Fast Inter Mode Decision Based on Hierarchical Homogeneous Detection and Cost Analysis for H.264/AVC Coders

Hung-Ming Wang, Ji-Kun Lin, and Jar-Ferr Yang

Department of Electrical Engineering
National Cheng Kung University
1 University Road, Tainan, Taiwan 701, R.O.C.

Abstract—**In H.264 advanced video coding (AVC), variable block size motion estimation plays an important role in compression of interframes. In this paper, we propose a fast inter prediction algorithm based on hierarchical homogeneous detection and cost analysis to select the best mode effectively. For each macroblock, we first detect that whether the macroblock is spatial homogeneous or not. For the non spatial homogeneous macroblock, we then perform the 16x16 motion estimation and examine if the 16x16 block is temporal homogeneous or not. Once the homogeneous macroblock is detected in the above process, the best mode will be chosen as 16x16 mode. For the non-homogeneous macroblock, we then execute 8x8 motion estimation and analyze the cost of 8x8 mode and 16x16 mode for deciding the best inter mode should be 16x16 mode or any other mode. The process for searching the best 8x8 block subtype is similar to the process for macroblocks. Finally, the best inter mode is decided by selecting the inter mode with least cost from the candidate modes. Experimental results show that our proposed algorithm can save about 32~54% computation time without introducing any noticeable performance degradation.**

I.   INTRODUCTION

In the digital epoch, people demand more and more high-quality multimedia services in their digital lives. Nevertheless, with quality increased, we need higher bandwidth for trade-off. However, in most portable devices like cellular phone and PDA, the bandwidth is limited. Therefore, to attain the high quality video with less bandwidth, there were many compression standards, such as MPEG-1, MPEG-2, MPEG-4, H.261, and H.263, being developed for effective video transmission and storage. To further increase the compression efficiency, the Joint Video Team (JVT), which is organized by ISO MPEG and ITU-T VCEG, has finalize a new video coding standard H.264/AVC in May 2003 [1]. Experimental studies have shown that H.264/AVC can achieve about 50% bit rate saving with the similar perceptual video quality compare to other existing standards.

H.264/AVC utilizes many advance features such as 4x4 integer transform, spatial intra-prediction in I-frame coding, 1/4 pixel motion compensation, multiple reference frames and multiple block size for P-frame coding, context-based adaptive variable length coding (CAVLC) and context-based adaptive binary arithmetic coding (CABAC) for entropy coding, etc [2]. These features help H.264/AVC to achieve significantly high compression efficiency. However, these features also increase the complexity in H.264/AVC. Take variable block size motion estimation for example, there are totally seven block sizes, 16x16, 16x8, 8x16, 8x8, 8x4, 4x8, and 4x4. A macroblock can be encoded using one 16x16 block, two 16x8 blocks, two 8x16 blocks or four 8x8 blocks as shown in Fig. 1(a). Furthermore, one 8x8 block can be divided into two 8x4 blocks, two 4x8 blocks, or four 4x4 blocks as shown in Fig. 1(b). To encode a macroblock, we have to test all the inter modes to find the best inter mode with the least RD cost, which is very time-consuming. To reduce the complexity of inter mode decision, there are many algorithms being developed to accelerate the encoding time with tolerant performance degradation [3], [4].

In this paper, we will develop a fast inter mode decision algorithm to choose the best block type for each macroblock. The paper is organized as follows. First, the inter mode decision in H.264/AVC will be briefly introduced in Section II. In Section III, we will develop a fast algorithm based on hierarchical homogeneous detection and cost analysis to select the best inter mode effectively. Experimental results are shown in Section IV. Finally, we conclude the paper in Section V.

