Fast Mode Decision Using Rate-Distortion Cost and Temporal Correlations in H.264/AVC

Yo-Sung Ho and Soo-Jin Hwang

Abstract In this paper, we propose a fast mode decision scheme in the H.264/AVC standard. The encoder uses variable block sizes to select the optimal best mode incurring the high computational complexity. Thus, we propose the fast mode decision algorithm to reduce the complexity. Using the rate-distortion cost and temporal correlations, we propose the additional early SKIP and inactive inter/intra mode decision methods in P frames. Experimental results show the proposed method provides about 60% encoding time savings on average without significant performance loss, compared to the H.264/AVC fast high complexity mode. Specifically, we can save the time about 22% and 39% in the intra and inter mode conditions, respectively.

Keywords H.264/AVC • Fast mode decision • Rate-distortion costs (RD cost) correlation for optimal and sub-optimal modes • Temporal correlation

1 Introduction

H.264/AVC is the latest international video coding standard developed by JVT (Joint Video Team) which is the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG) [1].

The standard achieves higher compression efficiency than the previous video coding standards with the rate-distortion optimized (RDO) technique for the mode decision. As shown in Fig. 1, H.264/AVC has several candidate modes: SKIP, 16 × 16, 16 × 8, 8 × 16, P8 × 8, intra 4 × 4, and intra 16 × 16 [2]. In the technique, by searching all combinations of the modes for each macroblock exhaustively, we

Y.-S. Ho (✉) • S.-J. Hwang
Gwangju Institute of Science and Technology (GIST), 261 Cheomdan-gwagiro, Buk-gu, Gwangju 500-712, Republic of Korea
e-mail: hoyo@gist.ac.kr; soojin@gist.ac.kr

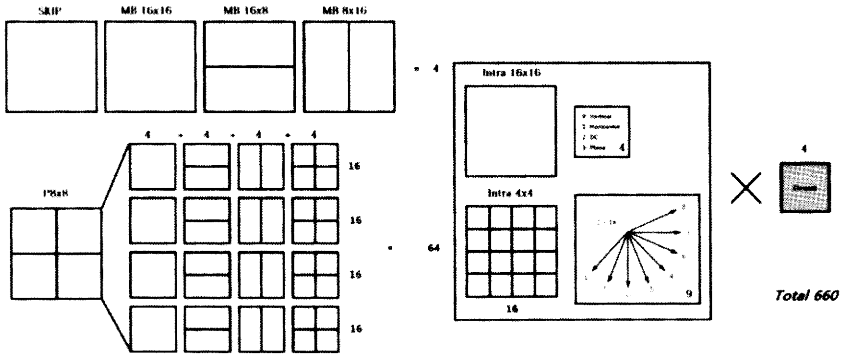

Fig. 1 The maximum number of computing RD cost in the mode decision

can achieve the optimal best coding quality while minimizing the bit rate. However, Fig. 1 more specifically shows that the RDO technique increases complexity and computation load drastically; it has to be calculated 660 times (inter mode: 68 and intra mode: 592) for each macroblock. Since the technique makes H.264/AVC unsuitable for real-time applications, the encoding time reduction is still an issue.

Most fast mode decision algorithms were already proposed using the correlations of spatial, temporal, RD cost, and so on. However, most proposed methods carry out experiments on only simple sequences or are used their own internal variables through several preliminary tests [3,4].

Therefore, we suggest a new approach based on the RD cost and temporal correlations without any internal variables and also experiment with both the simple and complex sequences. In this paper, we propose the fast mode decision method which adopts the additional early SKIP and inactive intra/inter mode conditions based on the correlations of temporal and RD costs between the optimal/sub-optimal best modes.

In the following, we introduce the overview of H.264/AVC mode decision method in Sect. 2. After that, we proposed the fast mode decision algorithm using the RD cost and temporal correlations in Sect. 3; In Sect. 4, we evaluate the experimental results and make analysis. Finally, we make conclusions in Sect. 5.

