MACROBLOCK FEATURE AND MOTION INVOLVED MULTI-STAGE FAST INTER MODE DECISION ALGORITHM IN H.264/AVC VIDEO CODING

Yiqing HUANG   *Qin LIU*   *Takeshi IKENAGA*

Waseda University, 2-7, Hibikino, Wakamatsu, Kitakyushu, Fukuoka, Japan

ABSTRACT

One fast inter mode decision algorithm is proposed in this paper. The whole algorithm is convoluted with block matching process. Firstly, before ME process, by exploiting spatial and temporal information, a skip mode early detection algorithm is proposed. Also, in this stage, edge gradient is used to filter out unpromising modes. Secondly, during the ME stage, the original search window is separated into several layers and our fast decision scheme works with motion information of each layer. Moreover, before stepping into small modes, the distribution of SAD (sum-of-absolute-difference) and RD (rate distortion) costs of big modes are analyzed in an early stage to accelerate the inter mode decision process. Experiments show that our algorithm can achieve a speed-up factor of up to 66.0% with trivial bit increment and quality degradation.

Index Terms— mode decision, H.264/AVC, feature analysis

1. INTRODUCTION

The H.264/AVC standard provides superior coding performance by using many new techniques. However, the negative side is that the computation complexity is also increased dramatically [1], especially the inter prediction part. As shown in Fig. 1, besides skip mode, there are 7 block modes for inter prediction. The encoding process will loop all these modes and select one with the minimum compression cost on the reference plane. 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.

Many works have been done to solve the problem. In [2], a pre-encoding scheme is proposed, which abstracts a down sampled small image and restrict the inter block modes within a small subset. [3] uses mean of absolute frame difference (MAFD) to filter out unpromising inter modes. In [4], an adaptive mode decision process based on all-zero coefficients block is proposed. Literature [5][6] focus on the optimization of early skip mode decision to release the complexity. However, the idea of introducing pre-processing in [2] and [3] 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 schemes [5][6], 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 blocks and skip modes will decrease greatly, which deteriorates the efficiency of inter mode filtering. In [7], a very fast decision algorithm is proposed, which dramatically reduces complexity for both low-motion and high-motion sequences. However, 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, the gradients of current macroblock (MB), spatial and temporal features of encoded MBs are inspected to filter out unpromising modes and detect skip mode. Secondly, during ME process, motion information of different layers on search window plane is collected for fast mode decision. We focus on the information of motion vector predictor's accuracy, the block matching on centering region, SAD distribution and RD cost.

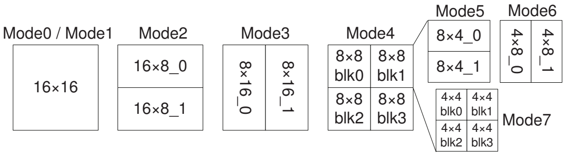


Fig. 1. Inter Prediction Modes in H.264/AVC

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 involved 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. EARLY STAGE INTER MODE DECISION

2.1. Gradient Analysis to Inter Mode Filtering

Edge detection is one useful technique in both image processing and pattern recognition fields. In [8] and [9], Sobel operator is used for fast mode decision on intra and inter prediction. In our paper, we also use edge detection to remove unpromising inter modes for homogeneous (homo) MB. Similar as [9], we apply edge detection on the MB level. As shown in Fig. 2(a), the edge detection is executed on four 8×8 sub-blocks within current MB. When gradients of each pixel within one specific 8×8 sub-block (blk8x8_i, i$\in\{0,1,2,3\}$) is within a predefined threshold, such 8×8 block is regarded as homo one; otherwise, it is an edge 8×8 block. For homo blk8x8, small inter modes (8×4 to 4×4) are discarded in ME process. When all four 8×8 blocks are homo ones, the entire 8×8 inter mode is bypassed. With increase of quantization parameter (QP), the smoothness of reconstructed frames is also increased, which results in decline of image's details. So, threshold for edge detection can be simply set linear with QP value ($4\times$QP based on our exhaustive experiments).

2.2. Early Decision based on Spatial and Temporal Analysis

For conventional video clips, redundancy always exists in form of spatial and temporal domains. In this paper, one spatial-temporal

---

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, JST.

