FAST INTER MODE DECISION ALGORITHM BASED ON MACROBLOCK AND MOTION FEATURE ANALYSIS FOR H.264/AVC VIDEO CODING

Yiqing HUANG, Qin LIU, and Takeshi IKENAGA

Information, Productions and System LSI, Waseda University
2-7, Hibikino, Wakamatsu, Kitakyushu, 808-0135, Fukuoka, Japan
phone: + (81-90-1191-0593), fax: + (81-93-692-5319), email: koisei@aoni.waseda.jp

ABSTRACT

One fast inter mode decision algorithm is proposed in this paper. The whole algorithm is divided in to two stages. In the pre-stage, by exploiting spatial and temporal information of encoded macrobocks (MBs), a skip mode early detection scheme is proposed. The homogeneity of current MB is also analyzed to filter out small inter modes in this stage. Secondly, during the block matching stage, a motion feature based inter mode decision scheme is introduced by analyzing the motion vector predictor's accuracy, the block overlapping situation and the smoothness of SAD (sum of absolute difference) value. Moreover, the rate distortion cost is checked in an early stage and we set some constraints to speed up the whole decision flow. Experiments show that our algorithm can achieve a speed-up factor of up to 53.4% with negligible bit increment and quality degradation.

1. INTRODUCTION

The latest H.264/AVC video coding standard provides us with superior coding performance by using many new techniques. However, the negative side is that the computation complexity is also increased dramatically [1]. In H.264, besides skip mode, there are 7 block modes (as shown in Fig. 1) for inter prediction and 9 modes for 4×4 intra prediction and 4 modes for 16×16 intra prediction. The encoding process will loop all these modes and select one with the minimum compression cost. When rate distortion is incurred, all the prediction modes will be involved in a real encoding process. So, the complexity is insurmountable considering the real-time application.

In [2], a fast intra prediction algorithm is proposed, which greatly fastens intra prediction process while still keep the quality. However, the decision on inter modes is more complicated compared with intra modes. It is because the motion estimation (ME) process adopts block matching on the plane of both current image and reference image, which incurs huge calculation on all the candidate points within the search window. The split, occlusion and fast motion in the moving video increase the ratio of temporal feature among frames, which makes it almost impossible to make a pre-decision on inter prediction.

Many works have been done to solve the problem. In [3], a pre-encoding scheme is proposed, which abstracts a down sampled small image and restrict the inter block modes within a small subset. The literature [4] tries the way of mean of absolute frame difference (MAFD) to filter out unpromising inter modes. In [5], an adaptive mode decision process based on all-zero coefficients block is proposed. Literature [6][7] focuses on the optimization of early skip mode decision to release complexity of inter mode decision. However, the idea of introducing pre-processing in [3] and [4] will intensify the computation burden of whole encoding system. With the expansion of image size, for example HDTV application, the 1/2 down sampling or the MAFD calculation of the original frame will increase power dissipation and system latency dramatically. As for all-zero block and skip mode early detection based algorithms [6][7], there exist obvious limitations. With several foreground objects moving irregularly on the complicated background, or the decrease of quantization parameter, the ratio of all-zero block and skip modes will decrease significantly, which deteriorates the efficiency of inter mode filtering. In [8], a very fast mode decision algorithm is proposed, which dramatically reduces complexity for both low-motion and high-motion sequences. However, the compression capability is worsened obviously since the bit rate increase is quite large. In our paper, we prefer to solve the inter mode decision in several stages. Firstly, in the pre-stage (before ME starts), the homogeneity of current macroblock (MB) and the features of encoded MBs on both current and previous frames are inspected to detect skip mode and filter out unpromising modes. Secondly, during ME process, the motion information is collected to filter out unpromising modes. We focus on the information of motion vector predictor's accuracy, the block overlapping situation , rate distortion cost and SAD's smoothness.

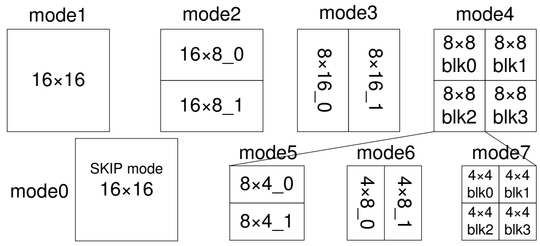


Figure 1: Inter Block Modes in H.264/AVC

The rest of paper is as follows. In section 2, the pre-stage inter mode decision scheme is described. Section 3 gives out a motion feature based inter mode decision algorithm which uses information obtained during ME process. The overall algorithm and experiments results are shown in section 4. This paper concludes with section 5.