2 Overview of Mode Decision in H.264/AVC

2.1 Mode Decision Method

Using the RD cost, H.264/AVC decides to select the best mode for each macroblock. The RD costs consist of J_{motion} and J_{mode} [5]. Generally, J_{motion} shows worse performance than J_{mode} in terms of the best mode decision. However, to reduce the complexity, the encoder calculates J_{motion} to determine the best motion vector (*MV*) and reference picture number (*REF*), since J_{motion} does not

calculate the actual coding bits but estimates them to reduce the complexity. The smallest of J_{motion} is called the sub-optimal mode. In the J_{motion} equation, it includes the sum of absolute difference (SAD) between the original and reference blocks. These are defined as follows:

$$J_{motion}(MV, REF|\lambda_{motion}) = SAD(s, r(MV, REF)) + \lambda_{motion} \cdot R(MV, REF). \quad (1)$$

$$SAD(s, r(MV, REF)) = \sum_{x \in H, y \in V}^{H,V} |s(x, y) - r(x - m_x, y - m_y)|. \quad (2)$$

where λ_{motion} is the Lagrangian multiplier that depends on the quantization parameters.

In addition, *R(MV, REF)* represents the coding bits used for the coding *MV* and *REF*, *s* and *r* indicate the current and reference blocks, respectively.

To decide the best mode for each macroblock, the encoder calculates the RD cost of each mode and then chooses the smallest one. The RD cost indicates J_{mode}, it is defined as follows:

$$J_{mode}(s, r, M|\lambda_{mode}) = SSD(s, r, M) + \lambda_{mode} \cdot R(s, r, M). \quad (3)$$

$$SSD(s, r, M) = \sum_{x \in H, y \in V}^{H,V} (s(x, y) - r(x - m_x, y - m_y))^2 \quad (4)$$

where M and λ_{mode} represent the macroblock coding mode and the Lagrange multiplier, respectively. Then, *SSD(s, r, M)* and *R(s, r, M)* denote the square of the distortion costs between the original and reconstructed signal, and the number of coding bits associated with the given mode, respectively.

2.2 Early SKIP Conditions

The fast high complexity mode of H.264/AVC JM reference software includes the fast inter mode decision. It is called the early SKIP conditions [1].

- *Reference frame = Previous frame*
- *SKIP MV = 16 × 16 MV*
- *Best motion compensated block size = 16 × 16*
- *Coded block pattern (CBP) = 0*

SKIP mode are encoded without motion and residual information. Hence, it can be reduce the complexity. The early SKIP conditions decide whether the optimal best macroblock mode is SKIP mode or not. When the above conditions are satisfied simultaneously, all remaining encoding parts should be omitted.

3 Proposed Fast Mode Decision Algorithm

3.1 *Fast Inter Mode Decision in P Frame*

In our proposed fast inter mode decision algorithm, we use two kinds of correlations: the temporal correlation and the RD cost correlation for the optimal/sub-optimal best modes. In here, the sub-optimal best mode is decided, when the mode has the minimum value of J_{motion}. As shown in Fig. 2, the frequency distribution of SKIP mode is quite higher than other inter modes; in here, the inter modes consist of SKIP, 16 × 16, 16 × 8, 8 × 16, and P8 × 8.

Jang et al. proposed a fast mode decision algorithm which used the correlation between the sub-optimal best mode and the optimal best mode [6]. In this sense, we inactivate the modes rarely selected as the optimal best mode using the sub-optimal mode. For example, when the sub-optimal best mode is 16 × 16 mode, the percentage which the optimal best mode is selected as 8 × 16, 16 × 8 or P8 × 8 mode is low.

Figure 3 shows each inter mode RD costs. It is distinct from each other. Besides, Ri et al. showed that there exists the strong Lagrangian cost correlation between the temporally collocated blocks [3]. As the candidate mode to predict the best mode for the current macroblock, we can use the best mode of the collocated macroblock in the previous frame. Therefore, we propose the additional early SKIP conditions and inactive mode decision methods of the inter modes.

