Fast Inter/Intra Mode Decision based on Cost and Temporal Correlations in P Frames

Soo-Jin Hwang, Jae-Il Jung, and Yo-Sung Ho
Gwangju Institute of Science and Technology (GIST)
123 Cheomdan-gwagiro (Oryong-dong), Buk-gu
Gwangju 500-712, Republic of Korea
{soojin, jijung, hoyo}@gist.ac.kr

***Abstract*—In this paper, we aim to reduce the encoding complexity in H.264/AVC. We focus on the rate-distortion cost (RD cost) and temporal correlations to reduce the complexity. We apply the proposed algorithm in inter modes: SKIP, 16×8, 8×16, and P8×8 and intra modes: Intra 4×4 and Intra 16×16 which the occurrence ratios are low. Experimental results show the proposed method provides approximately 50% encoding time savings on average with slightly increased bit-rate and with negligible loss of peak signal-to-noise ratio (PSNR), compared to the H.264/AVC fast high complexity mode.**

Keywords-component; H.264/AVC, Video coding, Rate-distortion optimization, Mode decision, Temporal correlation, Rate-distortion cost

I. Introduction

H.264/AVC is a popular international video coding standard that was developed by the Joint Collaborative Team on Video Coding (JCT-VC) [1]. H.264/AVC yields considerably high coding efficiency compared to the previous video coding standards, since accomplished by lots of advanced methods. The methods are variable block size motion compensation, multiple reference frames, and partial prediction in the intra prediction process, integer transform, deblocking filter, and context-based adaptive entropy coding [1].

In addition, one of the advanced methods is the rate-distortion optimization (RDO) technique. In order to decide the best mode, the encoder calculates the RD cost for all the possible modes and then chooses the best coding mode which has the lowest RD cost in the final coding process [1]. Based on the RDO technique, we are able to increase image quality, and also reduce coding bits by checking for all the possible mode combinations exhaustively. However, the RDO technique drastically increases coding complexity. Fig. 1 demonstrates that mode decision complexity is high [2].

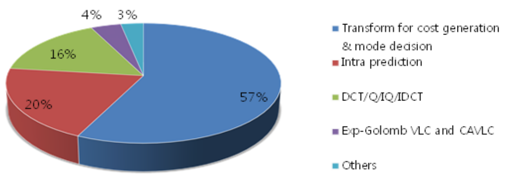

Figure 1. Run-time proportions of each major modules in H.264/AVC

To decide the best mode for each macroblock (MB), it has to be calculated totally 660 times; inter and intra modes take up 68 and 592 times in one P frame, respectively. It means H.264/AVC cannot available for real time applications.

The H.264/AVC JM reference software provides the fast inter mode decision. It is called the early SKIP conditions [1]. The MB is determined to SKIP is encoded without any additional information. In the fast high complexity mode of H.264/AVC, the early SKIP conditions decide whether the optimal best MB mode is SKIP mode or not. When the below 4 conditions are satisfied simultaneously, all remaining encoding parts should be omitted. The early SKIP method reduces the complexity.

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

Many people have proposed various fast mode decision algorithms. Pan et al. [2] proposed a fast intra mode decision algorithm using information of the edge directions, but this method needs an additional operation time to get the edge map. Lim et al. [3] and Pan et al. [4] used their own internal variables through several preliminary tests. Furthermore, most proposed methods carry out experiments on only simple sequences. Therefore, we suggest a new approach based on the RD cost and temporal correlations without any internal variables and spending any additional operation time. We also need to propose for any sequence not only simple sequences, but also complex sequences.

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

II. Overview of Mode Decision Scheme in H.264/AVC

As specified the mode decision method for luminance components, the reference software uses five different kinds of inter modes which consist of SKIP, 16×16, 16×8, 8×16, and

Copyright and reproduction permission: All rights are reserved by the authors. No part of this publication may be reproduced or distributed without permission from the respective authors.

P8×8 modes to consider the temporal correlation. It also uses two intra modes which are composed of Intra 16×16 and Intra 4×4 modes for the spatial correlation. In addition, only for High profile in H.264/AVC provides Intra 8x8 mode.

Generally, SKIP and 16×16 modes are coded in a homogeneous region. On the contrary to this, smaller block size modes such as P8×8 mode are used in the area containing the boundaries of moving objects.

