Fast Mode Selection to Reduce the Encoding Complexity of H.264/AVC

Bin Zhan, Baochun Hou and Reza Sotudeh
School of Electronic, Communication and Electrical Engineering
University of Hertfordshire
Hatfield, United Kingdom
{b.zhan, b.hou, r.sotudeh}@ herts.ac.uk

***Abstract*—Fast mode selection is one of the hottest research topics for H.264/AVC finalized in 2003. Since the computational complexity is still a major challenge for real-time applications, simplification of mode selection is therefore very important. A fast mode selection algorithm is proposed in this paper. By removing low-probability modes according to the correlative characteristic in macroblock mode selection and the statistical characteristic of sub-macroblock mode selection, the calculation of mode selection process is significantly reduced. Based on the experimental results, the encoding time can be saved 55.91% on average, and in most cases the degradation of video quality is limited in 0.15dB with less than 2% increase of bitrate.**

I. INTRODUCTION

H.264/AVC is a joint video coding standard finalized by ITU-T Video Coding Experts Group and ISO/IEC Moving Picture Experts Group in 2003 [1]. Evolved from the previous standards, H.264/AVC can provide the outstanding encoding performance [2]. The variable block size based rate distortion optimal (RDO) mode selection is one of the distinct techniques exploited in the new standard, which enables the excellent encoding efficiency in terms of video quality and compression ratio. However, this improvement is achieved at the expense of extremely increased complexity of calculation, and that is one of the most challenging problems in real-time and low power consumption implementations. Since a major application for H.264/AVC is the video communication over Internet and wireless environment, the fast mode selection becomes one of the hottest research topics in academies and industries. In the past several years, a number of efforts have been made by researchers. In [3], Grecos and Yang proposed a fast mode selection scheme for inter-prediction which exploits neighbourhood information and a set of constraints to enhance the Skip mode selection and other mode selection separately. Pan et al. introduced an approach to reduce the intra candidate modes according to a pre-established local edge direction histogram in [4]. J. F. Wang et al. in [5] used the dominant edge strength to estimate the possible directions to improve the intra mode selection efficiency. H. Wang et al. optimized the mode selection by early terminating unnecessary modes according to the thresholds based prediction in [6]. Recently, we also presented a fast mode selection algorithm based on the statistical characteristic of macroblock mode selection and adaptive adjustment [7]. In this paper, a fast mode selection algorithm was proposed to speed up the mode selection. By reducing the candidate modes according to the spatial and temporal correlation between macroblocks and incorporating with fast sub-macroblock mode selection scheme, the time saving is significant without obvious loss of encoding efficiency.

The rest of this paper is divided into four sections: a brief introduction of variable block size based mode selection is given in section II. In section III, the correlation in macroblock mode selection and the statistical characteristic of sub-macroblock mode selection are investigated, and the detailed algorithm is also stated in this section. Section IV lists the experimental results, and the following section is the conclusion of this paper.

II. VARIABLE BLOCK SIZE BASED MODE SELECTION

Compared with the unique block size employed in the former standards, each macroblock is partitioned into more block sizes in H.264/AVC in addition to use the 16×16 macroblock. Correspondingly, seven macroblock modes are supported in the new standard: SKIP, Inter 16×16, Inter 16×8, Inter 8×16, Inter P8×8, INTRA 16×16 and INTRA 4×4, and each Inter P8×8 mode can be further divided into four sub-macroblock modes: Inter 8×8, Inter 8×4, Inter 4×8 and Inter 4×4. For each macroblock, all candidate modes are tried and the one leads the minimal rate distortion (RD) cost is selected as the best suitable mode. The Lagrangian function (1) is frequently used to calculate the RD cost.

$$J_{MODE} = Distortion + \lambda_{MODE} \cdot Rate \tag{1}$$

In (1), J_{MODE} gives the Lagrangian cost of current macroblock for the candidate mode *MODE*. *Distortion* is contributed by the sum of the squared differences (SSD) between the current macroblock *s* and the reconstructed macroblock *c*. λ_{MODE} stands for the Lagrangian multiplier related to the quantization

978-1-4244-1684-4/08/$25.00 ©2008 IEEE

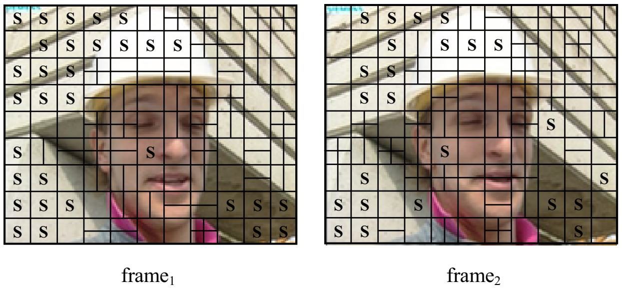


Figure 1. Mode selection in adjacent two frames for Foreman.