II. INTER MODE DECISION IN H.264/AVC

As we mentioned above, there are totally seven different block sizes can be used in interframe motion estimation. To select the best inter mode, the reference software JM10.1 provided two approaches: (1) RD optimization (RDO) and (2) non-RD optimization. The former uses the true distortion and true bits of each inter mode to calculate the RD cost while the later just adopts the sum of absolute Hadamard-transformed differences (SATD) and estimated motion vector cost of each inter mode to calculate the RD cost. It is obvious that we can achieve the

best coding efficiency by using the true RDO method. However, it is too time-consuming to implement in real-time applications. Hence, we will focus on non-RDO method, where motion estimation is performed for all possible block sizes to choose the best inter mode with the least RD cost. The procedure contains two main steps: (1) motion vector search of each inter mode, and (2) partition mode search. For each inter mode, motion estimation is first performed for each block size (16x16, 16x8, 8x16, 8x8, 8x4, 4x8) to find the best motion vector. The best motion vector (v_x, v_y) for a block size $N_1 \text{x} N_2$ is decided by minimizing a cost function:

$$J_{N_1 \text{x} N_2}(v_x, v_y) = SA(T)D_{N_1 \text{x} N_2}(v_x, v_y) + \lambda(Q_p)R(MVD_{N_1 \text{x} N_2}(v_x, v_y)), \quad (1)$$

where the *SAD* is used for integer pixel motion estimation and the *SATD* is used for sub-pixel motion estimation, are defined as follows:

$$SAD_{N_1 \text{x} N_2}(v_x, v_y) = \sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\left|\hat{f}(i+v_x, j+v_y) - f(i,j)\right|, \quad (2)$$

$$SATD_{N_1 \text{x} N_2}(v_x, v_y) = \sum_{i=1}^{N_1}\sum_{j=1}^{N_2}|c(i,j)|. \quad (3)$$

In (2), $\hat{f}(i+v_x, j+v_y)$ and $f(i,j)$ denote the $(i,j)^{\text{th}}$ element of the reference block $\hat{F}$ and the current block F, respectively. In (3), $c(i,j)$ is the $(i,j)^{\text{th}}$ element of C, which is the Hadamard transform of difference of $\hat{F}$ and F depicted as

$$C = \frac{T_H(\hat{F} - F)T_H^T}{2}, \quad (4)$$

with

$$T_H = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix}. \quad (5)$$

In (1), $\lambda(Q_p)$ is an approximated exponential function of the quantization parameter Q_p and $R(MVD)$ represents the rate cost contributed by motion vector difference (*MVD*), is expressed as:

$$MVD = \left|v_x - \bar{v}_x\right| + \left|v_y - \bar{v}_y\right|, \quad (6)$$

where $(\bar{v}_x, \bar{v}_y)$ is the predicted motion vector defined in the standard. After motion vector search, we can get the motion vector for each sub-block and its corresponding cost J_{sb}. The next step is to find the best partition mode with least cost. The cost is defined as follows:

$$J_{mode} = \sum_{sb \in macroblock} J_{sb}, \quad (7)$$

where J_{mode} denotes the cost of the partition mode and J_{sb} is the cost of the sub-block within the macroblock.

From the above, we can see that the computation complexity for deciding the best inter mode is very high since the motion estimation is performed for all possible block sizes. Therefore, how to accelerate the coding process and maintain the coding performance simultaneously is a very important subject in H.264/AVC interframe coding. In the next section, we will develop a fast algorithm for inter mode decision.

III. HIERARCHICAL HOMOGENEOUS MODE DETECTION AND COST ANALYSIS

A. Homogeneous Detection

In the original inter mode decision algorithm, we have to select the best inter mode from all possible inter modes. To avoid the exhaustive search, Wu et al. proposed a fast algorithm to early terminate the inter mode search process by using Sobel operator and a recommended threshold to detect homogeneous blocks [3]. The main concept of the algorithm is that homogeneous blocks in the picture would have similar motion and are very seldom split into smaller blocks. Although this concept is indeed very useful for accelerating the inter mode decision process, it is hard to deal with some sequences (ex. Mobile) which contain complex background because of fewer homogenous blocks. The detection of spatial homogeneous blocks always fairs to correctly detect the complex-but-still background as larger blocks. To handle this problem, we suggest that we should first utilize temporal homogeneous detection to further improve the coding process. For example, if we find that the current macroblock is not spatial homogeneous, we will then perform 16x16 motion estimation to get the residue block and decide if the macroblock is temporal homogeneous or not. If the block is temporal homogeneous, the inter mode will also be decided as 16x16 mode. In this paper, an NxN block is said to be temporal homogeneous if its residue block satisfies the following condition:

$$\sum_{i=1}^{N/2}\sum_{j=1}^{N/2}\left|r_k(i,j) - M_k\right| < Th, \quad k = 0,1,2,3. \quad (8)$$