2. PRE-STAGE INTER MODE DECISION SCHEMES

2.1 MV Oriented Spatial-temporal Inter Mode Check

In the conventional video sequences, spatial and temporal redundancy always exist within frame or between frames. In this paper, we proposes a spatial-temporal skip mode early detection scheme which is applied before encoding process (named pre-stage scheme).

Since the encoding process is executed in a raster scan order, the only spatial information we can use is from the encoded MBs. Therefore, as shown in Fig.2(a), the upper-left MB (UL.MB), left MB (L.MB) and upper MB (U.MB) of current MB (Cur.MB) are used for mode check. However, because only three MBs provide the mode information, the efficiency and correctness are quite limited. Therefore, the temporal information is also added. Besides co-located MB (Co.MB), there are 8 MBs around Co.MB in the previous frame and we classify all these MBs into three categories, as shown in Eq.1 to Eq.3. The Co.MB is the only element in the C0 category. The left-up (LU.MB), right-up (RU.MB), bottom-left (BL.MB) and bottom-right (BR.MB) MBs are ascribed to C1 category. As for C2 category, it includes four MBs in cross direction such as U.MB, L.MB, right MB (R.MB) and bottom MB (B.MB) around Co.MB. Our temporal mode check algorithm only depends on the dominant category of C1 and C2. Obeying the rule that the motion is continuous in the succeeding frames, we use the motion vector (MV) of MBs in C1 and C2 to decide dominant category. As shown in Eq.4 to Eq.6, we calculate the delta MV between C1 and C0 ($\Delta MV(C1,C0)$) based on the accumulation of absolute MV difference in x ($\Delta MV_x(C1,C0)$) and y ($\Delta MV_y(C1,C0)$) direction. The delta MV between C2 and C0 is calculated based on the same principle. As shown in Eq.7, the category with minimum MV difference will be chosen as candidate category. The pseudo codes of our spatial-temporal algorithm is shown in Fig.3(a). It mean that before ME, we apply mode check based on spatial and temporal information. The MV difference is used as a criterion to select candidate category (C') for temporal mode check. If all the modes of MBs in spatial (LU.MB, U.MB and L.MB in Fig. 2(a)) and temporal category (C' and C0) are skip modes (*mode*0), *mode*0 is selected as best inter mode. Otherwise, full modes of H.264/AVC are enabled.

$$C0 \in \{Co.MB\} \tag{1}$$

$$C1 \in \{LU.MB, RU.MB, BL.MB, BR.MB\} \tag{2}$$

$$C2 \in \{U.MB, L.MB, R.MB, B.MB\} \tag{3}$$

$$\Delta MV(C1,C0) = \Delta MV_x(C1,C0) + \Delta MV_y(C1,C0) \tag{4}$$

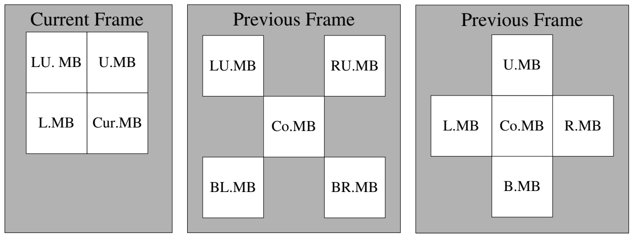
$$\Delta MV_x(C1,C0) = \Sigma|C1\{MV_x\} - C0\{MV_x\}| \tag{5}$$

$$\Delta MV_y(C1,C0) = \Sigma|C1\{MV_y\} - C0\{MV_y\}| \tag{6}$$

$$\begin{cases} \Delta MV(C1,C0) < \Delta MV(C2,C0), & C1 \text{ is dominant} \\ \Delta MV(C2,C0) \le \Delta MV(C1,C0), & C2 \text{ is dominant} \end{cases} \tag{7}$$

