A FAST INTER-MODE DECISION ALGORITHM BASED ON MACRO-BLOCK TRACKING FOR P SLICES IN THE H.264/AVC VIDEO STANDARD

Byung-Gyu Kim and Chang-Sik Cho

Embedded S/W Research Division,
Electronics and Telecommunications Research Institute (ETRI)
161 Gajeong-dong, Yuseong-gu, Daejeon 305-350, Korea
E-mail: bgkim@etri.re.kr, cscho@etri.re.kr

ABSTRACT

A fast inter-mode determination algorithm based on the macro-block (MB) tracking scheme and rate-distortion (RD) cost is proposed for the H.264/AVC video standard in which residual prediction is composed of intra-modes and inter-modes. In addition to intra-mode prediction, 8 block types exist for the best coding gain based on rate-distortion (RD) optimization in the inter-mode prediction. This scheme gives rise to exhaustive computations (search) in the coding procedure. To reduce the computational load of the inter-mode search at the inter-frame, we propose a new inter-mode determination algorithm based on the rate-distortion (RD) cost of the neighborhood MB that is tracked for the current MB in the previous frame. Based on the MB tracking scheme, an efficient sequential mode search approach is presented. We verify the performance of the proposed scheme through comparative analysis of experimental results using JM reference software.

Index Terms— Image coding, Prediction methods, Video coding, Video processing, Image motion analysis

1. INTRODUCTION

Developments in video coding techniques have accelerated over the last several years. H.264/AVC video coding is the newest standard defined by the Joint Video Team (JVT) [1] for which various techniques have been adopted to obtain a high coding efficiency, compared to previous standards.

Among the techniques of H.264/AVC video coding, the motion estimation routine is better than the previous video standards, such as H.261 and H.263 [1]. Generally, motion estimation for inter-prediction is performed only on a 16×16 or an 8×16 macroblock (MB), then each 16×16 or 8×16 MB is assigned one motion vector, which results in minimum block distortion.

The variable block size for inter-mode prediction allows maximization of the coding efficiency based on rate-distortion (RD) optimization in the H.264/AVC coding standard [1], including $SKIP$, 16×16, 16×8, 8×16, 8×8, 8×4, 4×8, and 4×4 blocks (Fig. 1). In addition, intra-mode prediction follows inter-mode prediction to determine the best residual image [1]. These structures give rise to complexity of the encoder for application in real-time situations. Therefore, a fast motion estimation scheme that can reduce the complexity of the H.264/AVC video encoder for the full-mode decision is needed.

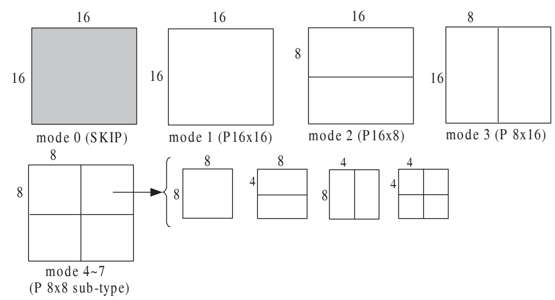

Fig. 1. Variable block types for motion estimation in the H.264/AVC video encoding.

Many kinds of fast inter-mode decision schemes with rate-distortion (RD) optimization have been reported [2]-[7]. Several approaches use the frame difference or the MB difference while others use a spatial homogeneity or a temporal prediction for $SKIP$ MB detection [2]-[7].

In this paper, we propose a fast inter-mode determination algorithm on the basis of the rate-distortion (RD) cost of the tracked MB for the current MB in the previous frame. To do this, we introduce a simple MB tracking scheme.

2. THE PROPOSED FAST MODE DECISION ALGORITHM

Two consecutive frames in a video sequence are highly correlated. Based on this temporal correlation, many techniques for fast inter-mode decision have been reported [5]-[6]. If there is a stationary background in the video sequence, MB mode types that belong to the background will be identical in the same regions (the colocated MB) of the previous frame.

With slow object motion in the video, the mode information for the same MB in the previous frame may still affect the mode determination process of the current MB because of the high temporal correlation. With fast motion there is little temporal correlation between successive frames (or MBs). In this case, a scheme based on spatial correlation rather than temporal correlation is needed.

In this study, we propose an MB tracking scheme using temporal correlation for all motion cases. Early extraction of the stationary block type, such as $SKIP$ and $P16 \times 16$, in a sequential mode search approach is used.

2.1. A Simple MB Tracking Strategy

To track an object is to locate the current object region in an adjacent frame. To do this, we need to define the desired object region in the image plane. We need a search procedure for the desired object region with a predefined tracking criterion in the temporal domain.

To apply this tracking scheme to block-based video coding, we consider each MB as a desired object in a mode decision procedure. As shown in Fig. 2, a $P16 \times 16$ MB type is used to locate the region in the previous frame with the highest correlation. This is an integer pel motion estimation procedure for the current MB.