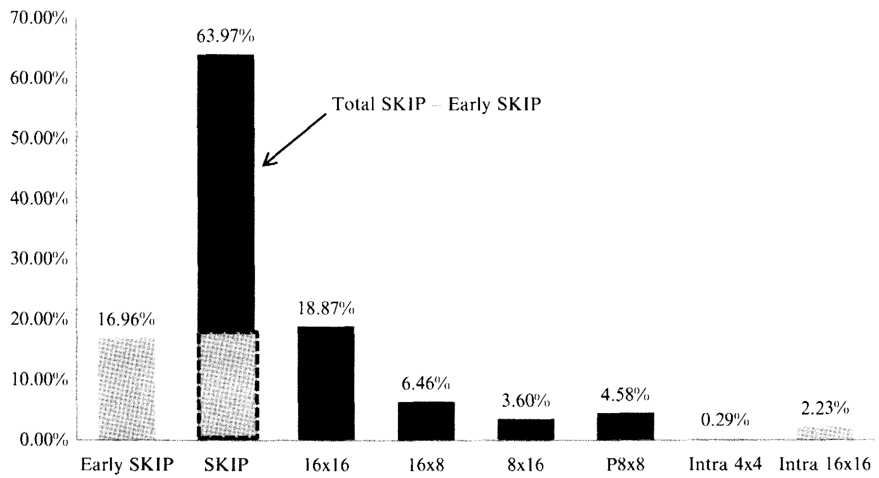

Fig. 2 Optimal best mode distribution in P frames (Sequence: hall_cif.yuv, QP: 27)

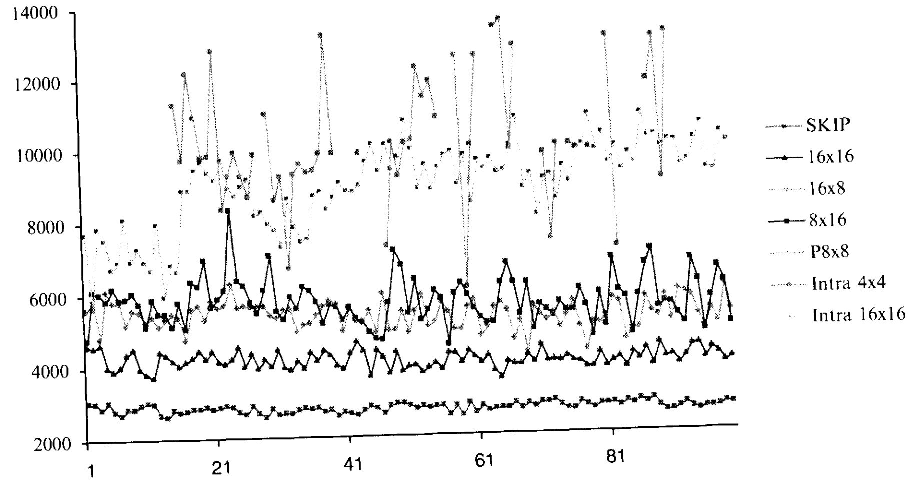

Fig. 3 J_{mode} of each best mode in P frames (Sequence: hall_cif.yuv, QP: 27)

3.1.1 Inter Condition 1: Additional Early SKIP Condition

The early SKIP conditions are used in the fast high complexity mode of the JM reference software; that is to say H.264/AVC omits unnecessary calculations via the early SKIP conditions (in Sect. 2.2). However, the number of SKIP modes by the early SKIP conditions is not sufficiently high percentage compared to the total number of SKIP modes. Figure 2 shows the satisfied percentage of the early SKIP conditions accounts for only one-third of SKIP mode. It means there is still some room for determining SKIP mode. The additional early SKIP conditions as follows.