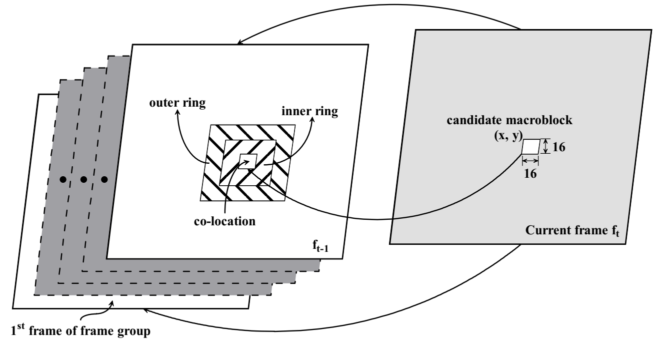


Figure 2. Location of macroblocks in reference frames.

parameter (QP). *Rate* represents the bit cost for encoding the additional information such as motion vectors. For each macroblock, the RD cost needs to be calculated for every candidate mode. It is therefore an obvious time consumption process of computation to compare the entire modes.

III. Fast Mode Selection Algorithm

The proposed algorithm consists of two schemes: the fast macroblock mode selection and the fast sub-macroblock mode selection:

A. Fast Macroblock Mode Selection

One segment of video sequence possesses strong spatial and temporal correlation between adjacent frames, and this characteristic is employed in the intra and inter predictions to remove the redundancy from the video sequence. From the intensive experiments the modes selected in the continued frames also present considerable correlation in both spatial and temporal domains. An example is given in Fig. 1. The best modes selected for two successively encoded frames ($frame_1$ is previously encoded and used as the reference for $frame_2$) are illustrated in the figure, in which *S* indicates the SKIP mode. For macroblock *(x, y)* in $frame_2$, if collected the best modes of its co-located and neighbouring macroblocks in $frame_1$ to generate a new candidate mode group $G_{(x, y)}$, the actual best mode of *(x, y)* is frequently one member of $G_{(x, y)}$. Furthermore, the number of modes in $G_{(x, y)}$ is generally less than the entire modes specified in the standard, the encoding time can be reduced by exploiting the $G_{(x, y)}$ instead of the entire mode group, and the percentage of time saving is depended on the number of modes included in $G_{(x, y)}$. Additionally, in order to keep the relevant information, the video sequence is divided into a series of frame groups (*FG*, the capacity of a *FG* equals to the frame rate). The first frame of *FG* is encoded by adopting the entire mode group, and the best modes are recorded to update $G_{(x, y)}$ for the rest frames in *FG*.

Based on the investigation given above, a fast macroblock mode selection scheme is then proposed. The fast scheme gives two different processing methods according to the location of macroblocks, separately. The definitions given in Fig. 2 helps to describe the algorithm. The neighbouring macroblocks that mostly close to the co-located macroblock in the reference frame form the inner ring macroblocks, and one circle of macroblocks surround the inner ring form the outer ring macroblocks. The proposed scheme is available except for the first frame in each *FG*.

1) Macroblocks located not on the edge of frame

For the current candidate macroblock *(x, y)*, if its outer ring macroblocks are complete, the mode group $G_{(x, y)}$ consists of the best modes of co-located, inner ring and outer ring of macroblocks in the most previously encoded frame and the first frame of current *FG*. Otherwise, the mode group $G_{(x, y)}$ consists of the best modes of co-located and inner ring of macroblocks in the most previously encoded frame and the first frame of current *FG*.

