International Journal of Soft Computing 5 (1): 1-6, 2010
ISSN: 1816-9503

Adaptive Transform Kernal Size Selection Algorithm for H.264/AVC Encoding

[1]Jongho Kim, [2]Byung-Gyu Kim and [2]Hilmi Badrul
[1]Department of Broadcasting and Telecommunications Media Research, Institute of Electronics and Telecommunications Research 138, Korea
[2]Department of Computer Science and Engineering, SunMoon University, Republic of Korea

Abstract: The DCT 8×8 transform mode has been adopted for high profile encoding in H264/AVC in addition to 4×4 integer transform. We propose an efficient DCT 8×8 transform mode decision method to reduce the computational complexity of the transform mode decision. We observed that the average variance of the Direct Current (DC) coefficients is quite different according to the best DCT kernel size. We use the amount of variation of the DC coefficients after DCT 4×4 procedure as a classification feature for the proposed algorithm and make a adaptive threshold using the average of variance of DC coefficients. We verify that the proposed algorithm can reduce the 8×8 transform mode decision by amount of 47.35-65.53%.

Key words: DC, DCT, quantization, parameters, kernel, Harbour sequences, Crew sequences, rate distribution

INTRODUCTION

MPEG-4 Part-10 H.264/AVC has been approved as the latest video coding standard by the Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG (Wiegand *et al*., 2003). It can achieve a >50% bit rate saving with the same quality, compared with previous standards. Many coding tools have been adopted to improve the compression efficiency.

The tools for high compression performance is including the variable block-size motion compensation, variable intra mode prediction modes based on spatial correlation, multiple reference pictures, Context-based Adaptive Binary Arithmetic Coding (CABAC) for entropy coding, weighted prediction, quarter-pel motion vectors and variable size integer discrete cosine transform kernels for High profile coding. But the computational complexity was increased due to these tools.

As increasing the demand on high quality video service, the importance of tools for High profile is also increasing, too. Most of tools which are used for High profile are very complex such as Context-based Adaptive Binary Arithmetic Coding (CABAC), weighted prediction and variable size integer discrete cosine transform kernels.

The variable block size motion estimation and the intra mode prediction cause serious computational complexity in the whole encoding system. Much effort has been devoted to reducing the complexity of the encoder in order to develop fast algorithms (Nie and Ma, 2002; Byung-Gyu, 2008b) and also many researches have been conducted for a long time.

An Adaptive Rood Pattern Search (ARPS), which is based on two sequential search stages has been suggested by Nie and Ma (2002) for fast motion vector estimation. Jing and Chau (2004) developed a fast inter-mode decision scheme by using both the frame difference and the MB difference. This is a useful scheme, since only frame difference and MB difference images are required to assign a mode. To detect an early SKIP block (Crecos and Yang, 2005) also used the prediction and thresholding scheme for fast inter mode decision process. Kim has proposed a fast Macro Block (MB) mode prediction and decision algorithm based on temporal correlation for P-slices in the H.264/AVC video standard (Byung-Gyu, 2008b). He used a MB tracking scheme to find the most correlated block and the R-D cost of that block are suggested for early inter mode determination.

For fast intra mode decision scheme, a directional field based approach has been reported by Pan *et al.* (2005) where several directions are selected by using an edge direction texture histogram according to block types. Choi *et al.* (2006) used an early SKIP detection method and a selective intra-mode search for inter-frames to speed-up the encoding system. This method has been adopted as a fast mode decision option in JM reference software. Kim has proposed a fast intra mode determination algorithm based on the Macro Block (MB)

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

tracking scheme and Rate-Distortion (RD) cost (Byung-Gyu, 2008a; Kim *et al*., 2006). He also developed a refinement process for speeding-up the intra mode decision process. Kim and Kuo (2007) suggested a fast intra mode decision method using joint spatial and transform domain feature. Cheng *et al*. (2005) have studied and proposed a fast intra SKIP detection using distortion values of the 3 spatially adjacent blocks. An intra mode decision algorithm that employees fast edge detection method which is based on Non-normalized Haar Transform (NHT) (Wei *et al*., 2007).