- *(Reference frame = Previous frame)*]&& *(SKIP MV = 16 × 16 MV)*
- $\{J_{mode(SKIP)} < avg\_J_{prev-mode(SKIP)}\} \&\& \{prev\_best\_mode(SKIP)\}$

where $J_{mode(i)}$ and $avg\_J_{prev\text{-}mode(i)}$ denote the RD cost in the current macroblock and the average RD cost in the previous frame for the corresponding mode, respectively. $prev\_best\_mode(i)$ represents the best mode of the collocated macroblock in the previous frame according to the mode i.

The first condition comes from the early SKIP conditions in the original JM reference software (in Sect. 2.2), and the second condition is proposed by the temporal correlation. If $J_{mode(\mathrm{SKIP})}$ is less than $avg\_J_{prev\text{-}mode(SKIP)}$ and the collocated macroblock mode is SKIP, we can expect SKIP mode to be selected as the best mode.

The original procedure in H.264/AVC is that $J_{mode(16\times16)}$ is calculated in the fast high complexity mode; and then the encoder calculates J_{mode} for the remaining modes at last. While the encoder check the early SKIP conditions, the encoder already has the $J_{mode(16\times16)}$ value; therefore, we can calculate $J_{mode(SKIP)}$ in this step. It does not give any time saving, but it will be used for inactive P8 × 8 mode decision.

3.1.2 Inter Condition 2: Inactivate 16 × 8 and 8 × 16 Modes

Compared to SKIP and 16 × 16 modes, 16 × 8 or 8 × 16 mode is not settled as the best mode frequently, as shown in Fig. 2. Nevertheless, H.264/AVC exhaustively calculates all modes to improve coding performance. In order to inactivate 16 × 8 and 8 × 16 modes, we use the correlation between the optimal and sub-optimal best modes and the temporal correlations. We check whether the mode is inactive or not as follows:

- $\{(J_{motion(16\times16)}<J_{motion(16\times8)})||(J_{motion(16\times16)}<J_{motion(8\times16)})\}$
- $\{prev\_best\_mode(SKIP)||prev\_best\_mode(16\times16)\}$

where $J_{motion(i)}$ represents J_{motion} value for the corresponding mode i at the current macroblock.

The first condition is based on the RD cost correlation between the sub-optimal and optimal best modes (in Sect. 3.1). The second condition is using the temporal correlation between the current and collocated macroblock modes. We can expect the percentage that 16 × 8 or 8 × 16 mode is determined as the best mode is quite low when the above conditions are satisfied simultaneously.

Therefore, we can inactivate 16 × 8 and 8 × 16 modes according to the above conditions. However, if the proposed condition is not satisfied, we calculate $J_{mode(16\times8)}$ and $J_{mode(8\times16)}$ values at this step. It also does not save any encoding time, but it will be used at the step of the inactive P8 × 8 mode decision.

3.1.3 Inter Condition 3: Inactivate P8 × 8 Mode

The encoding time for the P8 × 8 mode decision is same as that of all other inter modes; it means the complexity is high. As stated in Sect. 3.2, P8 × 8 mode also needs to inactivate if the following condition is satisfied.

$$Minimum\{J_{mode(SKIP)},\ J_{mode}(16\times16),\ J_{mode}(16\times8),\ J_{mode}(8\times16)\} < avg\_J_{prev-mode(P8\times8)}$$

$J_{mode(i)}$ is already calculated in the previous steps. As shown in Fig. 3, the RD cost distribution of P8 × 8 is higher than those of other inter modes. Based on this feature, when $avg\_J_{prev\text{-}mode(P8\times8)}$ is less than the minimum value of $J_{mode(i)}$, P8 × 8 mode rarely selected as the best mode; in this case, we can inactivate P8 × 8 mode.