In (8), $r_k(i,j)$ and M_k represent the $(i,j)^{\text{th}}$ element and the mean individually of the kth sub-block R_k of the residue block, which is shown in Fig. 2. *Th* is an under-determined threshold. Here, we use (8) for temporal homogeneous detection since it needs less computation than Sobel operator but can achieve the similar performance. To determine the threshold, *Th* defined in (8), we can use the similar concept like [5], which develops an algorithm for the early detection of all-zero blocks. By using its developed threshold and make the condition more strictly for avoiding performance degradation, we set the threshold is Q_{step} for a 4x4 block where Q_{step} is the quantization step in H.264/AVC. Hence, the formula (8) can be written as:

$$\sum_{i=1}^{N/2}\sum_{j=1}^{N/2}\left|r_k(i,j) - M_k\right| < \frac{N^2}{64}Q_{step}, \quad k = 0,1,2,3 \quad (9)$$

B. Cost Analysis

If the current macroblock is neither spatial homogeneous nor temporal homogeneous, we will then conduct 8×8 motion estimation to obtain the cost of 8×8 mode. If we can determine the best inter mode is 16×16 mode or not in advance by analyzing the costs of 16×16

mode and 8×8 mode, we can save more computation. To achieve this, we first assume that the cost of 16x16 mode is less than the cost of 8x8 mode, that is:

$$J_{16x16\,mode} < J_{8x8\,mode} \tag{10}$$

$$\Rightarrow SATD_{16x16mode} + MVcost_{16x16mode} < SATD_{8x8mode} + MVcost_{8x8mode}, \tag{11}$$

where $SATD_{mode}$ and $MVcost_{mode}$ denote the total SATD and motion vector costs of the corresponding inter mode, respectively. In the above condition, because 16×16 mode is better than 8x8 mode, it is reasonable to assume that the best mode should be selected from 16×16 mode, 16×8 mode, or 8×16 mode. However, if the following condition satisfies:

$$SATD_{16x16mode} - SATD_{8x8mode} < MVcost_{16x8mode} - MVcost_{16x16mode}. \tag{12}$$

We don't need to perform the 16×8 motion estimation to get the cost of mode 16×8. Since

$$SATD_{16x8mode} > SATD_{8x8mode} \tag{13}$$

$$\Rightarrow SATD_{16x16mode} - SATD_{16x8mode} < SATD_{16x16mode} - SATD_{8x8mode} \tag{14}$$

$$\Rightarrow SATD_{16x16mode} - SATD_{16x8mode} < MVcost_{16x8mode} - MVcost_{16x16mode} \tag{15}$$

$$\Rightarrow SATD_{16x16mode} + MVcost_{16x16mode} < SATD_{16x8mode} + MVcost_{16x8mode} \tag{16}$$

The inequality (16) means that the cost of 16×16 mode is less than that of 16×8 mode. For 8×16 mode, the condition is similar to 16×8 mode which just replaces $MVcost_{16x8\,mode}$ by $MVcost_{8x16mode}$ in (12). Thus, the remaining problem is how to estimate $MVcost_{16x8mode}$ or $MVcost_{8x16\,mode}$ from $MVcost_{16x16\,mode}$ and $MVcost_{8x8\,mode}$. It is obviously that $MVcost_{16x8mode}$ or $MVcost_{8x16mode}$ is between $MVcost_{16x16\,mode}$ and $MVcost_{8x8\,mode}$, and we can approximate them to the average of $MVcost_{16x16\,mode}$ and $MVcost_{8x8\,mode}$. Therefore, the formula (12) can be rewritten as:

$$SATD_{16x16mode} - SATD_{8x8mode} < \frac{MVcost_{8x8mode} - MVcost_{16x16mode}}{2} \tag{17}$$

By using (17), we can choose 16x16 mode as the best inter mode without performing 8x16 or 16x8 motion estimation. If (17) is not achieved, we then perform motion estimation for 8x16 and 16x8 modes and determine the final mode.

If the cost of 16x16 mode is larger than that of 8x8 mode, we then search the best 8x8 block subtype for each 8x8 block in the macroblock to get the candidate inter mode. The process from 8x8 mode to 4x4 mode is similar to the above process where each 8x8 block includes spatial homogeneous detection, temporal homogeneous detection, and cost analysis. Finally, the best inter mode is decided by selecting the inter mode with least cost from the candidate inter mode, 16x8 mode, and 8x16 mode.

