HIERARCHICAL MODE SEARCH WITH CLASSIFICATION OF BISECTIONAL PREDICTION MODES BASED ON THE POSITION OF MOTION BOUNDARY

Sadaatsu Kato, Kazuo Sugimoto, Fulvio Moschetti, and Choong Seng Boon

Multimedia Laboratories, NTT DoCoMo, Inc.
3-5, Hikarinooka, Yokosuka, Kanagawa, Japan
{kato, ksugi, fulvio, boon}@mml.yrp.nttdocomo.co.jp

ABSTRACT

Video coding schemes equipped with various kinds of prediction modes improves coding efficiency by offering efficient prediction modes that adapt to the motion in a macroblock (MB). Operation time for motion estimation and mode selection, on the other hand, increases with the number of prediction modes. We have proposed a motion compensated prediction method (Arbitrary-shape Partitioning Motion Compensation: APMC) equipped with various inter prediction modes which adapt to the motion boundary of objects in an MB basis. This paper confirms the coding efficiency of APMC in comparison to the final specification of H.264/AVC and introduces a method to improve the computational performance of APMC by using a hierarchical mode search method based on motion boundary position within MBs. Simulation results show that APMC improves about 5% of coding efficiency in comparison with the final specification of H.264/AVC, and the proposed hierarchical mode search method decreases the operation time to about 60%, while maintaining the coding efficiency.

1. INTRODUCTION

The new video coding standard H.264/AVC [1] has attained high coding efficiency compared with the previous coding standards such as MPEG-4 part2 [2] or H.263 [3]. One of the main factors of the improvement is Rate-Distortion optimization (R-D optimization) which chooses the prediction mode of the highest coding efficiency out of various prediction modes. In motion estimation (ME), seven kinds of block size (16x16, 16x8, 8x16, 8x8, 8x4, 4x8, and 4x4) are provided as inter prediction modes, and R-D optimization process improves coding efficiency by choosing the best prediction mode according to the motion in a macroblock (MB).

The computation time of ME, on the other hand, occupies about 90% of the total operation time, and the increase in the computation time of ME directly affects the overall operation time. Therefore, many techniques to reduce the amount of operations in ME have been proposed. When there are less inter prediction modes, such as in MPEG-4 part-2 or H.263, many methods for limiting the motion search point of each prediction mode, like diamond search [4] or motion vector field adaptive search [5], are proposed. However, when there are more prediction modes, such as in H.264/AVC, it is required to reduce the amount of operations by searching only the prediction modes which can expect high coding efficiency. In [6], an intensive search, based on the block size, of H.264/AVC inter prediction modes is proposed. First, three modes, 16x16, 8x8, and 4x4, are estimated, and then, only the modes which have the block size between the most efficient two prediction modes in the first search are further examined.

As an extension of inter prediction modes of H.264/AVC, we have proposed Arbitrary-shape Partitioning Motion Compensation (APMC) in [7]. In the proposed modes, the number of motion vectors available in an MB is restricted to two. The asymmetrical horizontal, vertical and diagonal segmentations of the bisection prediction modes capture various motions across multiple MBs with just a small number of motion vectors. While improvement in coding efficiency is achieved by using this technique, the computation time also increases due to the additional prediction modes.

In this paper, we propose a method to reduce the amount of operations by using hierarchical mode search by classifying the prediction modes based on the position of motion boundary within MBs. The rest of the paper is organized as follows. In Section 2, we describe the inter prediction modes of H.264/AVC and APMC, and show the latest results of coding efficiency of APMC in comparison with the final specification of H.264/AVC. In Section 3, complexity reduction method of APMC based on classification and intensive mode search is described. Simulation results are shown in Section 4, and Section 5 concludes this paper.