2) Macroblocks located on the edge of frame

For the current candidate macroblock *(x, y)*, both its inner ring and outer ring of macroblocks are not complete, which means that there are insufficient informations to generate the mode group $G_{(x, y)}$. If simply adopting the method introduced in the last paragraph, the actual best mode could be excluded from $G_{(x, y)}$, and the encoding efficiency could be obviously decreased. A SKIP mode based method is then used for those macroblocks:

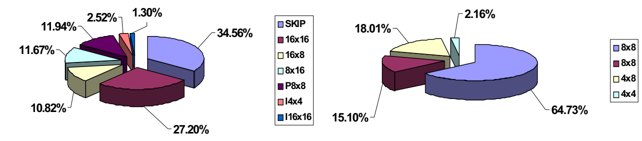
- If the best mode of co-located macroblock in the most previously encoded frame is SKIP, only SKIP mode and Inter 16×16 mode are contained in $G_{(x, y)}$.
- Otherwise, the entire candidate modes are contained in $G_{(x, y)}$.

B. Fast Sub-macroblock Mode Selection

In the section III.A of this paper, we introduced the fast mode selection scheme at the macroblock level. A sub-macroblock mode selection scheme is therefore proposed to further increase the encoding speed.

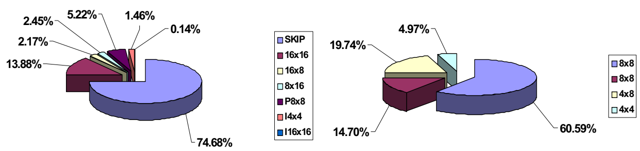
According to the two examples illustrated in Fig. 3 and Fig. 4, some candidate modes at the sub-macroblock level only possess very small percentage. Experiments also revealed that the encoding time could be saved if we remove some low-probability sub-modes. In [7], we indicated the percentage of modes employed in encoding process follows certain pattern for a given video sequence, and the encoder needs a learning period to obtain the statistic of mode distribution in order to remove the modes. In most cases, after fifteen P-frames have been encoded, the statistical characteristic tends to be stable as shown in Fig. 5. Therefore, we give the detailed scheme for sub-macroblock mode selection as follows:

- For the first 15 P-frames, encode them by exploiting all



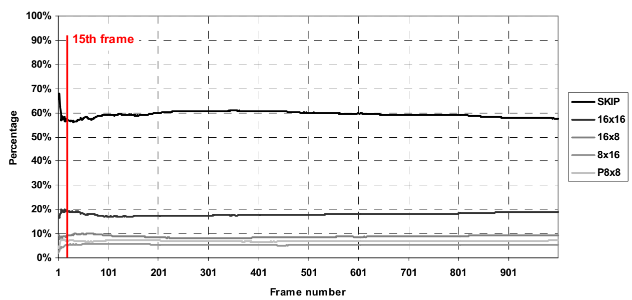
a). macroblock modes b). sub-macroblock modes

Figure 3. Percentage of selection of each mode for Foreman.

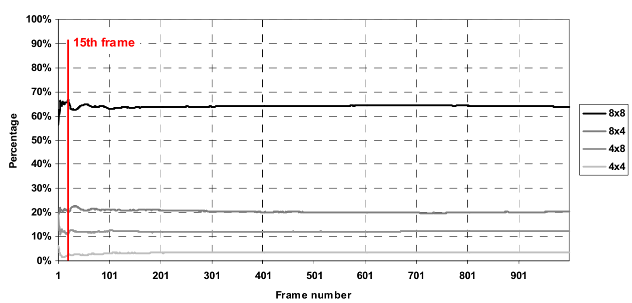


a). macroblock modes b). sub-macroblock modes

Figure 4. Percentage of selection of each mode for News.



a). macroblock modes



b). sub-macroblock modes

Figure 5. Percentage of selected modes vs. frame number for Highway (1000 QCIF).

sub-macroblock modes as specified in H.264/AVC, and record the number of selection for each sub-macroblock mode.

- After the 15th P-frame has been encoded, calculate the percentage (p_x) of sub-macroblock mode x.
  - x=1: Inter 8×8
  - x=2: Inter 8×4
  - x=3: Inter 4×8
  - x=4: Inter 4×4
- If p_x is less than the mode elimination threshold TH (TH = 5% yields better performance [7]), disable sub-macroblock mode x.
- From the 16th P-frame, only the sub-macroblock mode(s) remained in the last step will be available for Inter P8×8 macroblock mode.