To decide the best *MV* and *REF*, we use the RDO technique based on the Lagrangian multiplier. The minimum costs which are J_{motion} and J_{mode} are chosen the motion costs and the mode cost among all modes. Equation (1) demonstrates J_{mode} and Equation (2) denotes *SSD* included in Equation (1). The encoder calculates each mode RD cost, and then chooses the minimum one among all RD costs of all enable modes in H.264/AVC. The RD cost indicates J_{mode}, the mode having the smallest value of J_{mode} is called the optimal mode. It is defined as follows:

$$J_{mode}(s, r, M| \lambda_{mode}) = SSD(s, r, M) + \lambda_{mode} \cdot R(s, r, M). \quad (1)$$

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

where M and λ_{mode} represent the MB coding mode and the Lagrange multiplier, respectively. Then, *SSD(s, r, M)* and *R(s, r, M)* are the square of the distortion costs between the original and reconstructed signal, respectively. Equation (3) presents J_{motion} and (4) denotes *SAD*. The mode having the smallest J_{motion} is called the sub-optimal mode. It is denotes as follows:

$$J_{motion}(MV,REF|\lambda_{motion}) = SAD(s, r(MV, REF)) + \lambda_{motion} \cdot R(MV, REF). \quad (3)$$

$$SAD(s,r(MV,REF)) = \sum_{x\in H,\, y\in V}^{H,V} | s(x,y) - r(x-m_x, y-m_y) |. \quad (4)$$

where λ_{motion} is the Lagrangian multiplier that depends on the quantization parameters. Additionally, *s* and *r* mean the current and reference blocks, respectively. *R(MV, REF)* denotes the coding bits used for the coding *MV* and *REF*. Equation (4) is *SAD* which represents the sum of absolute difference between the original and reference blocks.

J_{motion} shows worse performance than J_{mode} in terms of the best mode decision. In order to determine the best *MV* and *REF*, the reference software calculates J_{motion}. J_{motion} can reduce the complexity compared to J_{mode}, because J_{motion} does not calculate the actual coding bits. Kim et al. [5] proposed the fast mode decision algorithm based on J_{motion}, since it helps to estimate the candidate mode. However, it is not accurate in order to decide the best mode.

III. Proposed Fast Mode Decision Algorithm

The temporal correlation and the RD cost correlation of each best mode are used in our proposed algorithm. As shown in Fig. 2, the portion of the decided best modes is skewed at SKIP and 16×8 in a sequence. It means the frequency distributions of 16×8, 8×16, and P8×8 modes are lower than its of SKIP and 16×16 modes. Also, the satisfied percentage of the early SKIP conditions is about 30% of all decided as SKIP mode. If the quantization parameter (QP) decreases, the percentage which meets the early SKIP conditions should be decreased. Furthermore, intra modes percentages are significantly low compared to inter modes in P frame, since a P frame has strong temporal correlation.

The original JM reference software calculates $J_{mode(SKIP)}$ at the first step in the fast high complexity mode, and then J_{mode} for the remaining modes at last for P frame. However, we calculate the RD cost for each mode, when the encoder decides the each J_{motion}.

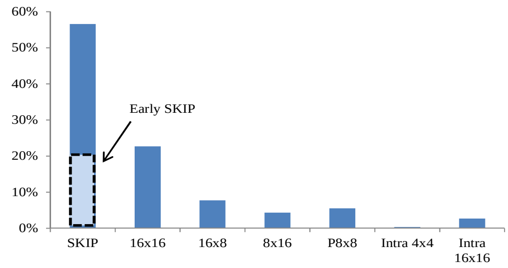

Figure 2. Percentage of the decided optimal best mode

A. Fast Inter Mode Decision

1) Supplementary early SKIP conditions

The early SKIP conditions are used in the fast high complexity mode of the H.264/AVC JM reference software; that is to say the encoder omits unnecessary calculations via the early SKIP conditions, especially in the fast high complexity mode. SKIP mode refers to 16x16 mode which any information of residual or motion is not encoded.

However, the number of SKIP modes satisfying the early SKIP conditions is not sufficiently high percentage compared to the total number of SKIP modes, as shown in Fig. 2. It accounts for only one-third of SKIP mode. Additionally, Kim et al. shows the frequency of the SKIP mode for each sequence [5]. This result shows there is still some room for determining SKIP mode in the early step by developing the additional early SKIP conditions. Equations (5) and (6) show the conditions for determining early SKIP.