3.2 *Fast Intra Mode Decision in P Frame*

The H.264/AVC standard allows the intra and inter modes in P frames. The intra mode uses the spatial correlation features, whereas the inter mode utilizes the temporal correlations. In general, since P frames contain higher correlation in time than in space, the occurrence frequencies of the intra modes are lower than those of the inter modes, as shown in Fig. 2. Therefore, we should make conditions for inactive intra modes in P frames.

The intra modes consist of intra 4 × 4 and intra 16 × 16 in the baseline profile. Features of these modes are different; first, intra 16 × 16 mode is used in a homogeneous region, whereas intra 4 × 4 mode is determined at a complexity region. Second, the ranges of the RD cost values are quite different. In Fig. 3, $J_{mode(intra\ 4\times4)}$ is quite different from $J_{mode(intra\ 16\times16)}$; $J_{mode(intra\ 4\times4)}$ is much larger than $J_{mode(intra\ 16\times16)}$. Therefore, we propose the distinct ranges which are divided by the decided thresholds.

Most group of picture (GOP) structure is IPPP or IBBB. It means we can estimate the brief distributions of $J_{mode(intra\ 4\times4)}$ and $J_{mode(intra\ 16\times16)}$ via the first I frame. The distributions of each intra mode RD costs in the first I frame are similar to the one in the first P frame. Therefore, we can decide two thresholds: the maximum and minimum thresholds. The maximum and minimum thresholds are decided based on the average RD cost values of intra 4 × 4 and intra 16 × 16, respectively. The selective conditions for inactive intra mode are described as follows.

- $J_{mode(i)} < minimum\ threshold$: inactivate intra 16 × 16 and intra 4 × 4
- $Minimum\ threshold < J_{mode(i)} < maximum\ threshold$: inactivate intra 4 × 4
- $J_{mode(i)} > maximum\ threshold$: inactivate intra 16 × 16

where $i \in \{16\times16, 16\times8, 8\times16\}$, $maximum\ threshold = avg\_J_{mode(intra\ modes)} \times 1.5$,and minimum threshold = $avg\_J_{mode(intra\ modes)}/1.5$.

Figure 4 shows the flow chart of the proposed algorithm. As shown in Fig. 4, we determine the maximum and minimum thresholds in I frame using $avg\_J_{mode(intra\ modes)}$; and then, check the conventional and additional SKIP conditions to early terminate the remain mode decision parts, 16 × 8, 8 × 16, and P8 × 8 mode conditions to inactivate with the inter condition 1, 2, and 3 based on the correlation between the optimal and sub-optimal best modes and the temporal correlations. In addition, we decide whether intra 4 × 4 and intra 16 × 16 modes are inactive or not, based on the intra conditions 1 and 2. Finally, we decide the final mode by comparing the RD cost.

4 Experimental Results and Analysis

To evaluate coding performance of the proposed algorithm, the reference software version JM 12.4 was modified [7]. Table 1 indicates the experimental conditions.

We compared coding performance in terms of time saving, bit saving rates, PSNR differences [8]. To evaluate these performance, we use the Eqs. (5)–(7).

