soccer), too. The proposed algorithm exhibited similar improvements in the encoding speed for the city and Foreman sequences. For sequences with fast object motion or a large global motion (Bus, football and soccer), the algorithm was faster than Li's method by a factor of up to 14% for the average encoding time. For QP_B = 40 and QP_E = 40 with the soccer sequence, the encoding speed was improved up to 64.88% with a bitrate decrement of 6.231%. Judging from total average values, the proposed algorithm produced a speed-up factor of >11% with a minimal loss of image quality and more bitrate-saving comparing to Li's method.

Table 4: The average performance for all QP values

		Performance parameters		
Contents	Test methods	ΔPSNR	ΔBits	ΔT
City	Li's method	-0.07	2.726	51.94
	Proposed method	-0.06	-2.567	61.73
Foreman	Li's method	-0.13	1.165	39.13
	Proposed method	-0.10	-2.511	56.19
Bus	Li's method	-0.06	2.565	39.13
	Proposed method	-0.08	-1.967	53.60
Football	Li's method	-0.10	-0.123	39.56
	Proposed method	-0.04	-1.026	48.94
Harbour	Li's method	-0.04	1.215	45.08
	Proposed method	-0.04	-2.890	52.88
Soccer	Li's method	-0.08	1.388	46.01
	Proposed method	-0.08	-1.945	54.92
Total	Li's method	-0.08	1.489	43.47
Average value	Proposed method	-0.06	-2.510	54.71

Table 5: The hit-probability of the predicted mode using the proposed algorithm for all QP values

Sequences	Hit-probability (%)
Foreman	92.97
Bus	85.95
City	94.47
Football	94.67
Harbour	97.75
Soccer	93.61

The average results for all algorithms are shown in Table 4. With Li's method, a speed-up gain of almost 43% is achieved while almost 54.92% gain is achieved using the algorithm (Table 4). In terms of image quality, the proposed algorithm is similar to Li's method. However, the proposed algorithm shows more bitrate-saving factor compared with Li's method from viewpoint of bitrate increment. This means the suggested algorithm is more efficient than Li's even though similar image quality and less bits in the given network environment.

Table 5 shows the result of the hit-probability using the proposed algorithm. The hit-probability is defined as the probability of selecting true solution as the block mode m when the algorithm predicts mode m as the best mode:

$$P(\text{Best Mode}_{\text{true}} = m \mid \text{Best Mode}_{\text{pred}} = m) \qquad (5)$$

where, m is one among the possible inter modes. The average hit-probability is almost 93%. For the Bus sequence, researchers achieved about 85% of the hit-probability due to much edges with very fast motion. In other sequences, the proposed fast mode decision algorithm shows the hit-probability of >92%. From this result, researchers can deduce the suggested algorithm is able to provide a large speed-up factor with a negligible quality loss and bitrate increment.

CONCLUSION

Researchers have proposed efficient mode decision algorithm based on Rate-distortion (RD) costs of BLS (Base Layer Skip) and 16×16 modes and mode information of a corresponding block of the base layer for spatial scalability of SVC. The proposed algorithm yield good performance because of adaptive RD thresholding scheme using average difference value of the RD costs between BL SKIP and 16×16 modes for the current MB. Based on comparative analysis, a speed-up factor of 36.07~66.51% was verified with a negligible bitrate increment or large bitrate saving and a minimal loss of image quality. Moreover if a fast mode decision for a single-layer (Base layer) with negligible loss of quality will be applied then, researchers can obtain higher speed-up gain in the overall encoding system.

ACKNOWLEDGEMENTS

This research was supported by the MKE (The Ministry of Knowledge Economy) Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National IT Industry Promotion Agency) (NIPA-2011-C1090-1131-0004).

REFERENCES

Crecos, C. and M.Y. Yang, 2006. Fast inter mode prediction for p slices in the H264 video coding standard. IEEE Trans. Broadcasting, 51: 256 -263.

Hilmi, B., C.S. Park and B.G. Kim, 2010. A survey of fast inter mode decision algorithms for H.264/AVC video encoding system. Int. J. Soft Comput., 5: 128-148.

Kim, B.G., 2008. Novel inter-mode decision algorithm ased on macroblock (MB) tracking for the p-slice in H.264/AVC video coding. IEEE Trans. Circuit Syst. Video Technol., 18: 273 -279.

Kuo, C.H., M. Shen and C.C.J. Kuo, 2006. Fast inter-prediction mode decision and motion search for H.264. Proc. IEEE Int. Conf. Multimedia Exp., 1: 663-666.

Li, H., Z.G. Li and C. Wen, 2006a. Fast mode decision algorithm for inter-frame coding in fully scalable video coding. IEEE Trans. Circuit Syst. Video Technol., 16: 889 -895.

Li, H., Z.G. Li, C. Wen and L.P. Chau, 2006b. Fast mode decision for spatial scalabiltiy video coding. Proceedings of IEEE International Symposium on Circuits and Systems, May 2006, Kos, Greece, pp: 3005-3008.

Lin, H.C., H.W. Peng, M.H. Hang and W.J. Ho, 2007. Layer-adaptive mode decision and motion search for scalable video coding with combined Coarse Granular Scalability (CGS) and temporal scalability. Proc. IEEE Int. Conf. Image Process., 2: 289-292.

Schwarz, H., D. Marpe and T. Wiegand, 2007. Overview of the scalable video coding extension of the H.264/AVC standard. IEEE Trans. Circuit Syst. Video Technol., 17: 1103-1120.

Wiegand, T., G.J. Sullivan, G. Bjontegard and A. Luthra, 2003. Overview of the H.264/AVC video coding standard. IEEE. Trans. Circuits Syst. Video Technol., 13: 560-576.

Wu, D., S. Wu, K.P. Lim, F. Pan, Z.G. Li and X. Lin, 2004. Block inter mode decision for fast encoding of H.264. Proc. IEEE Int. Conf. Acostics Speech Signal Process., 3: 181-184.

Yu, A.C.W., G.R. Martin and H. Park, 2008. Fast inter-mode selection in the H.264/AVC standard using a hierarchical decision process. IEEE Trans. Circuit Syst. Video Technol., 18: 186 -195.

Zhu, D., Q. Dai and R. Ding, 2004. Fast inter prediction mode decision for H.264. Proc. IEEE Int. Conf. Multimedia Exp., 1: 1123-1126.