2.2 Edge Gradient Based Inter Mode Filtering

The edge detection is another useful technique in both image processing and pattern recognition field. In [2], it uses Sobel edge operator to obtain candidate Intra modes. In fact, the same method can also be extended into inter mode filtering. Fig. 4 demonstrates the mode distribution among different sequences and Fig. 5 is the corresponding edge gradient histogram of each frame. Sobel operator is used to get the gradient of each frame. It is shown that the gradient distribution between mobile_qcif and container_qcif is quite big, which is



(a) Spatial Check (b) Temporal MB in C1 (c) Temporal MB in C2

Figure 2: Spatial-temporal Skip Mode Check

```
Spatial MB mode check
Loop MBs in C1 category
    Calculate delta MV (C1, C0)
End Loop
Loop MBs in C2 category
    Calculate delta MV (C2, C0)
End Loop
Decide temporal category (C' + C0)
If MBs in Spatial and Temporal category
  are all mode0
    Only mode0 for Cur.MB
Else
    Full modes of H.264/AVC
```

(a) Spatial-temporal Skip Mode Check

```
Loop blk8x8_i of current MB, i = [0, 3]
    Edge gradient analysis of blk8x8_i
    If homogenous blk8x8_i
        Discard 8x4, 4x8, 4x4 modes
End Loop
If all blk8x8_i are all homogenous
    Disable 8x8 inter modes
```

(b) Edge Based Inter Mode Filtering

Figure 3: Pseudo Codes of Pre-Stage Inter Mode Decision

in accordance with subjective observation. For mode distribution, the proportion of mode above 8×8 in container_qcif is much more than that of mobile_qcif. The situations of tempete_qcif and grandma_qcif are similar, where grandma_qcif is more favorable to big inter modes. So, the decrease of gradient in image will increase the possibility of big inter modes in the final mode decision stage.

To be compatible with H.264/AVC encoding flow, we apply Sobel edge detection based inter mode filtering on the MB level. Specifically, before block matching process, we analyze current MB and obtain the edge gradient of each pixel based on Eq. 8 to Eq. 10, where $P(m,n)$ is the pixel value on coordinate (m,n) of current MB. The G_x and G_y are the gradients in horizontal and vertical direction respectively. We simply sum up these two gradient to get $G(m,n)$ as gradient of $P(m,n)$.

In the H.264/AVC standard, the block matching is executed from *mode*1 to *mode*7 sequentially. Based on this mechanism, we set a threshold on each of the four 8×8 block (blk8x8_0 to blk8x8_3) within current MB. As shown in Eq. 11, if the gradient of every pixels within one 8×8 block (blk8x8_i, i∈0,1,2,3) is within a predefined threshold, this 8×8 block is regarded as homogeneous (homo) sub-block. Otherwise, it is an edge 8×8 block. The edge based inter mode filtering algorithm is shown in Fig.3(b). For homogenous 8×8 block, small inter modes (*mode*5 to *mode*7) are removed before ME process. If all the four 8×8 blocks are homogenous ones, even the *mode*4 inter mode is filtered.

As for the threshold setting, it is always a trade-off between quality and complexity. The prediction error e in block matching process can be assumed as a jointly Gaussian source with zero mean and variance σ^2. According to [9], the distortion of quantization D is approximated as $QP^2/3$, where QP is the quantization parameter. So, the rate distortion function [10] $R(D)$ can be represented as Eq. 12. It means that when D is bigger than σ^2, the related rate distortion approaches to zero. This conclusion is in accor-