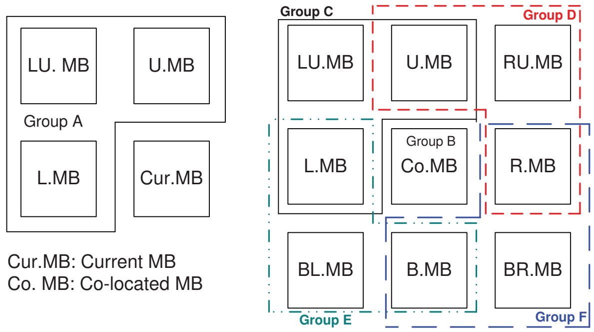
```
Loop blk8x8_i of current MB, i = [0, 3]
    Edge gradient analysis of blk8x8_i
    If homogenous blk8x8_i
        Discard Mode5 to Mode7
End Loop
If all blk8x8 are homogenous
    Disable Mode4 for Cur.MB

※ Use Sobel operator for edge analysis
※ Mode0 is always included
```

(a) Edge Oriented Mode Filtering

```
Spatial order check in GA
Check MB in GB
If MBs in {GA, GB} != Mode0
    No early decision
Else
    Loop GC to GF
        Check MBs' modes
        Gradient analysis
    End Loop
If Criterion Satisfied
    Only mode0 for Cur.MB
Else
    No early decision
```

(b) Early Skip Mode Decision

Fig. 2. Pseudo Coding of Pre-stage Decision



(a) Spatial Order on Cur_Frm (b) Temporal Order on Prev_Frm

Fig. 3. Spatial and Temporal Order for Early Decision

skip mode early decision scheme is applied before encoding process.

For raster scan based processing order, the only spatial information available is from the MB already encoded. As shown in Fig. 3(a), the left-up MB (LU.MB), left MB (L.MB) and upper MB (U.MB) around current MB (Cur.MB) are useful candidates on current frame (Cur_Frm). In our algorithm, these MBs are included in group A (GA), as shown in Eq. 1.

$$\{LU.MB, L.MB, U.MB\} \text{ on } Cur\_Frm \in GA \quad (1)$$

To improve efficiency and correctness, temporal information is also needed. As shown in Fig. 3(b), on previous frame (Prev_Frm), besides co-located MB (Co.MB), eight MBs namely LU.MB, RU.MB, BL.MB, BR.MB, U.MB, B.MB, L.MB and R.MB of different locations around Co.MB are used. We assign these nine MBs into five groups (Group B to Group F), as shown from Eq. 2 to Eq. 6. The detail decision flow is shown in Fig.2(b). The whole process consists of two stages. Firstly, the modes of MBs in GA and GB are checked. If modes of GA or GB are not all skip modes, no early decision will be made and the ME starts with remaining modes in section 2.1. Secondly, if "else" branch is activated in first stage, the second stage will be executed by looping MBs from GC to GF. During this stage, besides mode information, gradients of MBs are used for similarity check between Cur.MB and MBs in four groups. When criterion in Eq. 7 is satisfied, Cur.MB will be decided as skip MB. In detail, all the modes in current group (Cur.GP) must be skip modes. Here, Cur.GP is the group under analysis, which can be GB, GC, GD or GE. Moreover, the gradient of Cur.MB must between $G_{Min}(Cur.GP)$ and $G_{Max}(Cur.GP)$ which are minimum and maximum gradients of MB in Cur.GP, respectively. Both mode and gradient criterions in Eq. 7 must be satisfied for one single group.

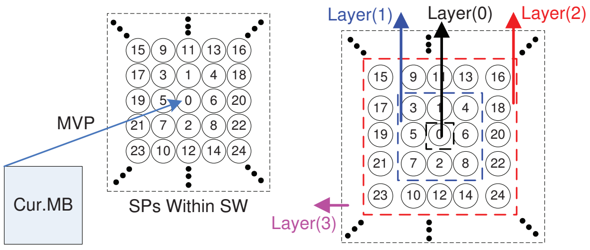


Fig. 4. Matching Process and Layer Division

If none of the four groups reach constraints in Eq. 7, no early decision will be made at this stage.