(*Reference frame: Previous*) & (*SKIP MV == 16×16 MV*) (5)

$\{J_{mode(SKIP)} < J_{prev\text{-}mode(SKIP)}\}$ & { *prev_col_MB(SKIP)* } (6)

where $J_{mode(i)}$ and $J_{prev\text{-}mode(i)}$ denote the RD cost in the current MB and the average RD cost in the previous frame for the corresponding mode, respectively. The term of *prev_col_MB(i)* represents the best mode of the collocated MB in the previous frame according to the mode *i*.

Video sequences are highly correlated in time because almost experimental sequences are over 30 frames per second. Equation (5) comes from the early SKIP conditions in the original JM reference software using temporal correlation and the mode feature. We propose the second condition as equation (6). If $J_{mode(SKIP)}$ is less than $J_{prev\text{-}mode(SKIP)}$ and the collocated MB mode is SKIP, we can expect SKIP mode to be selected as the best mode; it comes from the idea which the RD cost distributions are similar by each mode. S. Ri et al. also showed there exists the strong Lagrangian cost correlation between the temporally collocated blocks [3]. The probability which the collocated MB mode is same as current MB mode is high.

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.

2) Inactivate 16×8 mode and 8×16 mode conditions

We can check 16×8 and 8×16 modes have 2 similar features, as shown in Fig. 2 and Fig. 3. The first is the frequency distribution is low. The second is the RD cost distributions are similar between 16×8 and 8×16 mode. Equation (7) and (8) are the proposed conditions for 16×8 and 8×16 modes.

$$min(J_{mode(SKIP\ or\ 16\times16)}) < avg(J_{prev\text{-}mode(16\times8)}, J_{prev\text{-}mode(8\times16)}). \quad (7)$$
$$\{prev\_col\_MB(SKIP)\ \|\ prev\_col\_MB(16\times16)\}. \quad (8)$$

where $min(J_{mode(i\ or\ j)})$ represents the smallest J_{mode} value between i mode and j mode. The term of $avg(J_{prev\text{-}mode(i)}, J_{prev\text{-}mode(j)})$ denotes the mean value between $J_{prev\text{-}mode(i)}$ and $J_{prev\text{-}mode(j)}$.

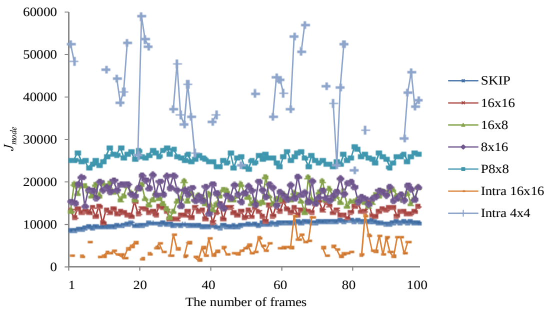

Figure 3. The best inter mode RD cost values of each MB

3) Inactivate P8×8 mode condition

We know P8×8 mode also in the same sense as like 16×8 and 8×16 modes. Therefore, we can apply same idea to find inactivate mode adaptively. In this case, we had the RD cost of SKIP, 16×16, 16×8, and 8×16 modes. In Fig. 3, we know the RD cost distributions for 16×8 and 8×16 modes do not have significant differences. The condition for P8×8 mode is proposed as following equation (9).

$$min(J_{mode(SKIP,16\times16,16\times8,\ or\ 8\times16)}) < J_{prev\text{-}mode(P8\times8)}. \quad (9)$$

B. Fast Intra Mode Decision

H.264/AVC allows both the intra and inter modes in P frames. The inter modes utilize the temporal correlations, whereas the intra modes use the spatial correlation features. As shown in Fig. 2, the occurrence frequencies of the intra modes are lower than those of the inter modes in P frame.

The intra modes consist of Intra 4×4 and Intra 16×16. As shown in Fig. 3, we can notice the ranges of the intra mode RD cost values are quite different between Intra 4×4 and Intra 16×16. $J_{mode(intra\ 4\times4)}$ is much higher than $J_{mode(intra\ 16\times16)}$. In this sense, we propose the distinct ranges which can be divided by the decided thresholds. The decided thresholds are the adaptive threshold. After I frame coding, we have the RD cost values of Intra 4×4 and Intra 16×16 modes. Using these RD costs, we can decide the thresholds to reduce rarely selected intra modes.