In H.264/AVC video standard, the best mode decision is recommended by the Rate Distortion Optimization (RDO) function (Nie and Ma, 2002; Byung-Gyu, 2008b). Based on the RD cost, the best MB mode is determined based on minimizing the bit rate and maximizing the image quality. The RD cost can be represented as follows:

$$J_{RD} = D_{Mode|QP} + \lambda \cdot R(\text{Header} + \text{residual})_{Mode|QP} \quad (1)$$

Where:

JRD	=	Bitrate-distortion value used as a cost function
Dmode|QP	=	Sum of the Absolute Differences (SAD) or the Sum of the Squared Difference (SSD) for the given mode
λ	=	Lagrangian multiplier
R(x)	=	Bit amount for encoding
x	=	Header provides header information and residual represents residual data for the given mode (QP) at the current MB

As mentioned in the above, most of studies were focused on improving the coding efficiency and reducing the complexity of inter and intra mode selection but the studies to reduce the computational complexity of High profile tools were few. There are two integer Discrete Cosine Transform (DCT) modes in H264/AVC. Integer DCT 4×4 is basically used for all profiles and integer DCT 8×8 is used for High profile coding. Especially the transform 8×8 mode occupies large portion among many adding tools for High profile. When the 8×8 transform mode is enabled in High profile, both integer DCT 8×8 and intra 8×8 mode searches are performed to improve the coding efficiency, resulting in an increased complexity for coding, which occupies 25-30% of the total encoding time.

In this case, the Rate Distortion Optimization (RDO) function for the best mode type can be rewritten as follows using Eq. 1:

$$J_{RD} = D_{Mode|QP|DCT} + \lambda \cdot R(\text{Header} + \text{residual})_{Mode|QP|DCT} \quad (2)$$

where DCT means each DCT transform mode (Integer DCT 4×4 or integer DCT 8×8 transform).

This transform mode can also be large computational burden for encoding the HD or Ultra HD sequences. Thus, it is important part to reduce the complexity of the transform process, effectively.

We present an adaptive kernel size selection method for the transform mode in high profile to reduce the 8×8 transform mode time. The 8×8 transform mode (Integer DCT 8×8) has advantages for blocks that contain a relatively large amount of high frequencies in the residual image. We can detect these blocks using the variance of the DC coefficients after a DCT 4×4 mode decision and transform. The adaptive thresholds are also proposed based on Bayesian theory.

MATERIALS AND METHODS

Adaptive kernel size selection algorithm

Observation: As we mentioned, only the integer DCT 4×4 was used in Baseline profile of H.264/AVC. Both DCT 4×4 and DCT 8×8 are used for some block modes in Main or high profile, then more suitable DCT kernel is selected as the best kernel of these modes. It means that the texture condition which each DCT kernel is well conducted under can be different. Thus if we can distinguish the proper texture condition for each DCT kernel size, we can reduce the consuming time which is used for DCT process.

We conducted several experiments to find a proper classifier. Table 1 shows the occupation ratio for each transform mode with various Quantization Parameters (QPs). The portion of the integer DCT 8×8 is increased with a decreasing QP.

Generally, when the QP value is decreased the amounts of the high frequencies that are removed by quantization are also decreased and energy compaction can be decreased. Based on this characteristic, the 8×8 transform mode is good with blocks that have a relatively large high frequency region. We identify 16 DC coefficients after the integer DCT 4×4 mode decision for the classifier which can show the texture characteristics.

Because DC coefficients can represent the average characteristics of each 4×4 block. If the amount of high frequencies is large in the residual block, the homogeneity of the block is small. This means that the variance between the DC coefficients will also be increased.