$$\{Co.MB\} \text{ on } Prev\_Frm \in GB \quad (2)$$

$$\{LU.MB, L.MB, U.MB\} \text{ on } Prev\_Frm \in GC \quad (3)$$

$$\{RU.MB, R.MB, U.MB\} \text{ on } Prev\_Frm \in GD \quad (4)$$

$$\{BL.MB, L.MB, B.MB\} \text{ on } Prev\_Frm \in GE \quad (5)$$

$$\{BR.MB, R.MB, B.MB\} \text{ on } Prev\_Frm \in GF \quad (6)$$

$$\begin{cases} \{modes \in Cur.GP\} = 0 \\ G_{Min}(Cur.GP) < G(Cur.MB) < G_{Max}(Cur.GP) \\ GP = \{GC, GD, GE, GF\} \end{cases} \quad (7)$$

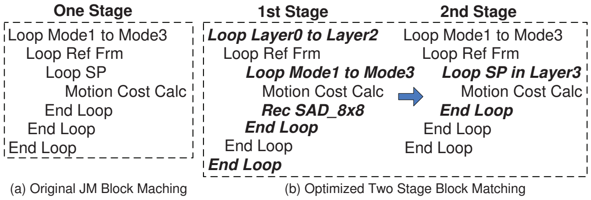
3. MOTION CONVOLUTED INTER MODE DECISION

In pre-stage decision, unpromising inter modes are filtered based on features of current MB, spatial and temporal previous encoded MBs. However, the complexity reduction is quite limited due to the abundant motion MBs. In fact, motion information is another important feature for inter mode decision. With better understanding of motion feature, the candidate modes can be further narrowed down.

3.1. Layer Based Motion Region Division

In JM software, the block matching process of H.264/AVC standard is implemented by 2 steps, as shown in left part of Fig. 4. The motion vector predictor (MVP) is firstly decided based on neighboring MBs or sub-blocks. Then, the matching process is executed based on certain criterion and searching order on the search window plane. The criterions can be SAD, SATD or SSD while searching orders such as full search and UMHexagon search are widely adopted. The search window in our algorithm is divided into different layers, as shown in right part of Fig. 4. Since most motion MBs are center biased, the centering 25 points are split into three layers (Layer(0) to Layer(2)). The Layer(3) consists of points outside of layer(2). During the ME process, motion feature of all inter modes is analyzed layer by layer.

Also, the JM algorithm has to be modified into layer based one. Fig. 5(a) is the original matching flow which consists of three big loops. The block matching process on all 7 inter modes is divided into two parts. Big modes such as Mode1 to Mode3 are assigned separately from small modes (Mode4 to Mode7). In our paper, we only focus on big modes and leave processing of small modes unchanged. It is obvious that original algorithm is a mode based one-stage scheme. Once the specific mode is decision, the matching process is executed on the whole search window directly. In our algorithm, the whole matching process is divide into two stage. In the first stage, matching process is executed on Layer(0) to Layer(2). For search point (SP) on Layer(3), scheme in JM algorithm which



```
Loop Mode1 to Mode3
    Loop Ref Frm
        Loop SP
            Motion Cost Calc
        End Loop
    End Loop
End Loop
```

(a) Original JM Block Maching

```
Loop Layer0 to Layer2
    Loop Ref Frm
        Loop Mode1 to Mode3
            Motion Cost Calc
            Rec SAD_8x8
        End Loop
    End Loop
End Loop
```

```
Loop Mode1 to Mode3
    Loop Ref Frm
        Loop SP in Layer3
            Motion Cost Calc
        End Loop
    End Loop
End Loop
```

(b) Optimized Two Stage Block Matching

Fig. 5. Original Algorithm and Our Modification

loops candidate mode is used. Fig. 5(b) shows our modification to JM scheme (marked with bold font) and we record SAD value in 8×8 size (Rec SAD_8x8) during 1st stage. The purpose of modification is to facilitate our proposed motion convoluted mode decision schemes. The details will be given in following parts.