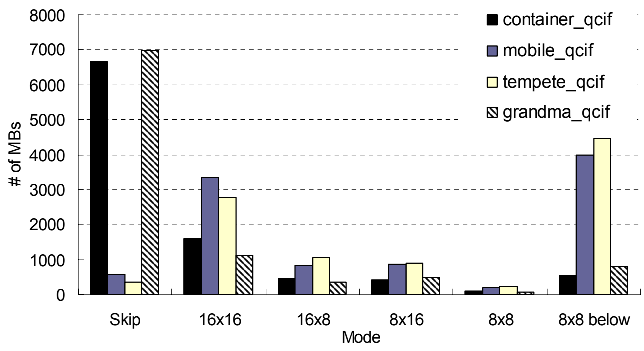


Figure 4: Inter Mode Distributions

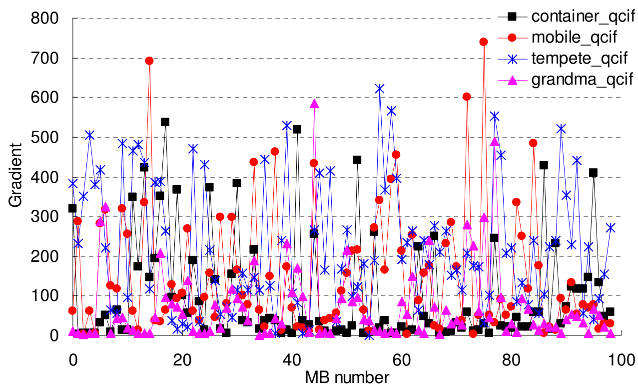


Figure 5: Gradient Distributions of 20th Frame

dance with QP setting in H.264 encoding system. With the increase of QP, the smoothness of reconstructed frames is increased, which results in decline of image's details. The related residue is also decreased or vanished. Thus, from theoretical analysis of [9] and [10], we can simply set threshold as linear relationship with QP value. With exhaustive experiments, the Thr_G in Eq. 11 is defined as $4\times$QP to balance the quality and complexity reduction. In the following sections, the related thresholds are also set linearly with QP.

$$\begin{aligned} G_x(m,n) = |P(m-1,n-1)+2P(m-1,n) \\ +P(m-1,n+1)-P(m+1,n-1) \\ -2P(m+1,n)-P(m+1,n+1)| \end{aligned} \quad (8)$$

$$\begin{aligned} G_y(m,n) = |P(m-1,n-1)+2P(m,n-1) \\ +P(m+1,n-1)-P(m-1,n+1) \\ -2P(m,n+1)-P(m+1,n+1)| \end{aligned} \quad (9)$$

$$G(m,n) = G_x(m,n)+G_y(m,n) \quad (10)$$

$$P(m,n)\in blk8x8\_i, i\in\{0,3\}$$
$$\begin{cases} G(m,n) < Thr\_G, & homo\ \ blk8x8\_i \\ otherwise, & edge\ \ blk8x8\_i \end{cases} \quad (11)$$

$$R(D) = \begin{cases} \frac{1}{2}log\frac{\sigma^2}{D}, & 0\le D\le \sigma^2 \\ 0, & D>\sigma^2 \end{cases} \quad (12)$$

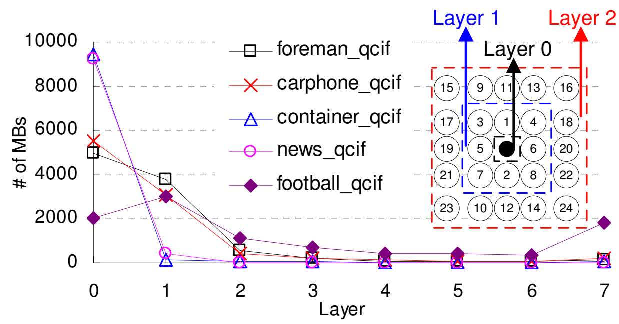


Figure 6: BIP Distribution of 16×16 Mode in 100 Frames

3. MOTION FEATURE BASED FAST INTER MODE DECISION SCHEMES

In the pre-stage, the unpromising inter modes are filtered before ME starts. However, as mentioned above, the reduction of complexity is quite limited due to the motion feature of MB. In fact, during the ME procedure, it is also possible to skip unnecessary inter modes. The more motion information we obtained, the more we can narrow down the candidate modes.