2. MULTIPLE INTER PREDICTION MODES

2.1 Prediction Modes of H.264/AVC and APMC

H.264/AVC supports various block sizes of inter prediction modes as shown in Fig. 1. 16x16 pixels of an MB can be divided into 8x16, 16x8, and 8x8. 8x8 blocks can be further divided into 4x8, 8x4, and 4x4. Multiple reference frames are also supported. The encoder can use more than one previously decoded picture as the reference to encode the current picture. A motion vector is allocated to the each segment of the prediction modes, and each segment equal or larger than 8x8 is allowed to be predicted from different

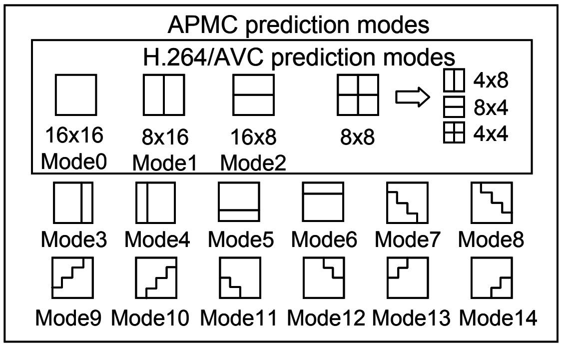

Fig.1. Inter prediction modes of H.264/AVC and APMC.

reference frames. The current reference software of H.264/AVC [8] uses R-D optimization for ME and mode decision. ME is first performed for all segments of prediction modes. The optimal motion vectors and reference frames are selected by minimizing both the bits for sending motion vectors and Sum of Absolute Differences (SAD) between the original signal and referenced signal using R-D optimization. After the calculation of the coding cost of all prediction modes, including that of intra prediction modes and SKIP mode, the mode of the minimum coding cost is selected as the optimum mode. As the inter prediction modes of H.264/AVC divides an MB or 8x8 block in a symmetrical manner, it is not capable of efficient representation of complicated shape of motion boundary, and results in the need to send many motion vectors. Thus it works well when sufficient bits allowance is available for motion representation, but not under stringent bits limitation due to the need to offer high compression ratios.

In addition to the H.264/AVC prediction modes, we have proposed additional bisectional segmentation modes, from Mode 3 to Mode 14, as shown in Fig. 1. We call the motion prediction scheme incorporate all these modes as APMC. A motion vector and reference frame is allocated to the each segment of an APMC bisectional prediction mode. Thus, the number of motion vectors available in an MB is restricted up to two. The feature of these modes is the introduction of asymmetrical horizontal, vertical and diagonal segmentations of an MB. In consideration of the 4x4 transform employed in H.264/AVC, so as to avoid signal discontinuity, the smallest unit of motion vector allocation is fixed as 4x4 pixel block for all modes.

ME and mode decision are the same as H.264/AVC. ME is first performed for all segments of APMC prediction modes, and the optimal motion vectors and reference frames are selected by minimizing the coding cost by using R-D optimization. After that, the mode of the minimum coding cost is selected as the final mode.

2.2 Performance of APMC

We incorporate the APMC prediction modes to the H.264/AVC inter prediction modes and evaluate its coding efficiency. The reference software used for this comparison corresponds to the final specification of H.264/AVC, which is slightly different during first proposal of APMC in [7]. The main difference of Inter prediction modes from [7] is that tree structured 8x8 sub-partitioning modes are introduced instead of the fixed-size MB partitioning modes of 8x8, 4x8, 8x4, and 4x4 in this simulation. Coding conditions are shown in Table 1. The search method on integer-pel positions is a full search, and for 1/2 and 1/4 pel search, only the closest 8 positions to the search center are checked. [9] is used for the comparison of coding efficiency. Simulation results in Table 2 shows that the proposed APMC prediction modes reduce the total coding bits up to 5.3% in comparison with the final H.264/AVC.

Table1. Simulation conditions.