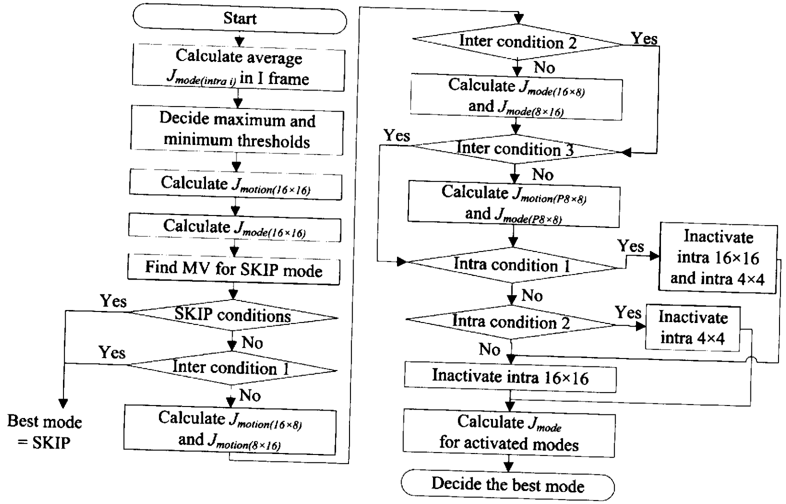

Fig. 4 Flowchart of proposed algorithm

Table 1 Experimental conditions

Classify	Encoder description
RDOptimization	Fast high complexity mode
Encoding frames	100
Quantization parameter	22, 27, 32, 37
The number of reference frames	5

$$\Delta Time(\%) = \frac{Time_{proposed} - Time_{original}}{Time_{original}} \times 100. \quad (5)$$

$$\Delta Bitrates(\%) = \frac{Bitrates_{proposed} - Bitrates_{original}}{Bitrates_{original}} \times 100. \quad (6)$$

$$\Delta PSNR(dB) = PSNR_{proposed} - PSNR_{original}. \quad (7)$$

In order to verify efficiency of each proposed algorithm, our experiment included three sections: intra, inter, and intra + inter modes. Table 2 shows the performance comparison results between the proposed intra/inter mode and the fast high complexity mode decision of H.264/AVC. We can reduce the encoding time approximately 22% and 39%, respectively.

In Table 3, we compare the performances between Jang's algorithm and our inter mode algorithm. The average saving time of Jang's algorithm is faster than our proposed algorithm for inter mode decision, but other performances which are PSNR and bit rate changes are better than Jang's algorithm. Additionally, our

Table 2 Experimental results of each intra/inter mode

		Intra mode			Inter mode		
Sequence	QP	ΔTime (%)	ΔBitrate (%)	ΔPSNR (dB)	ΔTime (%)	ΔBitrate (%)	ΔPSNR (dB)
Mobile	22	−25.54	0.00	−0.01	−38.87	1.56	−0.07
	27	−23.63	−0.02	−0.01	−43.23	1.88	−0.07
	32	−21.70	0.10	−0.01	−46.65	2.05	−0.06
	37	−19.90	0.10	0.00	−49.62	1.34	−0.02
Paris	22	−30.63	0.24	−0.01	−49.48	1.79	−0.07
	27	−25.51	0.29	−0.03	−45.71	2.07	−0.07
	32	−21.09	0.76	−0.01	−43.72	1.75	−0.05
	37	−19.07	0.81	−0.03	−43.59	0.91	−0.05
Flowergarden	22	−22.46	0.01	−0.08	−23.52	0.51	−0.02
	27	−21.59	0.15	−0.06	−24.24	0.53	−0.03
	32	−20.07	0.22	−0.03	−27.13	0.39	−0.05
	37	−17.33	0.05	−0.03	−31.84	−0.56	−0.07
Tempete	22	−26.11	0.69	0.00	−30.19	1.23	−0.04
	27	−24.02	1.06	0.01	−33.63	1.30	−0.03
	32	−21.40	2.40	0.01	−37.38	0.68	−0.04
	37	−19.29	2.05	−0.03	−37.48	−0.30	−0.02
Stefan	22	−23.98	0.50	−0.02	−38.99	1.29	−0.04
	27	−21.85	0.54	−0.03	−41.97	1.89	−0.05
	32	−19.92	0.82	−0.02	−44.80	2.32	−0.04
	37	−17.78	0.80	−0.02	−46.68	0.87	−0.03
Akiyo	22	−24.04	0.17	−0.01	−39.90	0.93	−0.04
	27	−21.40	−0.42	−0.01	−38.88	−0.56	−0.04
	32	−17.98	−0.82	−0.02	−32.08	−1.59	−0.02
	37	−16.22	−0.17	−0.04	−26.01	−0.44	−0.05
Hall	22	−27.79	1.35	0.00	−37.65	0.45	−0.05
	27	−24.94	0.29	−0.04	−48.40	−2.49	−0.06
	32	−19.09	1.75	−0.04	−44.84	−1.24	−0.06
	37	−17.13	1.45	−0.07	−38.12	0.48	−0.01
Container	22	−28.05	0.46	−0.02	−51.11	0.06	−0.07
	27	−23.09	0.49	−0.02	−49.72	−0.54	−0.04
	32	−17.73	0.14	−0.02	−40.15	−1.27	−0.04
	37	−16.52	−0.08	−0.01	−32.14	−1.85	−0.03
Total Average		**−21.78**	**0.51**	**−0.02**	**−39.30**	**0.48**	**−0.04**