3.2. MVP and Motion Analysis on Centering Region

For sequence with smooth and regular motion, the prediction of MVP is very accurate. From our experiments, large proportion of best integer position (BIP) is located in MVP even for case of foreman_qcif and carphone_qcif. The high accuracy in MVP also indicates that the current MB is seldom split into small blocks. To avoid being trapped into local minimum, we only focus on information of 16×16 mode, as shown in Eq. 8. It means that after matching of mode1 to mode3 on Layer(0), the best mode (best_md) is Model1 and the related motion cost ($mcost$) is within Thr_MVP(20×QP), then it is regarded as big mode MB and Mode4 to Mode7 are discarded.

$$\begin{cases} best\_md = Mode1 \\ mcost\_(Mode1) \leq Thr\_MVP \end{cases} \tag{8}$$

Besides MVP, we also apply early mode decision on Layer(1) and Layer(2). Criterions, as shown in Eq. 9, are given out for matching analysis of big modes on each layer. Firstly, Eq. 9 requires that best modes on neighboring two layers ($Layer(k)(best\_md)$ and $Layer(k-1)(best\_md)$) are the same. Secondly, the difference of motion cost ($Layer(k)\_mcost$) on current layer must be less than $\alpha(k)$ of its related one on previous layer ($Layer(k-1)\_mcost$). Here, $\alpha(1)$ and $\alpha(2)$ are two empirical factors for Layer(1) and Layer(2). When Eq. 9 is satisfied in both Layer(1) and Layer(2), current MB will also be treated as big mode MB.

$$\begin{cases} Layer(k)(best\_md) = Layer(k-1)(best\_md) \\ Layer(k)\_mcost \leq \alpha(k) \times Layer(k-1)\_mcost \\ \alpha(1) = 0.6,\ \ \alpha(2) = 0.8,\ \ k \in \{1, 2\} \end{cases} \tag{9}$$

3.3. SAD Distribution Analysis

After ME of big modes on centering region, 2nd stage in Fig. 5(b) will be executed, which involves matching on Layer(3). Before 2nd stage starts, the SAD distribution is further checked. With decrease of SAD value, the smoothness degree of current MB is increased, which indicates that further matching on small modes is not necessary. Since big sub-block is more representative for motion feature, we use 8×8 sized sub-block for SAD distribution analysis, as shown in Eq. 10 to Eq. 13. Both horizontal and vertical differences are checked on best mode. The current MB is regarded as big mode MB if Eq. 10 to Eq. 13 are all satisfied.

$$|SAD8 \times 8\_LU - SAD8 \times 8\_RU| < Thr\_SAD \tag{10}$$

$$|SAD8 \times 8\_BL - SAD8 \times 8\_BR| < Thr\_SAD \tag{11}$$

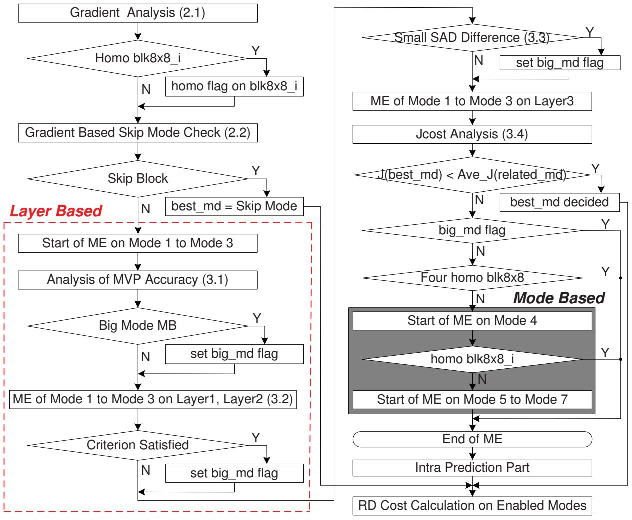


Fig. 6. Flow Chart of Proposed Algorithm

$$|SAD8 \times 8\_LU - SAD8 \times 8\_BL| < Thr\_SAD \tag{12}$$

$$|SAD8 \times 8\_RU - SAD8 \times 8\_BR| < Thr\_SAD \tag{13}$$

3.4. Rate Distortion Analysis of Big Inter Modes

When matching of big modes on Layer(3) is finished, we apply rate distortion check before matching process on small modes, as shown in Eq. 15. In JM algorithm, mode decision is executed after block matching of all 7 inter modes on reference plane. Eq. 14 is the criterion in JM, where 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. In our scheme, RD cost analysis is pre-checked after matching of big modes on Layer(3). In detail, if the cost of best mode ($J(best\_md)$) is smaller than average cost of related mode on previous frame ($Ave\_J(related\_md)$), the final best mode is set as current $best\_md$ and the rest process is skipped.