3.1 MVP Accuracy and Block Overlapping Analysis

In H.264/AVC, the block matching process starts in the motion vector predictor (MVP), which is obtained by the neighboring coded MBs. For sequence with smooth or regular motion, the prediction of start point is very accurate. Fig. 6 shows the distribution of best integer point (BIP) of 16×16 mode among typical clips. The search window is divided into several layers. The layer 0 is the MVP point while layer 1 indicates the 8 points around MVP. The meaning of other layers can be traced by analogy. It is shown that large proportion of BIP are located in MVP even for foreman_qcif and carphone_qcif. The high accuracy in MVP also indicates that the current MB is seldom split into small blocks. Since small modes is easily to be trapped into local minimum, we only use the information of 16×16 mode. As shown in Eq. 13, the motion information of 16×16 mode is analyzed after it's ME search. If the MV of 16×16 mode ($MV\_16\times16$) is the same as its own MVP ($MVP\_16\times16$) and its motion cost ($mcost$) is less than Thr_MVP (20×QP), then it is regarded as big mode MB and $mode4$ to $mode7$ are discarded. For the rest MBs, whose best 16×16 MVs are not MVPs, we further analyze the motion information after $mode3$'s block matching. The related mode decision criterion is shown in Eq.14 to Eq.16. It means that when the MVs of block0 in $mode2$ and $mode3$ (16×8_0 and 8×16_0 in Fig. 1) are the same with MVs of $mode1$; and the MVs of block1 in $mode2$ and $mode3$ are only 8 pixel displacement in x or y direction. Then the previous three inter modes are overlapped each other, which indicates that current inter modes are well enough to express the motion trend. In this case, ME of $mode5$ to $mode7$ is bypassed. The $mode4$ is remained to keep the quality.

3.2 Smoothness of Sum of Absolute Difference (SAD)

The SAD value which is obtained after ME is another useful information. With approaching to the potential best point, the SAD value decreases gradually, which leads to small motion bits. On the contrary, the occurrence of big SAD value can indicate the necessity of ME on further small modes, which results in split of MB. In this paper, we fully utilize SAD information to guide mode decision process. During IME on 16×16 mode, the four 8×8 SAD blocks is recorded, namely left-up 8×8 SAD (SAD8×8_LU), right-up 8×8 SAD (SAD8×8_RU), bottom-right 8×8 SAD (SAD8×8_BR), and bottom-left one (SAD8×8_BL). If Eq. 17 to Eq. 20 are all satisfied, which means that the distribution of SAD value is quite smooth. So, further process on small modes is rarely needed (*mode*5 to *mode*7 are discarded). When any of Eq. 17 to Eq. 20 is dissatisfied, we skip *mode*2, *mode*3 and directly turn to *mode*4 to *mode*7 of ME process. The Thr_SAD is set as 15×QP based on our experiments.

$$\begin{cases} MV\_16\times16 = MVP\_16\times16 \\ mcost\_16\times16 \leq Thr\_MVP \end{cases} \tag{13}$$

$$MV\_16\times16 = MV\_16\times8\_0 = MV\_8\times16\_0 \tag{14}$$

$$MV_x\_16\times16 = MV_x\_16\times8\_1 = MV_x\_8\times16\_1 - 8 \tag{15}$$

$$MV_y\_16\times16 = MV_y\_8\times16\_1 = MV_y\_16\times8\_1 - 8 \tag{16}$$

$$|SAD8\times8\_LU - SAD8\times8\_RU| < Thr\_SAD \tag{17}$$

$$|SAD8\times8\_BL - SAD8\times8\_BR| < Thr\_SAD \tag{18}$$

$$|SAD8\times8\_LU - SAD8\times8\_BL| < Thr\_SAD \tag{19}$$

$$|SAD8\times8\_RU - SAD8\times8\_BR| < Thr\_SAD \tag{20}$$

3.3 Rate Distortion Cost Analysis on Big Inter Modes

In the high complexity mode of H.264/AVC, after ME and intra prediction loop over all inter and intra modes, the rate distortion (RD) costs of each mode are checked exhaustively by minimizing the Lagrangian function, as shown in Eq. 21. The *SSD* is sum of squared difference between original source MB (*s*) and its reconstructed one (*c*). The *R* represents the rate after quantization and λ_{mode} is the Lagrange multiplier. It is shown that all the factors are related with *QP* and mode which is decided by inter and intra predictions. To ensure that best mode is not missed, the whole encoding process causes huge computation resources among several exhaustive loops. In our paper, the RD cost analysis is involved in the inter prediction process. Specifically, at the end of ME on *mode*1, its RD cost (*Jmode*1) is compared with average RD cost of *mode*1 MBs (*Ave_Jmode*1) in the previous encoded frames. If Eq. 22 is satisfied, only *mode*0 and *mode*1 are candidate inter modes. Otherwise, we obtained the RD cost of *mode*2 (*Jmode*2) and *mode*3 (*Jmode*3) and use Eq. 23 to judge whether current inter modes are well enough. Since inter mode decision on modes below 8×8 is quite complicated, we only insert RD cost check after ME on big modes such as *mode*1 to *mode*3.