In general, the encoder structure has to encode a sequence from the first I frame. It means the group of pictures (GOP) can be IPPP, IBBP, or IIII. In this sense, we can briefly anticipate the specific RD cost ranges of intra modes. We checked the average RD costs of intra modes are similar based on preliminary tests. Hence, the maximum and minimum thresholds are defined as follows.

$$max\_thr = avg\_J_{mode(intra\ 4\times4)}. \quad (10)$$
$$min\_thr = avg\_J_{mode(intra\ 16\times16)}. \quad (11)$$

where $avg\_J_{mode(intra\ i)}$ is the average RD cost of the corresponding intra mode in the first I frame. The terms of min_thr and max_thr are minimum and maximum thresholds, respectively.

Eq. (12), Eq. (13), and Eq. (14) are the fast intra mode conditions based on the distinctive RD cost ranges between Intra 16×16 and Intra 4×4 modes.

$$J_{mode(i)} < min\_thr\text{: inactivate Intra 16×16 and 4×4.} \quad (12)$$
$$min\_thr < J_{mode(i)} < max\_thr\text{: inactivate Intra 4×4.} \quad (13)$$
$$J_{mode(i)} > max\_thr\text{: inactivate Intra 16×16.} \quad (14)$$

where i is a inter mode among 16×16, 16×8 and 8×16 modes.

Fig. 4 is the flow chart of the proposed fast mode decision algorithm.

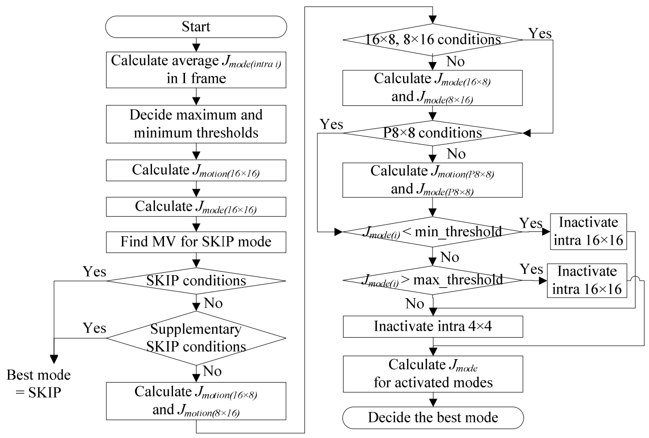

Figure 4. Flow chart of the proposed algorithm

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The H.264/AVC reference software version JM 12.4 [6] was modified to evaluate coding performance of the proposed algorithm. The encoding parameters are in Table 1. In our experiment results, we compared performance in terms of time saving rates, changes of bit-rate and PSNR using the following equations (15), (16) and (17) [7]:

TABLE II. EXPERIMENTAL RESULTS

Sequences (CIF)	Intra			Inter			Kim's algorithm [5]			Intra + Inter		
	ΔTime (%)	*ΔBit rate (%)*	*ΔPSNR (dB)*	*ΔTime (%)*	*ΔBit rate (%)*	*ΔPSNR (dB)*	*ΔTime (%)*	*ΔBit rate (%)*	*ΔPSNR (dB)*	*ΔTime (%)*	*ΔBit rate (%)*	*ΔPSNR (dB)*
Akiyo	-14.92	0.03	-0.01	-31.17	-0.15	-0.04	-21.05	-0.45	-0.22	-44.43	-0.18	-0.06
Container	-21.74	0.27	-0.01	-43.85	-0.04	-0.05	-15.46	-0.46	-0.01	-57.12	-0.14	-0.06
Tempete	-16.97	0.41	-0.01	-27.93	0.51	-0.06	-10.61	0.02	0.00	-43.49	1.16	-0.07
Hall	-16.67	-0.06	-0.02	-41.00	-0.36	-0.09	-11.03	-0.32	0.00	-55.21	-0.36	-0.09
Mother	-14.50	0.09	-0.01	-53.09	0.62	-0.06	-20.23	-0.21	-0.01	-46.51	0.67	-0.06
News	-15.19	0.11	-0.01	-57.74	0.85	-0.06	-14.54	-0.43	0.00	-46.98	1.32	-0.06
Paris	-17.67	0.21	-0.02	-42.64	1.69	-0.08	-8.38	-0.05	-0.01	-54.33	1.88	-0.08
Mobile	-10.86	0.06	-0.01	-33.51	2.32	-0.10	-0.64	-0.03	0.00	-50.24	2.54	-0.11
Foreman	-14.58	0.33	-0.3	-37.19	1.74	-0.06	-18.65	-0.18	0.00	-52.53	2.32	-0.08
Average	**-15.90**	**0.16**	**-0.04**	**-40.90**	**0.79**	**-0.06**	**-13.40**	**-0.23**	**-0.03**	**-50.09**	**1.02**	**-0.07**