Table 1: The occupation ratio of each transform mode with various QP values

		Integer DCT	
Sequence	QP	4×4	8×8
Rush hour (HD)	20	53.19	46.80
	24	67.88	32.11
	28	80.75	19.24
	32	89.31	10.68
	36	94.59	5.40
Tractor (HD)	20	41.39	58.60
	24	43.21	56.78
	28	61.46	38.53
	32	75.4	24.59
	36	83.76	16.23
Crew (4CIF)	20	32.37	67.62
	24	44.06	55.93
	28	61.73	36.26
	32	75.34	24.65
	36	85.46	14.53
Harbour (4CIF)	20	15.16	84.83
	24	18.1	81.89
	28	29.51	70.48
	32	46.85	53.14
	36	66.46	33.53

Figure 1 shows the average variance of 16 DC coefficients after a DCT 4×4 mode decision when each DCT transform mode is selected as the best transform mode.

As shown in Fig. 1, the average variance of 16 DC coefficients after a DCT 4×4 mode decision is quite different according to which DCT kernel is selected as the best kernel size.

From Fig. 1, we are able to observe that graphs of two DCT kernels are clearly separated. It means that the each kernel has a different distribution. To use this characteristic in this research, two parameters are defined as follows:

$$\text{Mean}\ (\delta^2_{DC\,4\times4\text{-}DCT\,8\times8})$$

is the average variance of DC coefficients after DCT 4×4 when DCT 8×8 is selected as the best transform mode.

$$\text{Mean}\ (\delta^2_{DC\,4\times4\text{-}DCT\,4\times4})$$

is the average variance of the DC coefficients after DCT 4×4 when DCT 4×4 is selected as the best transform mode.

Based on these parameters, we propose an adaptive threshloding scheme to reduce the complexity of the DCT transform process in H.264/AVC video encoding system.

The proposed DCT kernel selection algorithm: Based on the defined parameters, the proposed algorithm can be summarized as follows:

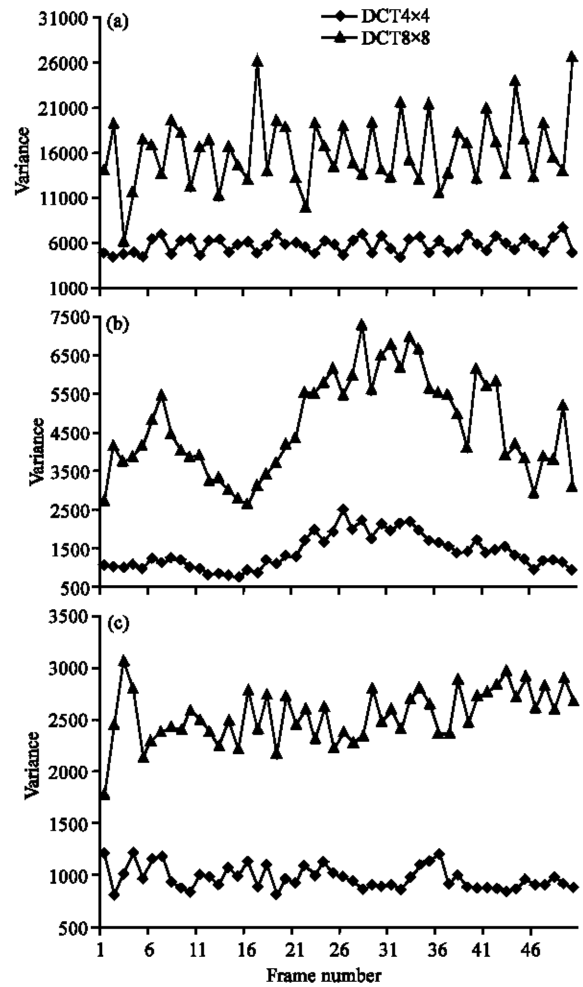

Fig. 1: The average variance of 16 DC coefficients after DCT 4×4 when each DCT mode is selected as the best transform mode. a) Harbour (4CIF), b) Pedestrian (HD) and c) Tractor (HD)

Step 1: Perform the DCT 4×4 mode and calculate the average δ^2 between the 16 DC coefficients of the current block ($\delta^2_{DC\,4\times4\text{-}current}$).