$$\begin{aligned} J(s,c,mode|QP,\lambda_{mode}) = {} & SSD(s,c,mode|QP) \\ & +\lambda_{mode} \times R(s,c,mode|QP) \end{aligned} \tag{21}$$

$$Jmode1 < Ave\_Jmode1 \tag{22}$$

$$\{Jmode2, Jmode3\} < \frac{Ave\_Jmode2 + Ave\_Jmode3}{2} \tag{23}$$

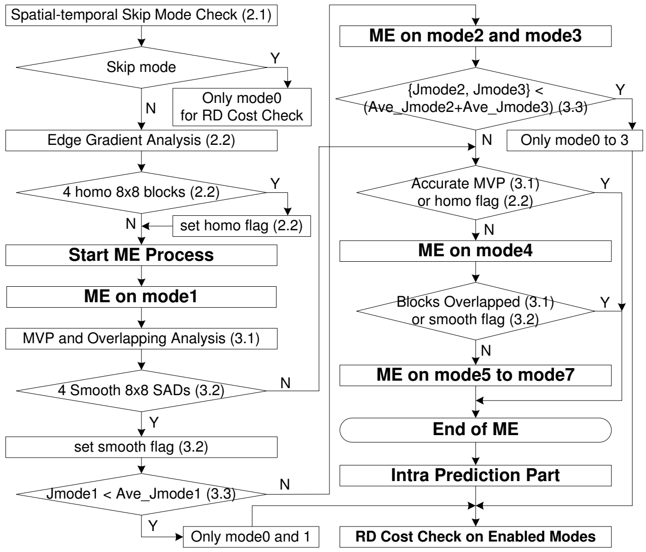


Figure 7: Overall Flow Chart of Proposed Algorithm

4. OVERALL ALGORITHM AND EXPERIMENTS

The overall flow chart of our algorithm is shown in Fig. 7. The parts with bold font are original JM mode decision flow and each of our schemes described in this paper is noted with its section number in brackets. It is shown that schemes in section 2.1 and 2.2 work before ME start and the rest schemes are involved with the ME process.

The proposed algorithm is implemented in JM 11.0 software. Several QCIF and CIF clips with different features are used for simulation. We encode 200 frames with RD optimization enabled. The QP value ranges from 28 to 40 with 4 intervals. The encoding structure is IPPP under baseline profile and 1 reference frame. The search range for QCIF and CIF are ±16 and ±24 respectively.

The experiments and comparisons are shown in Tab. 1 and Tab. 2. We use Eq. 24 to analyze the ratio of motion estimation time (*MET*) and bit increment. The Γ_{pro} is the element of proposed method (ours or others) and Γ_{jm} is the related one caused by original JM full search which loops all inter modes. The Γ can be *MET* or *Bits*. As for $\Delta PSNR$, it is calculated by subtraction of JM's PSNR and that of proposed one. The '+' in Tab. 2 represents PSNR gain and increment of bits. The meaning of '-' in Tab. 2 can be deduced by analogy. It is shown that our scheme is superior to [4] in terms of complexity reduction, especially clips with slow motion feature like container_qcif. In case of [6], it can achieve high complexity reduction for clips such as container_qcif and paris_cif. However, the situation of fast motion (football_qcif/cif), complex background (coastguard_cif) or camera's shaking (foreman_qcif/cif) will deteriorates the efficiency of this algorithm greatly. As for [8], the ΔMET is always large among different clips. However, the quality trade-off is also very serious. As shown in Tab. 2, besides our result, we also use bold font to mark the $\Delta PSNR$ which surpasses -0.1dB and $\Delta Bits$ which is larger than 1%. It is shown that most cases fall into [8] and the bits increment in fast motion clips (football_qcif/cif) is extremely large. In our scheme, the quality loss and bits gain are always trivial while we also achieve large reduction for clips with static feature