$$\begin{aligned} J(s, c, mode|QP, \lambda_{mode}) &= SSD(s, c, mode|QP) \\ &\quad + \lambda_{mode} \times R(s, c, mode|QP) \end{aligned} \tag{14}$$

$$J(best\_md) < Ave\_J(related\_md) \tag{15}$$

4. OVERALL ALGORITHM AND EXPERIMENTS

The overall flow chart is shown in Fig. 6. Each proposed scheme is marked with its section number in parentheses. It is shown that ME of big modes is layer based searching while matching of small modes is mode based way. Also, schemes in section 2.1 and 2.2 serves as a pre-process and the rest parts are involved with motion process.

The overall algorithm is implemented in JM 11.0 software. QCIF and CIF clips with different features are used for simulation. We encode 200 frames with RD optimization enabled. The QP ranges from 28 to 40 with interval of 4, IPPP structure and 1 reference frame. The search ranges for QCIF and CIF are ±16 and ±24, respectively.

Table. 1 and Table. 2 are experiments and comparisons with others. ΔMET, $\Delta PSNR$ and $\Delta BitR$ are used for analysis of ME

Table 1. Quality Analysis based on $\Delta PSNR$ (dB) and $\Delta Bits$ (%)

clips	criterion	QP=28			QP=32			QP=36		
		[7]	[3]	our	[7]	[3]	our	[7]	[3]	our
c1	$\Delta PSNR$	**-0.24**	-0.04	**-0.06**	**-0.30**	-0.03	**-0.06**	**-0.34**	-0.03	**-0.05**
	$\Delta Bits$	+**4.78**	+0.90	+**0.32**	+**4.88**	+0.59	+**0.42**	+**4.29**	+0.77	+**0.63**
c2	$\Delta PSNR$	**-0.18**	-0.07	**-0.09**	**-0.21**	-0.05	**-0.10**	**-0.17**	-0.05	**-0.19**
	$\Delta Bits$	+**1.00**	+0.10	+**0.30**	+0.41	+0.48	+**0.20**	+**1.64**	+0.28	+**0.20**
c3	$\Delta PSNR$	-0.07	-0.00	**-0.08**	-0.03	-0.02	**-0.07**	-0.02	-0.01	**-0.06**
	$\Delta Bits$	+**1.57**	+0.01	**-0.59**	+**1.25**	+0.35	**-0.10**	+**1.17**	+**1.01**	**-0.11**
c4	$\Delta PSNR$	**-0.17**	-0.01	**-0.03**	**-0.15**	-0.03	**-0.05**	-0.07	-0.02	**-0.04**
	$\Delta Bits$	+0.32	+0.32	+**0.28**	+**1.42**	+0.77	+**0.53**	+**1.10**	+0.00	+**0.65**
c5	$\Delta PSNR$	**-0.24**	-0.02	**-0.11**	**-0.25**	-0.04	**-0.12**	**-0.25**	-0.00	**-0.21**
	$\Delta Bits$	+**1.90**	+0.21	+**0.22**	+**1.36**	+0.26	+**0.32**	+**1.28**	+0.13	+**0.35**
c6	$\Delta PSNR$	-0.08	-0.04	**-0.17**	**-0.12**	-0.04	**-0.18**	**-0.11**	-0.06	**-0.12**
	$\Delta Bits$	+**2.09**	+0.36	+**0.15**	+**1.43**	+0.06	+**0.13**	+**1.00**	+0.07	+**0.11**
c7	$\Delta PSNR$	**-0.20**	-0.03	**-0.03**	**-0.22**	-0.04	**-0.03**	**-0.17**	-0.03	**-0.09**
	$\Delta Bits$	+**1.28**	+0.00	+**0.60**	+**2.71**	+0.15	+**0.62**	+**2.85**	+0.13	+**0.46**
c8	$\Delta PSNR$	**-0.34**	-0.01	**-0.03**	**-0.41**	-0.03	**-0.06**	**-0.45**	-0.01	**-0.18**
	$\Delta Bits$	+**4.62**	+0.42	+**0.77**	+**4.74**	+0.14	+**0.75**	+**4.61**	+0.20	+**0.28**

c1: football_qcif, c2: foreman_qcif, c3: container_qcif, c4: carphone_qcif, c5: foreman_cif, c6: paris_cif, c7: coastguard_cif, c8: football_cif