algorithm includes the fast intra mode decision scheme. We can get the combined experimental results as shown in Table 4, when we experiment about the fast intra and inter modes simultaneously. It shows the total average of the saving time rate is about 58.86% with negligible BDPSNR and BDBR, compared to the H.264/AVC fast high complexity mode [9]. The average BDBR increases approximately 2.49% and BDPSNR decreases 0.12 dB. Figure 5 shows the rate-distortion curves.

Table 3 Performance comparison between Jang's and our proposed inter mode algorithm

	Proposed inter mode algorithm			Jang's algorithm [6]		
Sequence	ΔTime(%)	ΔBitrate(%)	ΔPSNR()	ΔTime(%)	ΔBitrate(%)	ΔPSNR(dB)
Mobile	−22.69	0.05	−0.01	−18.20	1.52	−0.05
Paris	−24.07	0.53	−0.02	−27.80	3.26	−0.05
Flowergarden	−20.36	0.11	−0.05	−26.03	2.27	−0.02
Stefan	−20.88	0.66	−0.02	−24.56	1.30	−0.04
Akiyo	−19.91	−0.31	−0.02	−31.35	0.66	−0.02
Hall	−22.24	1.21	−0.04	−39.20	1.03	−0.07
Container	−21.35	0.25	−0.02	−33.13	2.04	−0.02
Average	−21.64	0.36	−0.03	−28.61	1.72	−0.04

Table 4 Experimental results of inter + intra modes

		Inter mode and Intra mode				
Sequence	QP	ΔTime(%)	ΔBitrate(%)	ΔPSNR(dB)	BDBR(%)	BDPSNR(dB)
Mobile	22	−63.20	1.57	−0.07	1.82	−0.12
	27	−65.31	−1.24	−0.08		
	32	−66.55	2.26	−0.07		
	37	−66.48	1.21	−0.04		
Paris	22	−66.60	2.04	−0.07	3.63	−0.19
	27	−62.57	2.42	−0.08		
	32	−59.94	2.29	−0.07		
	37	−59.05	1.49	−0.09		
Flowergarden	22	−47.05	0.52	−0.11	1.71	−0.11
	27	−46.20	0.57	−0.08		
	32	−46.07	0.30	−0.09		
	37	−47.81	−0.33	−0.09		
Tempete	22	−56.42	1.86	−0.05	3.20	−0.15
	27	−56.73	2.18	−0.03		
	32	−58.27	2.71	−0.05		
	37	−57.70	3.06	−0.02		
Stefan	22	−62.13	1.77	−0.08	3.92	−0.20
	27	−62.80	2.30	−0.08		
	32	−62.98	3.17	−0.06		
	37	−62.77	2.32	−0.06		
Akiyo	22	−59.84	1.00	−0.04	0.70	−0.03
	27	−57.63	0.00	−0.05		
	32	−49.68	−1.52	−0.05		
	37	−40.93	−0.96	−0.09		
Hall	22	−63.97	1.84	−0.06	3.48	−0.10
	27	−66.75	−1.34	−0.09		
	32	−59.60	0.93	−0.10		
	37	−52.80	2.47	−0.11		
Container	22	−71.30	0.84	−0.07	1.42	−0.05
	27	−68.00	0.43	−0.06		
	32	−57.36	−1.02	−0.05		
	37	−49.78	−1.80	−0.03		
Total Average		**−58.57**	**1.04**	**−0.07**	**2.49**	**−0.12**