Tested sequences	Foreman (QCIF) Mobile & Calendar (CIF) Flower garden (CIF)
Bitrate	30-130 Kbps (QCIF) 100-400 Kbps (CIF)
GOP	IPPP
Motion search accuracy	1/4 pel
Motion search range	16(QCIF), 32(CIF)
Number of reference frames	4
Entropy coding	CAVLC

Table 2. Bit reduction and Encoder execution time of APMC. (vs H.264/AVC)

Sequence	Foreman (QCIF)	Mobile (CIF)	Flower (CIF)
Bit reduction (%)	4.15	4.81	5.32
Encoder execution time	x 1.59	x 1.51	x1.43

3. IMPROVEMENT OF COMPUTANIONAL PERFORMANCE OF APMC

As seen in the preceding section, APMC with bisectional prediction modes achieves high coding efficiency. The total execution time of encoding, at the same time, increases about 50% from that of H.264/AVC, because the amount of operations in ME increased due to the additional 12 prediction modes. Therefore, computational performance of ME for APMC prediction modes should be improved, while maintaining the coding efficiency.

APMC prediction modes are designed based on the bisectional motion model where a motion boundary in an MB separates the region into segments of rectangular or diagonal shape. By changing the position of the boundary in an MB, various kinds of bisectional prediction modes are determined. Therefore, we propose a hierarchical mode search scheme which classifies the bisectional prediction modes based on the position of the boundary in an MB.

1) Classification of APMC prediction modes

If there is a motion boundary in a picture, a bisectional prediction mode whose border lies close to the motion boundary can adapt well to the motion field on both sides of the boundary, and show high coding efficiency. Thus, bisectional prediction modes with close boundary positions are classified into the same group.

2) Determination of parent modes

Among the bisectional prediction modes classified into the same group, a "parent" mode which represents the highest coding efficiency and the characteristic of the boundary position of the group is determined. The remaining prediction modes in the group are called "children" modes.

3) Hierarchical mode search

By comparing the coding cost of the parent modes, the best parent mode of the minimum coding cost is determined. Next, only the children modes attached to the best parent mode are further examined. The mode of the minimum coding cost is chosen as the best bisectional prediction mode.

By using this scheme, suitable prediction mode can be determined without the need to examine all the prediction modes, and consequently the overall amount of operations in ME can be reduced.

3.1 Determination of Parents Modes

Table 3 shows the probability of the best APMC mode selection in the simulation results of Section 2.2. This shows that probabilities of vertical and horizontal partitioning modes (Mode 1 - 6) are higher than those of diagonal prediction modes (Mode 7 - 14). In order to avoid the degradation of coding efficiency due to the hierarchical mode search, prediction modes with high probability are assigned as parent modes. Since the parent modes are always searched, it guarantees the best coding efficiency. Therefore, Mode 1 - 6 are determined as parent modes. Mode 0 is also treated as a parent mode.

Table3. Probability of the APMC mode selection

Skip	29.1%	Mode7	1.5%	8x8SubMB	10.0%
Mode0	19.7%	Mode8	1.6%	Intra4x4	2.1%
Mode1	7.2%	Mode9	1.8%	Intra16x16	1.0%
Mode2	6.3%	Mode10	1.8%		
Mode3	3.6%	Mode11	1.1%		
Mode4	4.1%	Mode12	1.2%		
Mode5	3.3%	Mode13	0.8%		
Mode6	2.9%	Mode14	0.7%		

3.2 Classification of Children Modes

Based on the 6 parent modes, the remaining prediction modes (children modes: Mode7 - 14) are classified into 6 groups. The closeness of the boundary position in an MB reflects the closeness of the parent and children modes. In this case, the extent of deviation between the two boundaries can be represented by the number of 4x4 pixel block. As shown in Fig.2, if the parent mode is Mode 1 and the child mode is Mode 7, for instance, the level of deviation between the two boundaries is 4 units of 4x4 pixel blocks. The smaller the level of deviation, the closer the boundary positions of the parent and children modes.

Table 4 shows the level of deviation between the boundary of a parent and a child mode. From this result, a child mode is attached to the parent modes which have the smallest level of deviation. Fig.3 shows the classification results of all prediction modes.