Table 2. Complexity Analysis based on ΔMET (%)

clips	QP=28			QP=32			QP=36			QP=40		
	[7]	[3]	our	[7]	[3]	our	[7]	[3]	our	[7]	[3]	our
c1	38.4	24.6	33.9	42.3	24.8	35.0	41.9	25.6	36.8	41.8	28.9	39.0
c2	45.9	22.5	34.7	47.6	26.2	38.1	49.6	28.5	50.5	49.5	34.6	61.5
c3	45.3	26.7	45.2	42.4	28.6	49.1	41.8	33.9	59.7	40.6	32.3	66.0
c4	47.6	21.4	37.3	49.9	23.8	38.8	50.1	28.1	48.2	48.3	30.3	48.5
c5	47.2	26.2	50.0	47.9	27.3	49.5	47.9	30.9	51.7	48.5	36.3	53.0
c6	45.4	38.1	43.5	44.8	37.3	42.3	45.3	36.3	42.8	46.5	36.8	46.0
c7	51.0	24.8	45.1	50.6	30.8	44.3	49.2	35.3	47.4	48.4	35.3	50.5
c8	47.4	31.8	39.6	44.6	26.7	41.0	45.5	27.2	48.8	45.1	27.4	50.1

time saving and video quality between original JM and corresponding proposals (ours or others). The '+' in Tab. 1 represents PSNR gain and increment of bits. The meaning of '-' in Tab. 1 can be deduced by analogy. It is shown that our scheme is superior to [3] in terms of complexity reduction, especially clips with slow motion feature like container_qcif. Compared with [7], our algorithm can not always achieve large complexity reduction for all QPs, especially in several sequences such as foreman_qcif and carphone_qcif when QP is 28 or 32. However, the problem of [7] is that, its quality tradeoff is always very serious, especially for motion abundant sequences. As shown in Tab. 1, 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 [7] and the bits increments in fast motion clips (football_qcif/cif) are extremely large, which means that the decision process in this work may not so accurate compared with ours. In our scheme, the quality loss and bits gain are always trivial while we also achieve large reduction for clips with static feature and comparative big reduction for clips of various types under different QPs. In all, about 33.9% to 66.0% complexity is reduced for inter mode decision process.

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 detected skip mode in an early stage. The homogeneity of current MB is also abstracted to filter out unpromising modes. In the motion stage, the MVP's accuracy and block matching result on centering region are extracted as motion information for fast decision process. Moreover, we apply early RD costs check for big inter modes and use SAD distribution as a criterion for skipping unpromising inter modes. Experiments show that our algorithm can achieve up to 66.0% ME time reduction with trivial quality loss and bits increment.

6. REFERENCES

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

[2] D. Zhu, Q. Dai, and R. Ding, "Fast inter prediction mode decision for h.264," in *ICME '04*, Jun 2004, pp. 1123–1126.

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

[4] Y. Kim, J. Yoo, S. Lee, J. Shin, J. Park, and H. Jung, "Adaptive mode decision for h.264 encoder," *Electron Letter*, vol. 40, no. 19, pp. 1172–1173, Sept 2004.

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

[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] L. Salgado and M. Nieto, "Sequence independent very fast mode decision algorithm on h.264/avc baseline profile," in *ICIP '06*, Oct 2006, pp. 41–44.

[8] F. Pan, X. Lin, S. Rahardja, K. Lim, Z. Li, D. Wu, and S. Wu, "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.

[9] D. Wu, S. Wu, K. Lim, F. Pan, Z. Li, and X. Lin, "Block inter mode decision for fast encoding of h.264," in *ICASSP '04*, May 2004, pp. 181–184.