Step 2: Compare the average ($\delta^2_{DC\,4\times4\text{-}current}$) with the defined parameters and perform the followings:

- $(\delta^2_{DC\,4\times4\text{-}current}) < \text{Mean}\ (\delta^2_{DC\,4\times4\text{-}DCT\,4\times4})$

We skip execution of the transform 8×8 mode.

- $(\delta^2_{DC\,4\times4\text{-}current}) > \text{Mean}\ (\delta^2_{DC\,4\times4\text{-}DCT\,8\times8})$

We perform the 8×8 transform mode.

- $(\delta^2_{DC\,4\times4\text{-}DCT\,4\times4}) \le \delta^2_{DC\,4\times4\text{-}current} \le \text{Mean}\ (\delta^2_{DC\,4\times4\text{-}DCT\,8\times8})$

The distribution of $\delta^2_{DC\ 4\times4\text{-}DCT\ 4\times4}$ values and $\delta^2_{DC\ 4\times4\text{-}DCT\ 8\times8}$ values are assumed to be Gaussian distributions.

Let the input x be $\delta^2_{DC\ 4\times4\text{-}current}$ to calculate the probability p(x) Eq. 3 in each distribution as the following:

$$p(x) = \frac{1}{\sqrt{2\pi\delta^2}} e^{\frac{-(x-m)^2}{2\delta^2}} \quad (3)$$

Where:

x = Variable of $\delta^2_{DC\ 4\times4\text{-}DCT\ 4\times4}$

m = Mean of variable x.

Based on this probability model if the distribution of $\delta^2_{DC\ 4\times4\text{-}DCT\ 8\times8}$ has a higher probability, than $\delta^2_{DC\ 4\times4\text{-}DCT\ 4\times4}$ the 8×8 transform is performed. Otherwise, the DCT 8×8 transform mode can be omitted.

RESULTS AND DISCUSSION

Experimental results: Various MPEG standard sequences (4CIF, HD) were used to verify the performance of the proposed algorithm.

The analysis was performed with encoding frames = 100, reference picture = 1, sequence type = IPPP, QP = 24, 28 and 32 with CABAC enabled and the High profile. The software platform was JM11.0 reference software by JVT. All algorithms for comparison were run on a Pentium 4 PC hardware platform with a 2.41 Ghz core of 2 quad CPUs and 2.0 Gbytes of RAM.

We defined several measures for evaluating the performance of the proposed scheme, including the average ΔPSNR, the average ΔBits and ΔTS.

The average ΔPSNR and ΔBits show the differences in quality (average PSNR) or total bits between the proposed method and the corresponding values of the full mode search. The average PSNR is defined as:

$$\overline{PSNR} = \frac{4PSNR_Y + PSNR_{Cb} + PSNR_{Cr}}{6} \quad (4)$$

where, PSNR*y*, PSNRCb and PSNRCr are peak-to-noise ratios of the luminance and two chroma components, respectively. As performance improves this criterion becomes larger.

The ΔTS is a complexity comparison factor used to indicate the time saving when comparing to the total time of transform process at each QP, as follows:

$$\Delta TS = \frac{Time(reference) - Time(proposed)}{Time(reference)} \times 100 \quad (5)$$

Table 2: Experimental results with the proposed algorithm

Sequences	QP	ΔPSNR (dB)	ΔBits (%)	ΔTS (%)
Crew (4CIF)	24	-0.02	0.069	52.53
	28	-0.04	0.213	60.57
	32	-0.05	-0.115	57.10
Harbour (4CIF)	24	-0.01	0.018	47.35
	28	0.00	0.089	49.58
	32	-0.01	-0.032	50.44
Rush hour (HD)	24	-0.05	-0.361	58.48
	28	-0.04	-0.496	60.49
	32	-0.05	-0.499	63.38
Tractor (HD)	24	-0.01	-0.042	54.17
	28	-0.02	0.035	56.56
	32	-0.03	-0.045	65.53
Average	-	-0.03	-0.097	56.35

Where time (reference) denotes the consumed time of the whole transform process and Time (proposed) means that of the whole transform process using the suggested algorithm. This value increases with the performance speed.