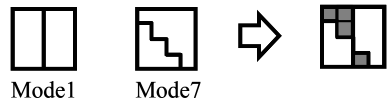

Fig.2. Deviation between the boundaries.

Table 4. Level of deviation between the boundaries of each parent mode and child mode

child \ parent	Mode1	Mode2	Mode3	Mode4	Mode5	Mode6
Mode7	4	4	6	4	4	6
Mode8	4	4	4	6	6	4
Mode9	4	4	6	4	6	4
Mode10	4	4	4	6	4	6
Mode11	5	5	7	3	3	7
Mode12	5	5	3	7	7	3
Mode13	5	5	7	3	7	3
Mode14	5	5	3	7	3	7

3.3 Hierarchical Mode Search

According to the determination of parent modes and classification of children modes, hierarchical mode search of APMC prediction modes are performed, and the best inter prediction mode in Mode 0 – 14 is chosen.

(Step1):
Calculate the coding cost of Mode 0 – 6 and choose the best parent mode.

(Step2):
if (best parent mode == Mode 0),
choose Mode 0 as the best mode.
otherwise,
calculate the coding cost of the children modes attached to the best parent mode to decide the best mode.

3.4 Analysis of the Amount of Operations in Hierarchical Mode Search of APMC

In a full search method, the amount of operations for ME on integer-pel position is proportional to the number of prediction modes. With the proposed hierarchical mode search scheme, if the best parent mode is Mode 0, only 7 parent prediction modes are searched. Otherwise, a total of 11 prediction modes are to be searched because each parent mode has 4 children. On the other hand, without the proposed hierarchical mode search scheme, all 15 prediction modes have to be searched. Therefore, the ratio of the operation reduction of ME for prediction modes of APMC using the hierarchical mode search is estimated as:

$$R = \frac{7 \times P\_mode0 + 11 \times (1 - P\_mode0)}{15} ,$$

where P_mode0 is the probability that Mode 0 is chosen as the best parent mode.

4. EXPERIMENTS

We adapted the proposed hierarchical mode search scheme in ME of APMC. Simulation conditions are the same as that of Table 1. Table 5 shows the ratio of the operation time reduction of ME for APMC prediction modes by using the hierarchical search measured from the

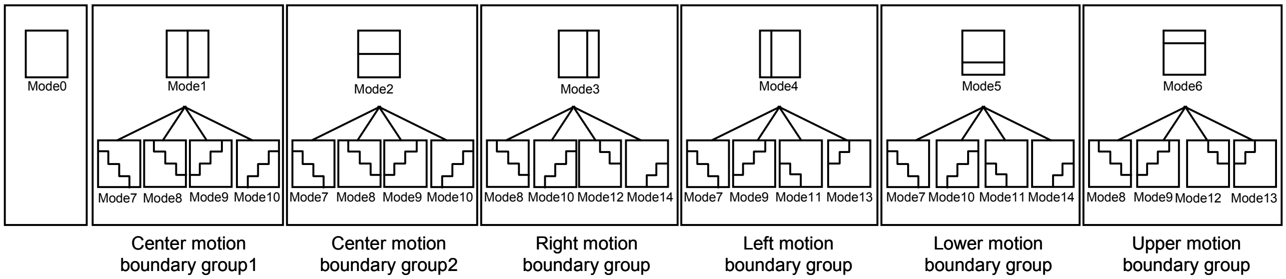

Fig. 3 Classification results of all prediction modes.

simulation results. From the table, operation time for the prediction modes is reduced to about 60%. Table5 also shows the *P_mode0* and theoretical estimated ratio of the operation time reduction. We observe that the expected operation time reduction is achieved from the results.

Table 6 shows the results of bit reduction and the ratio of overall encoder execution time of the H.264/AVC encoder with APMC (using the hierarchical search) and without APMC. The average of bit reduction is 4.5% and the degradation due to the hierarchical mode search is less than 0.5%. On the other hand, the increase of the encoder execution time from H.264/AVC is reduced to about 25%. From Fig.4, we can see that the proposed hierarchical mode search of APMC reduces overall encoder execution time, while maintaining the coding efficiency.