IV. EXPERIMENTAL RESULTS

To evaluate our proposed algorithm, we test three sequences, Foreman, Mobile, and Bus with CIF format (352x288). And the experimental conditions are as follows: (1) the sequence type is IPPP…; (2) number of reference frame is 1; (3) search range is 33x33; (4) search algorithm is full search; (5) entropy coding method is CAVLC; (6) RD

optimization (RDO) is disabled; (7) Hadamard transform is used; and (8) no random intra macroblock refresh in P pictures; and (9) the frame rate is 30 fps. Fig. 3 shows the RD curves achieved by the original method, only spatial homogeneous detection method (OSHDM), and the proposed method in different sequences. For the OSHDM, the threshold for homogeneous detection is set to 20000 for 16x16 blocks and 5000 for 8×8 blocks, which is recommend by [3]. For the proposed method, the threshold is set to 5000 for 16×16 block and 1250 for 8×8 block to lessen the PSNR degradation. From Fig. 3, we can see that the three curves are almost the same in three sequences which means our proposed method does not introduce any noticeable PSNR loss compare to the original method. Table I shows the time saving of motion estimation process of our proposed method and the OSHDM in different Q_p. From Table I, we can see that in sequences with complex scene (Mobile, Bus), the time saving of the OSHDM is not so much. Especially in Mobile, the time saving is just about 8~9 % due to fewer homogeneous blocks. As for our proposed method, we can see that the time saving is still high even in complex sequences, which also demonstrates our proposed method is robust for various kinds of sequences. It is noticed that the time saving of our method is increase as Q_p increase, which is owing to the temporal homogeneous detection by using formula (9). Table II shows the mode number per macroblock of the original method, the OSHDM, and the proposed method. The mode number per macroblock is defined as the equivalent 16×16 motion estimation times for each macroblock. Hence, for the original method, the mode_number/macroblock is always 7. From Table II, we can see the similar result and our proposed method can reduce nearly 50% mode computation for different kinds of sequences.

V. CONCLUSIONS

In this paper, we first introduced the process for inter mode decision in H.264 and then proposed a fast inter mode decision algorithm base on hierarhical homogeneous detection and cost analysis. Experimental results show that our proposed algorithm can deal with not only sequences with spatial homogeneous blocks (ex. Foreman) but also complex sequences (ex. Mobile, Bus). By using the proposed algorithm, we can save about 32~54% coding time without introducing any noticeable performance degradation for various kinds of sequences, which is very helpful for H.264 interframe coding.

REFERENCES

[1] ITU-T Rec. H.264 / ISO/IEC 11496-10, "Advanced Video Coding", Final Committee Draft, Document JVTG050, March 2003.

[2] Thomas Wiegand, Gary J. Sullivan, Gisle Bjontegaard, and Ajay Luthra, "Overview of the H.264 / AVC Video Coding Standard," in IEEE Transactions on Circuits and Systems for Video Technology, July 2003.

[3] D. Wu, F. Pan, K. P. Lim, S. Wu, Z. G. Li, X. Lin, S. Rahardja, and C. C. K," Fast Intermode Decsion in H.264/AVC Video Coding," in IEEE Transactions on Circuits and Systems for Video Technology, July 2005.

[4] Christos Grecos and Ming Yuan Yang," Fast Inter Mode Prediction for P Slices in the H264 Video Coding Standad," in IEEE Transactions on Broadcasting, June 2005.

[5] Yong Ho Moon, Gyu Yeong Kim, and Jae Ho Kim ,"An Improved Early Detection Algorithm for All-Zero Blocks in H.264 Video Encoding," in IEEE Transactions on Circuits and Systems for Video Technology, August 2005.
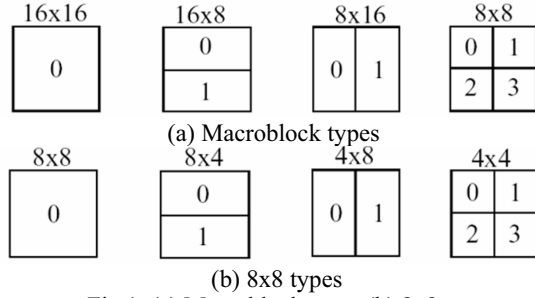
(a) Macroblock types