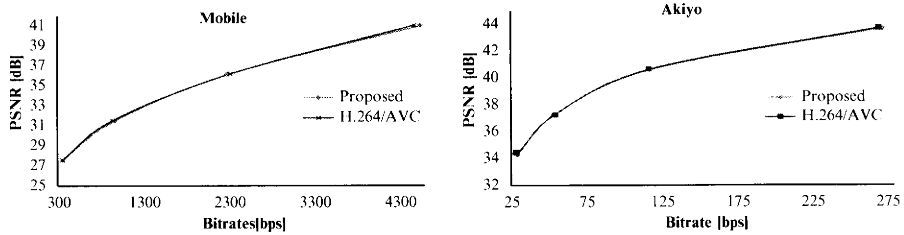

Fig. 5 Rate-distortion curves

5 Conclusion

H.264/AVC has been developed by focusing on coding performance instead of the complexity. For this reason, it is hard to use in real-time applications. One of the abundant computational parts is the mode decision part. Therefore, in this paper, we proposed the fast mode decision algorithm to reduce the encoding time by the additional early SKIP conditions, inactive inter mode (16 × 8, 8 × 16, and P8 × 8), and intra mode conditions (intra 4 × 4 and intra 16 × 16), based on the RD cost correlation for the optimal/sub-optimal best modes and the temporal correlation. From the experimental results, it reduced the encoding time by approximately 60%. However, the PSNR decreased only about 0.07 dB and bit rate increased only about 1%. Also, BDPSNR decreased 0.12 dB and BDBR increased 2.49% compared to the fast high complexity mode on of the H.264/AVC. Furthermore, our proposed algorithm for inter and intra is speeded up approximately 30% and the PSNR and bit rate are better as compare with Jang's algorithm.

Acknowledgement This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2011-0030822).

References

1. Wiegand, T., Sullivan, G.J., Bjontegaard, G., Luthra, A.: Overview of the H.264/AVC Video Coding Standard. IEEE Transactions on Circuits and Systems for Video Technology, 13(7), pp. 560–576 (2003)
2. Richardson, I.E.G.: H.264 and MPEG-4 Video Compression - Video Coding for Next-generation Multimedia. Wiley. Chichester (2003)
3. Ri, S.H., Ostermann, J.: Fast Inter-Mode Decision in an H.264/AVC Encoder Using Mode and Lagrangian Cost Correlation. IEEE Transactions on Circuits and Systems for Video Technology, 19(2), pp. 302–306 (2009)
4. Pan, L.J. and Ho, Y.S.: Fast Mode Decision Algorithm for H.264 Inter Prediction. IET Electronics Letters, 43(24), pp. 1351–1353 (2007)

5. Lim, K.P., Sullivan, G.J., Wiegand, T.: Text Description of Joint Model Reference Encoding Methods and Decoding Concealment Methods. JVT-N046, JVT of ISO/IEC MPEG and ITU-T VCEG (2005)
6. Jang, W.S., Heo, J., Ho, Y.S.: Efficient Selection of Candidates for Fast Inter Mode Decision in H.264. In: International Symposium on Signal Processing, Image Processing and Pattern Recognition (SIP), pp. 222–225, (2008)
7. JVT Reference Software version JM 12.4,: http://iphome/hhi.de/shehring/tml/download/old_jm/jm12.4.zip.
8. Bjontegaard, G.: Calculation of Average PSNR Differences between RD-curves. ITU-T Q.6/16, Doc. VCEG-M33 (2001)
9. Bjontegaard, G.: Improvement of the BD-PSNR Model. ITU-T Q6/16, Doc. VCEG-AI11, (2008)