Table 5. Ratio of the operation time reduction of ME for APMC prediction modes by using the hierarchical prediction mode search.

Sequence	Foreman (QCIF)	Mobile (CIF)	Flower (CIF)
P_mode 0	0.40	0.53	0.51
R (estimated) (%)	62.7	57.2	57.9
R (measured) (%)	62.7	59.1	59.9

Table 6. Bit reduction and overall encoder execution time of APMC with hierarchical mode search. (vs H.264/AVC)

Sequence	Foreman (QCIF)	Mobile (CIF)	Flower (CIF)
Bit reduction (%)	3.97	4.40	5.01
Encoder execution time	x 1.32	x 1.24	x1.22

5. CONCLUSION

In this paper, we confirmed the performance of APMC in comparison with the final specification of H.264/AVC. An improvement of about 5% has been observed. We, then, proposed a hierarchical mode search scheme by classifying the bisectional prediction modes based on the position of the boundary in an MB. It was shown that the proposed hierarchical mode search method decreases the operation time of APMC to about 60%, while maintaining the coding efficiency.

Further study includes additional complexity reduction of APMC prediction modes, considering the effect of restricting the number of reference frames.

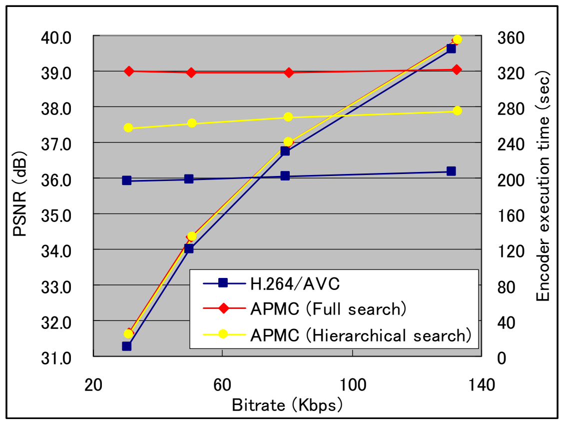

Fig.4. R-D curve and overall encoder execution time of H.264/AVC, APMC full mode search, and APMC hierarchical mode search. (Foreman, QCIF)

REFERENCES

[1] Joint Vide Team (JVT), "Study of Final Committee Draft of Joint Video Specification (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC)", Document JVT-F100d2, Generated on December 2002.

[2] "Information technology – Coding of audio-visual objects – Part 2: Visual, Amendment 1: Visual extensions", ISO/IEC JTC1/SC29/WG11/N3056, Dec, 1999

[3] G.Sullivan, "Draft Text of Recommendation H.263 Version 2("H.263+") for Decision", Jan, 1998

[4] S.Zhu, and K.K.Ma, "A New diamond search algorithm for block motion estimation", IEEE Trans. Image Processing, Vol. 9, No.2, pp.287-290, Feb, 2000

[5] A.M. Tourapis, O.C. Au, and M.L. Liou, "Predictive motion vector field adaptive search technique (PMVFAST) - Enhancing Block Based Motion Estimation", Proc. VCIP2001, Vol.4310, pp.883-892. Jan, 2001

[6] P.Yin, H. Tourapis, A.Tourapis, and J. Boyce, "Fast Mode Decision and Motion Estimation for JVT/H.264", Proc. ICIP2003, pp.853-856, Sep, 2003

[7] S.Kato, S.Sekigutchi, S.Adachi, and M.Etoh, "Performance evaluation on H.26L-based motion compensation with segmented multiple reference frames", Proc. ICIP2002, Vol.1, pp.22-25, Sep, 2002

[8] JVT Reference Software version 6.1e, http://bs.hhi.de/~suehring/tml/download/

[9] G.Bjontegaard, "Calculation of average PSNR differences between RD-curves", VCEG-M33, Apr, 2001