IV. Experimental Results and Analysis

To evaluate the proposed fast mode selection algorithm, the experiments were performed on the JM CODEC Version 10.1 [8], and the experimental environments are listed as following:

- Baseline profile
- ±16 for motion estimation search range
- IPPP sequence type and 300 frames in each sequence
- Full search motion estimation scheme
- One reference frame for motion estimation
- Pentium IV 2.66 GHz PC with 2GB memory

Ten commonly used test video sequences were included in the experiments and divided into three groups according to the characteristics of contents:

Distance:	{bridge-far (Q), bridge-close (Q)}
Background:	{foreman (Q), carphone (Q), Claire (Q)}
Movement:	{mobile (Q), coastguard (Q), container(C), highway (C), Stefan(C)}

where Q indicates QCIF, and C indicates CIF. $\Delta Time$ (%), $\Delta PSNR$ (dB) and $\Delta Bitrate$ (%) are three parameters used to point out the variation of encoding performance under the proposed fast mode selection algorithm compared with the exhaustive algorithm of JM 10.1, and their definitions are given in (2), (3) and (4):

$$\Delta Time = \frac{Time_{proposed} - Time_{JM}}{Time_{JM}} \times 100\% \tag{2}$$

$$\Delta PSNR = PSNR_{proposed} - PSNR_{JM} \tag{3}$$

$$\Delta Bitrate = \frac{Bitrate_{proposed} - Bitrate_{JM}}{Bitrate_{JM}} \times 100\% \tag{4}$$

The experimental results are given in Table I – Table IV based on the different QP values: 28, 32, 36 and 40. From the results, the encoding time can be significantly reduced 55.91% on average. In most cases, the degradation of video quality is limited in 0.15 dB with less than 2% increase of bitrate. The RD curves of the sequences (Foreman, Coastguard, Stefan and Highway) are illustrated in Fig. 6. Furthermore, Fig. 7 shows the comparison of time saving rate between the proposed algorithm and Grecos's algorithm [3], which is under the similar experimental environments, and the figure indicates the proposed fast mode selection algorithm gives an impressive improvement.

V. Conclusions

A fast algorithm is proposed to reduce the computational complexity incurred by the variable block size based mode selection exploited in H.264/AVC. Correlation in macroblock mode selection and statistical characteristic of sub-macroblock mode selection are combined to remove some candidate modes with low-probability. According to the experimental

TABLE I. EXPERIMENTAL RESULTS (QP=28)

Sequence	ΔTime (%)	ΔPSNR (dB)	ΔBitrate (%)
bridge-far	-82.81	-0.04	1.08
bridge-close	-69.26	-0.03	0.35
foreman	-37.60	-0.05	0.29
carphone	-42.07	-0.11	0.40
Claire	-71.29	-0.13	1.65
mobile	-39.61	-0.06	0.12
coastguard	-38.58	-0.09	0.81
container(C)	-61.05	-0.07	0.92
highway(C)	-52.59	-0.12	1.74
Stefan(C)	-32.88	-0.03	0.97

TABLE II. EXPERIMENTAL RESULTS (QP=32)

Sequence	ΔTime (%)	ΔPSNR (dB)	ΔBitrate (%)
bridge-far	-82.19	0	0
bridge-close	-77.61	-0.05	1.28
foreman	-42.81	-0.05	1.52
carphone	-44.53	-0.13	0.84
Claire	-67.30	-0.17	1.35
mobile	-42.11	-0.05	0.08
coastguard	-40.84	-0.14	1.02
container(C)	-67.05	-0.07	0.92
highway(C)	-57.89	-0.11	1.38
Stefan(C)	-29.72	-0.05	0.63

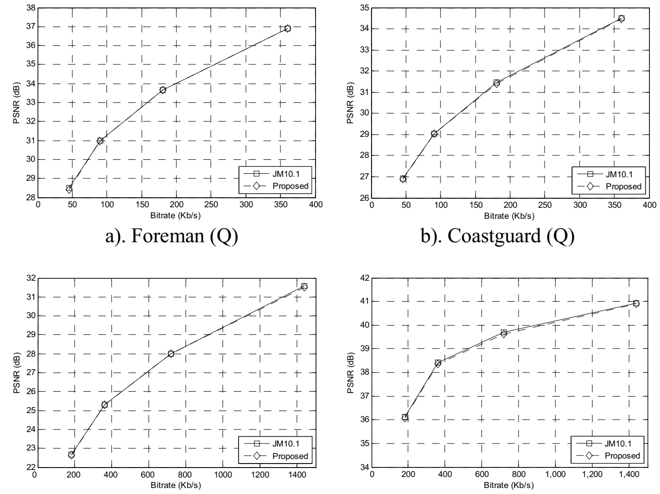
TABLE III. EXPERIMENTAL RESULTS (QP=36)