The overall performance: Table 2 shows the results for the proposed algorithm. For the Harbour and Crew sequences, there was very small increment of bits while achieving a large speed-up gain of the DCT transform mode process.

In the Rush Hour sequence, the bits saving effect of >0.499% was observed. For the Tractor sequence, a speed-up gain of 65.53% was achieved with a negligible loss of quality and bit increment.

The average loss of quality in PSNR was -0.01~ -0.05 dB and the total bit increment was approximately -0.499~0.213% compared with the full mode search. The proposed scheme achieved an improvement of 47.35~65.53% in time saving for the 8×8 transform and mode decision time and a 15~18% time reduction in the total encoding time.

Comparing to the full transform process, a speed-up gain of 56.35% was observed in the average time while maintaining the loss of quality of 0.03 and bit decrement of 0.097%.

The proposed method has good performance in the larger QP values. From Table 1, we can know that the portion of the DCT 8×8 is small in the lower QP values and vice versa in higher QP values. Therefore, we can confirm that the proposed method can reduce the unnecessary 8×8 transform procedure effectively.

RDO performance: Figure 2 shows the Rate-Distortion (RD) curves for the IPPP structure derived from the full DCT mode (DCT 4×4 and DCT 8×8) and the proposed algorithm. The rate-distortion performance of the proposed method was similar to the original full JM 11.0 of the original encoder.

This means the proposed algorithm is very reliable to encode H.264/AVC video contents. Especially, the graphs were perfectly overlapped in the Harbour

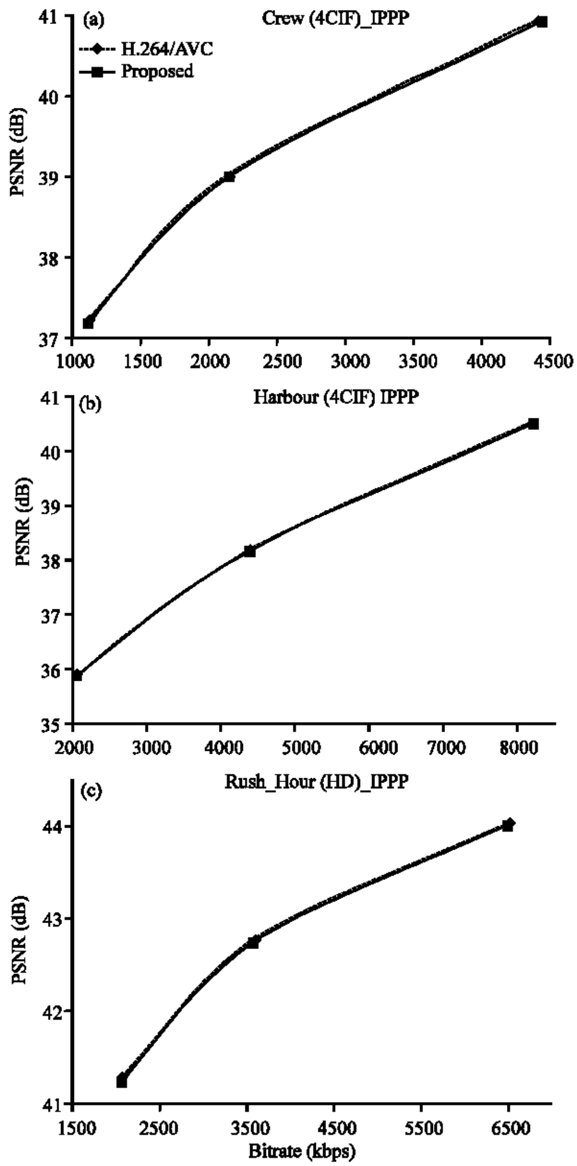

Fig. 2: RD curves for the IPPP type. a) Crew (4CIF), b) Harbour (4CIF) and c) Rush Hour (HD)

sequences. The proposed algorithm achieves the good time reduction performance with a negligible loss in quality.