Table 1: Complexity Analysis based on $-\Delta MET$ (%)

clips	QP=28				QP=32				QP=36				QP=40			
	[6]	[8]	[4]	our	[6]	[8]	[4]	our	[6]	[8]	[4]	our	[6]	[8]	[4]	our
[1]	5.7	38.4	24.6	**24.9**	8.5	42.3	24.8	**23.0**	10.1	41.9	25.6	**21.6**	12.7	41.8	28.9	**23.9**
[2]	9.5	45.9	22.5	**30.6**	11.2	47.6	26.2	**27.2**	13.4	49.6	28.5	**34.2**	17.9	49.5	34.6	**35.0**
[3]	51.8	45.3	26.7	**51.7**	56.6	42.4	28.6	**52.3**	57.0	41.8	33.9	**50.2**	66.5	40.6	32.3	**53.4**
[4]	9.6	47.6	21.4	**35.2**	13.9	49.9	23.8	**33.7**	20.0	50.1	28.1	**33.0**	32.5	48.3	30.3	**30.3**
[5]	8.0	47.2	26.2	**49.0**	11.0	47.9	27.3	**46.5**	15.0	47.9	30.9	**45.6**	21.0	48.5	36.3	**46.1**
[6]	28.9	45.4	38.1	**40.2**	33.8	44.8	37.3	**39.1**	37.8	45.3	36.3	**40.3**	41.5	46.5	36.8	**44.7**
[7]	2.1	51.0	24.8	**39.8**	6.5	50.6	30.8	**38.1**	9.3	49.2	35.3	**40.5**	12.5	48.4	35.3	**46.5**
[8]	7.3	47.4	31.8	**37.3**	4.7	44.6	26.7	**29.9**	4.4	45.5	27.2	**31.8**	10.9	45.1	27.4	**32.8**

[1]: football_qcif, [2]: foreman_qcif, [3]: container_qcif, [4]: carphone_qcif,
[5]: foreman_cif, [6]: paris_cif, [7]: coastguard_cif, [8]: football_cif

Table 2: Quality Analysis based on $\Delta PSNR$ (dB) and $\Delta Bits$ (%)

clips	criterion	QP=28				QP=32				QP=36			
		[6]	[8]	[4]	our	[6]	[8]	[4]	our	[6]	[8]	[4]	our
[1]	$\Delta PSNR$	-0.01	**-0.24**	-0.04	**-0.06**	-0.02	**-0.30**	-0.03	**-0.07**	-0.05	**-0.34**	-0.03	**-0.09**
	$\Delta Bits$	+0.00	+**4.78**	+0.90	**-0.17**	-0.02	+**4.88**	+0.59	**-0.23**	-0.10	+**4.29**	+0.77	**-0.31**
[2]	$\Delta PSNR$	-0.06	**-0.18**	-0.07	**-0.09**	**-0.15**	**-0.21**	-0.05	**-0.09**	-0.09	**-0.17**	-0.05	**-0.00**
	$\Delta Bits$	-0.33	+**1.00**	+0.10	**-0.27**	-0.45	+0.41	+0.48	**-0.28**	-0.64	+**1.64**	+0.28	+**0.00**
[3]	$\Delta PSNR$	-0.09	-0.07	-0.00	**-0.05**	-0.04	-0.03	-0.02	**-0.03**	-0.06	-0.02	-0.01	**-0.03**
	$\Delta Bits$	+**1.72**	+**1.57**	+0.01	**-0.16**	-0.04	+**1.25**	+0.35	**-0.11**	+**2.27**	+**1.17**	+**1.01**	**-0.10**
[4]	$\Delta PSNR$	-0.02	**-0.17**	-0.01	**-0.06**	-0.03	**-0.15**	-0.03	**-0.07**	-0.01	-0.07	-0.02	+**0.01**
	$\Delta Bits$	+0.22	+0.32	+0.32	**-0.17**	+0.63	+**1.42**	+0.77	**-0.20**	+0.80	+**1.10**	+0.00	+**0.04**
[5]	$\Delta PSNR$	-0.02	**-0.24**	-0.02	**-0.01**	-0.05	**-0.25**	-0.04	**-0.02**	-0.05	**-0.25**	-0.00	**-0.01**
	$\Delta Bits$	+0.29	+**1.90**	+0.21	+**0.25**	+**1.10**	+**1.36**	+0.26	+**0.24**	+**1.18**	+**1.28**	+0.13	+**0.25**
[6]	$\Delta PSNR$	-0.06	-0.08	-0.04	**-0.07**	-0.08	**-0.12**	-0.04	**-0.08**	**-0.12**	**-0.11**	-0.06	**-0.04**
	$\Delta Bits$	+0.05	+**2.09**	+0.36	+**0.08**	+0.14	+**1.43**	+0.06	+**0.03**	+0.61	+**1.00**	+0.07	+**0.01**
[7]	$\Delta PSNR$	-0.00	**-0.20**	-0.03	**-0.04**	-0.01	**-0.22**	-0.04	**-0.04**	-0.00	**-0.17**	-0.03	**-0.05**
	$\Delta Bits$	+0.05	+**1.28**	+0.00	+**0.01**	+0.00	+**2.71**	+0.15	+**0.52**	+0.02	+**2.85**	+0.13	+**0.04**
[8]	$\Delta PSNR$	-0.01	**-0.34**	-0.01	**-0.02**	-0.01	**-0.41**	-0.03	**-0.05**	-0.02	**-0.45**	-0.01	**-0.03**
	$\Delta Bits$	+0.05	+**4.62**	+0.42	+**0.23**	+0.16	+**4.74**	+0.14	+**0.20**	+0.21	+**4.61**	+0.20	+**0.15**