Sequence	ΔTime (%)	ΔPSNR (dB)	ΔBitrate (%)
bridge-far	-80.38	0	0
bridge-close	-80.95	-0.04	1.54
foreman	-41.31	-0.09	1.99
carphone	-52.17	-0.10	0.63
Claire	-70.82	-0.16	2.04
mobile	-38.38	-0.06	0.22
coastguard	-45.23	-0.06	1.23
container(C)	-68.77	-0.06	0.89
highway(C)	-68.94	-0.18	1.67
Stefan(C)	-30.47	-0.07	0.55

TABLE IV. EXPERIMENTAL RESULTS (QP=40)

Sequence	ΔTime (%)	ΔPSNR (dB)	ΔBitrate (%)
bridge-far	-81.07	0	0
bridge-close	-77.96	-0.03	1.08
foreman	-47.34	-0.12	0.76
carphone	-56.95	-0.15	0.61
Claire	-66.87	-0.15	0.80
mobile	-41.03	-0.07	-0.28
coastguard	-50.63	-0.09	-1.22
container(C)	-64.99	-0.05	0.24
highway(C)	-65.85	-0.21	-0.07
Stefan(C)	-26.49	-0.08	0.49

results, the encoding time can be reduced up to 82.81%, and the degradation of encoding efficiency is negligible.



a). Foreman (Q) b). Coastguard (Q)

c). Stefan (C) d). Highway (C)

Figure 6. RD curves.

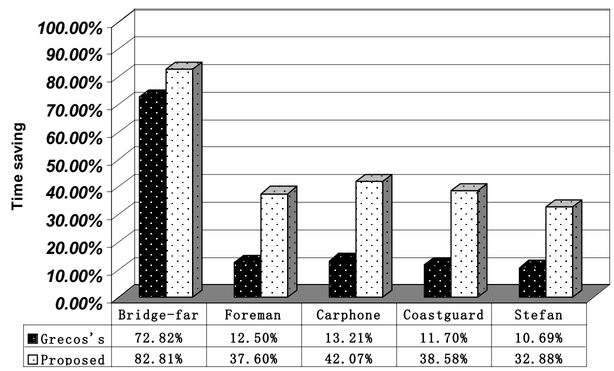


Figure 7. Comparison between Grecos's fast mode selection algorithm and proposed mode selection algorithm.

REFERENCES

[1] "Draft ITU-T Rec. and FDIS of Joint Video Spec. (H.264 | ISO/IEC 14496-10 AVC)," JVT of MPEG and VCEG, Doc. JVT-G050rl, May 2003.

[2] T. Wiegand, G. J. Sullivan, G. Bjntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," IEEE Trans. Circuits and Systems for Video Technology, vol. 13, pp.560-576, July 2003.

[3] C. Grecos and M. Y. Yang, "Fast inter mode prediction for P slices in the H264 video coding standard," IEEE Trans. Broadcasting, vol.51, pp. 256-263, June 2005.

[4] F. Pan, X. Lin, S. Rahardja, K. P. Lim, Z. G. Li, D. Wu, and S. Wu, "Fast mode decision algorithm for intraprediction in H.264/AVC video coding," IEEE Trans. Circuits and Systems for Video Technology, vol.15, pp. 813-822, July 2005.

[5] J. F. Wang, J. C. Wang, J. T. Chen, A. C. Tsai, and A. Paul, "A novel fast algorithm for intra mode decision in H.264/AVC encoders," Proc. IEEE ISCAS 2006, pp. 3498-3501, May 2006.

[6] H. Wang, S. Kwong, and C. W. Kok, "An efficient mode decision algorithm for H.264/AVC encoding optimization," IEEE Trans. Multimedia, vol.9, pp. 882-888, June 2007.

[7] B. Zhan, B. Hou, and R. Sotudeh, "A novel fast inter mode decision algorithm based on statistic and adaptive adjustment for H.264/AVC," Proc. IEEE SoftCOM 2007, September 2007.

[8] JVT reference software unofficial version JM10.1, http://iphome.hhi.de/suehring/tml/download/jm10.1.zip.