In the Crew sequence, we can observe very small loss of quality. At high bit-rate area, the loss of quality becomes larger than lower bit-rate region. For the Rush Hour sequence, similar performance was achieved for all bit-rate range while obtaining the speed-up gain of up to 63.38%.

From these results, we can deduce that the proposed adaptive DCT kernel size determination algorithm is very efficient to speed-up the DCT transform process of the H.264/AVC video encoder.

CONCLUSION

We have proposed an efficient DCT kernel size selection algorithm for high profile in H.264/AVC video. The algorithm is based on the characteristic of the variance of the DC coefficient after DCT 4×4 transform procedure. An adaptive thresholding scheme was designed to decide whether performing the DCT 8×8 transform or not. Through comparative analysis, a speed-up factor of 47.35- 65.53% for IPPP sequences was verified with a negligible bit increment and a minimal loss of image quality.

ACKNOWLEDGEMENTS

This research was supported 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-2009-C1090-09020020).

REFERENCES

Byung-Gyu, K., 2008a. Fast selective-intra mode search algorithm based on adaptive thresholding scheme for H.264/AVC encoding. IEEE Trans. Circuits Syst. Video Technol., 18: 127-133.

Byung-Gyu, K., 2008b. Novel inter-mode decision algorithm based on Macroblock (MB) tracking for the P-slice in H.264/AVC video coding. IEEE Trans. Circuits Syst. Video Technol., 18: 273-279.

Cheng, Y., Z.J.G. Wmg and K. Dai, 2005. Research on intra modes for inter-frame coding in H.264. Proceedings of the IEEE 9th International Conference on Computer Supported Cooperative Research in Desi, May 24-26, Guilin China, pp: 740-740.

Choi, I., J. Lee and B Jeon, 2006. Fast coding mode selection with rate-distortion optimization for MPEG-4 part-10 AVC/H.264. IEEE Trans. Circuits Syst. Video Technol., 16: 1557-1561.

Crecos, C. and M.Y. Yang, 2005. Fast inter mode prediction for P-slices in the H264 video coding standard. IEEE Trans. Broadcasting, 51: 256-263.

Jing, X. and L.P. Chau, 2004. Fast approach for H.264 inter mode decision. Elect. Lett., 40: 1050-1052.

Kim, B.G., J.H. Kim and C.S. Cho, 2006. A fast intra skip detection algorithm for H.264/AVC video coding. ETRI J., 28: 721-731.

Kim, C. and C.C.J. Kuo, 2007. Feature-based intra-/inter coding mode selection for H.264/AVC. IEEE Trans. Circuits Syst. Video Technol., 2: 441-453.

Nie, Y. and K.K. Ma, 2002. Adaptive rood pattern search for fast block-matching motion estimation. IEEE Trans. Image Proces., 11: 1442-1449.

Pan, F., X. Lin, S. Rahardja, K.P. Lim, Z.G. Li, D. Wu and S. Wu, 2005. Fast mode decision algorithm for intraprediction in H.264/AVC video coding. IEEE Trans. Circuits Syst. Video Technol., 15: 813-822.

Wei, Z., H. Li and K.N. Ngan, 2007. An efficient intra mode selection algorithm for H.264 based on fast edge classification. IEEE International Symposium of Circuits System, (May 27-30), Hong Kong, 3630-3633.

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.