TABLE I. EXPERIMENTAL CONDITIONS

Classify	***Condition***
RDOptimization	*Fast high complexity mode (2)*
Sequence size	*CIF (352 x 288)*
Encoding frames	*100*
Frame structure	*IPPP*
Quantization parameter	*22, 27, 32, 37*
Search mode	*UMhexagonS*
The number of reference frames	*5*

$$\Delta Time\ (\%) = \frac{Time_{proposed} - Time_{original}}{Time_{original}} \times 100 \quad (15)$$

$$\Delta Bitrates\ (\%) = \frac{Bitrates_{proposed} - Bitrates_{original}}{Bitrates_{original}} \times 100 \quad (16)$$

$$\Delta PSNR\ (dB) = PSNR_{proposed} - PSNR_{original} \quad (17)$$

In order to verify efficiency of each proposed algorithm, we present the experimental results in Table II. There are three parts of results. Firstly, it indicates the performance comparison results between the proposed and the fast high complexity mode decision of H.264/AVC reference software JM 12.4. Furthermore, it shows performances of each intra and inter mode decision independently and combined intra and inter mode decision, respectively. Also, it shows the comparison between Kim's method and our proposed algorithm in respect to only inter mode decision.

The encoding time for inter and intra mode decisions is reduced approximately 40% and 13%, respectively. The average of the total saving time rate is approximately 50%. Nevertheless, the bitrate is only increases 1% and the PSNR decreases 0.07dB. Also, it shows the time saving ratio of Kim's algorithm is not effective at complex sequences such as *Mobile* and *Paris*. Comparing Kim's algorithm to only our inter mode algorithm, the saving time of our proposed algorithm is faster than Kim's algorithm over 27.5%. Furthermore, our algorithm includes the fast intra mode decision scheme. Therefore, our proposed algorithm is speeded up 37% compared to Kim's algorithm. Even though the performance qualities of our proposed algorithm are slightly worse, it can be negligible.

V. CONCLUSION

H.264/AVC has been developed for the better coding performance than the fast encoding time. For this reason, it is hard to use in real-time applications. One of the abundant computational parts is the mode decision part. Hence, we proposed the fast mode decision algorithm to reduce the complexity by the inactivate mode conditions using the RD cost correlation and the temporal correlation. We applied the proposed algorithm at inter modes: supplementary early SKIP mode, 16×8, 8×16, and P8×8 and intra modes: Intra 4×4 and Intra 16×16. Experimental results show that it reduced the encoding time by approximately 50%. Although the bit rate was increased less than 1% and also PSNR was decreased only about 0.07 dB, compared to the fast high complexity mode.

ACKNOWLEDGMENT

This research is supported by Ministry of Culture, Sports and Tourism (MCST) and Korea Creative Content Agency (KOCCA) in the Culture Technology (CT) Research & Development Program 2011.

REFERENCES

[1] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," IEEE Transactions on Circuits and Systems for Video Technology, vol. 13(7), pp. 560-576, July 2003.

[2] F. Pan, X. Lin, S. Rahardja, K.P Lim, Z.G.Lim D.Wu, and S. wu, "Fast mode decision algorithm for intraprediction Fast Mode Decision Algorithm for Intraprediction in H.264/AVC Video Coding," IEEE Transactions on Circuits and Systems for Video Technology, vol. 15(7), pp. 813-822, July 2005.

[3] S. Ri, Y. Vatis, and J. Ostermann, "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, vol. 19(2), pp. 302-306, Feb. 2009.

[4] L. Pan and Y. Ho, "Fast Mode Decision Algorithm for H.264 Inter Prediction," Electronics Letters, vol. 43(24), pp. 1351-1353, Nov. 2007.

[5] G. Kim, B. Yoon, Y. Ho, "A Fast Inter Mode Decision Algorithm in H.264/AVC for IPTV Broadcasting Services," *Proc. of Visual Communication and Image Processing*, vol. 6508, pp. 65081L-1-10, 2007.

[6] JM 12.4: http://iphome/hhi.de/shehring/tml/download/old_jm/jm12.4.zip.

[7] G. Bjontegaard, "Calculation of Average PSNR Differences between RD-curves (VCEG-M33)," in VCEG Meeting (ITU-T SG16 Q.6), Austin, Texas, USA, April 2001.