and comparative big reduction for clips of different types. For the QCIF format, the bits increment is always negative with negligible PSNR loss. In all, about 21.6% to 53.4% complexity is reduced for the inter mode decision process.

$$\Delta\Gamma = \frac{\Gamma_{pro} - \Gamma_{jm}}{\Gamma_{jm}} \times 100\%,\ \Gamma \in \{MET, Bits\} \tag{24}$$

5. CONCLUSION REMARKS

One fast inter mode decision algorithm is contributed in this paper. In the pre-stage, the spatial-temporal information is used to detect skip mode in an early stage. The current MB's homogeneity is also abstracted to filter out unpromising small modes. In the motion stage, the MVP's accuracy, the block overlapping and SAD distribution are analyzed to bypass unnecessary inter modes. Furthermore, the RD costs of big modes are obtained in an early stage and compared with historical ones to fasten mode decision procedure. Experiments show that our algorithm can achieve up to 53.4% speed-up ratio with trivial quality loss and bit increment.

6. ACKNOWLEDGMENTS

This research was supported by "Ambient SoC Global COE Program of Waseda University" of the Ministry of Education, Culture, Sports, Science and Technology, Japan. The research was also supported by Core Research for Evolutional Science and Technology (CREST), JST.

REFERENCES

[1] T. Wiegand and etc, "Overview of the H.264/AVC video coding standard," *IEEE Trans. on CSVT*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[2] F. Pan and etc, "Fast mode decision algorithm for intraprediction in h.264/avc video coding," *IEEE Trans. on CSVT*, vol. 15, no. 7, pp. 813–822, Jul. 2005.

[3] D. Zhu, Q. Dai, and R. Ding, "Fast inter prediction mode decision for h.264," in *Proc. ICME 2004*, Taipei, Taiwan, Jun. 27-30. 2004, pp. 1123–1126.

[4] X. Jing and L. Chau, "Fast approach for h.264 inter-mode decision," *Electron Letter*, vol. 40, no. 17, pp. 1050–1052, Sept. 2004.

[5] Y. Kim and etc, "Adaptive mode decision for h.264 encoder," *Electron Letter*, vol. 40, no. 19, pp. 1172–1173, Sept. 2004.

[6] C. Grecos and M. Yang, "Fast inter mode prediction for p slices in the h264 video coding standard," *IEEE Trans. on Broadcasting*, vol. 51, no. 2, pp. 256–263, Jun. 2005.

[7] I. Choi, J. Lee, and B. Jeon, "Fast coding mode selection with ratedistortion optimizaion for mpeg-4 part-10 avc/h.264," *IEEE Trans. on CSVT*, vol. 16, no. 12, pp. 1557–1558, Dec. 2006.

[8] L. Salgado and M. Nieto, "Sequence independent very fast mode decision algorithm on h.264/avc baseline profile," in *Proc. ICIP 2006*, Atlanta, USA, Oct. 27-30. 2006, pp. 41–44.

[9] H. Gish and J.N. Pierce, "Asymptotically efficient quantizing," *IEEE Trans. on Information Theory*, vol. 14, no. 5, pp. 676–683, Sept. 1968.

[10] T. Berger, *Rate Distortion Theory*. Prentice Hall, 1971.