(b) 8x8 types

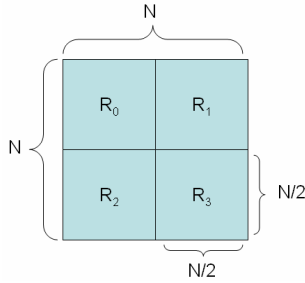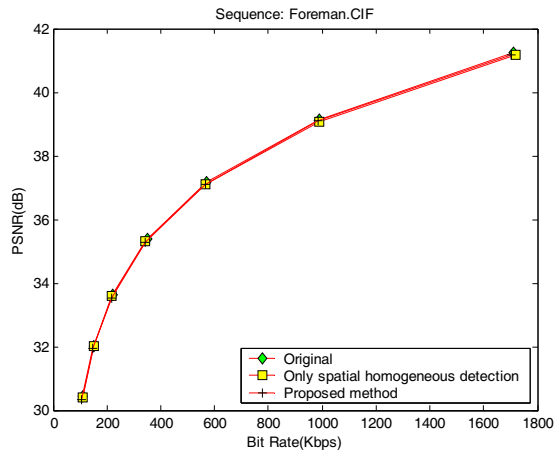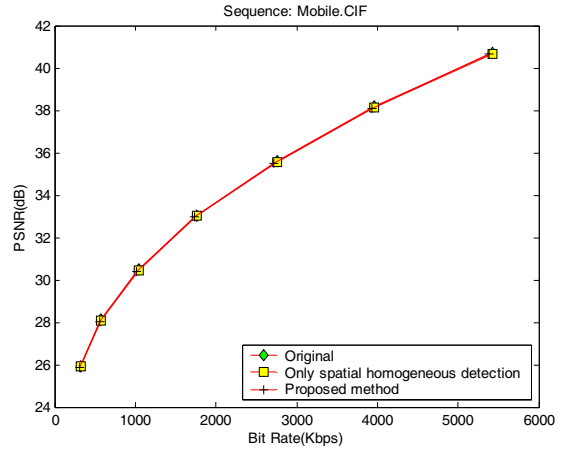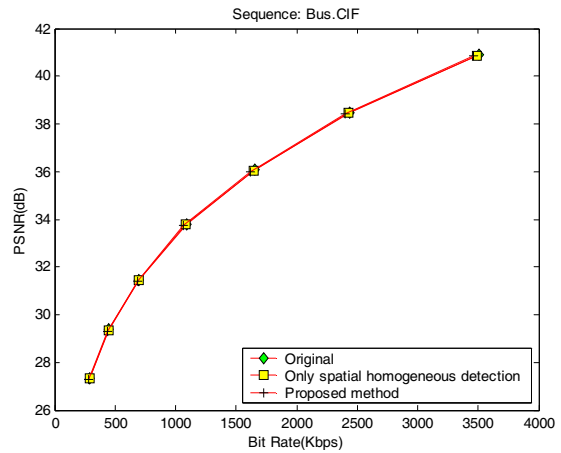Fig 1. (a) Macroblock types (b) 8x8 types

Fig 2. A Residue block with its sub-blocks.

(a) Foreman

(b) Mobile

(c) Bus

Fig 3. RD curve of sequence (a) Foreman (b) Mobile (c)Bus

Table I. Time-saving (ME) in different Qp

Sequence Qp	Foreman		Mobile		Bus	
	OSHDM	Proposed	OSHDM	Proposed	OSHDM	Proposed
26	39%	40%	9%	33%	19%	32%
32	32%	46%	8%	38%	17%	40%
38	30%	54%	9%	45%	12%	44%

Table II. Mode_number/macroblock in different Qp

Sequence Qp	Foreman			Mobile			Bus		
	Original	OSHDM	Proposed	Original	OSHDM	Proposed	Original	OSHDM	Proposed
26	7	4.11	3.51	7	6.07	4.21	7	5.55	4.10
32	7	4.11	2.92	7	6.07	3.82	7	5.55	3.53
38	7	4.11	2.31	7	6.07	3.28	7	5.55	3.02