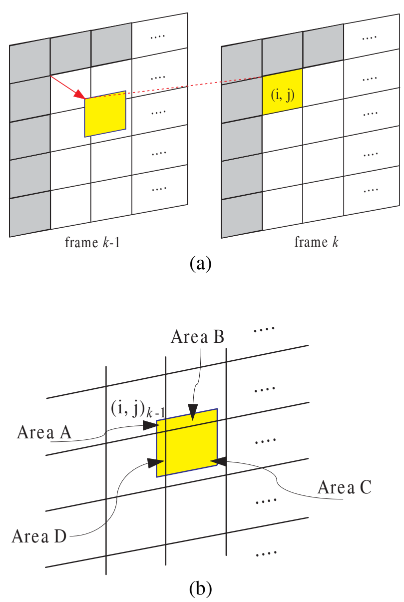

Fig. 2. A MB tracking scheme using $P16 \times 16$ block motion estimation.

Once the best motion vector and most highly correlated region are obtained, we determine the most highly correlated MB of the current MB in the previous frame, as follows:

$$\text{Max}_{\text{MB}(k,\, l)}\{\text{Area A, Area B, Area C, Area D}\}, \quad (1)$$

where (k, l) denotes an index of MB that contains the overlapped (correlated) region of the current MB. We use the above equation to determine that MB which has the maximum correlation with the current MB and also it is more than some predefined threshold τ. We have chosen as $\tau = 0.85 \sim 0.9$ to guarantee enough correlation between two MBs.

2.2. Sequential Mode Search Approach Using Temporal Correlation

Temporal correlation between successive frames can provide good information for an efficient mode search. In many cases many stationary MBs or regions exist. If we can detect these MBs early using a simple prediction technique, the speed of the mode decision procedure will increase.

Experimental results are presented in Table 1 as statistical tests for the two kinds of MBs that can be thought of as part of the stationary region. We can see that the probabilities of the selected mode type of the current MB under that the colocated MBs as $SKIP$ or $P16 \times 16$ block types. We have used four sequences (*Carphone*, *Foreman*, *Stefan*, and *Salesman*) with 100 frames each for experimental analysis.

Table 1. Statistical results for stationary MBs (unit: %).

Modes		$QP = 24$	$QP = 28$	$QP = 32$
$SKIP$	$SKIP$	60.86	65.21	69.94
	$P16 \times 16$	20.03	17.49	16.54
	$P16 \times 8$	5.82	5.44	4.69
	$P8 \times 16$	5.62	6.00	5.33
	$P8 \times 8$ sub-type	6.08	4.69	2.75
$P16 \times 16$	$SKIP$	22.04	24.59	30.03
	$P16 \times 16$	35.83	36.04	35.87
	$P16 \times 8$	10.39	10.28	10.51
	$P8 \times 16$	11.08	11.31	11.61
	$P8 \times 8$ sub-type	20.32	17.39	11.35

With the SKIP mode of the colocated MB, the probability that the current MB will be $SKIP$ or $P16 \times 16$ mode is greater than 80%. The probability that the MB will be $SKIP$, $P16 \times 16$, $P16 \times 8$, and $P8 \times 16$ is greater than 91%. There is a high probability for the MB to be $SKIP$, $P16 \times 16$, $P16 \times 8$, and $P8 \times 16$ if the mode of the colocated MB is $SKIP$ in the previous frame. When the colocated MB is $P16 \times 16$ mode, the probability to be $SKIP$, $P16 \times 16$, $P16 \times 8$, and $P8 \times 16$ is less than the $SKIP$ colocated MB. This probability of about 80% is still valuable information for evaluation of a fast algorithm.

Using these properties, we can consider the candidate modes of the current MB as modes that are sequential down to the mode of the colocated MB. For example, candidate modes of the current MB will be $\{SKIP,\ \ P16 \times 16\}$ if the mode of the colocated MB is $P16 \times 16$. Also, candidate modes of the current MB will be $\{SKIP,\ \ P16 \times 16,\ \ P16 \times 8\}$ if the mode of the colocated MB is $P16 \times 8$. When the mode of the colocated MB is a $P8 \times 8$ sub-type, we consider all eight possible modes ($SKIP \sim P4 \times 4$) as candidate modes.

Using the suggested candidate selection method and the simple MB tracking concept, we can summarize our algorithm based on an adaptive RD thresholding technique for the mode of the colocated MB, as follows:

Step 1) If the colocated MB is $SKIP$ mode: Set candidate mode as $\{SKIP\}$, then compute the minimum RD cost of the candidate mode. If the minimum RD cost is less than the cost of the colocated MB, the final mode will be the one that has the minimum RD cost and no intra mode search. Otherwise, $\{P16\times16,\ P16\times8,\ P8\times16\}$ will be added as candidate modes. The mode that has the minimum RD cost will be the final mode for the current MB.

Step 2) If the colocated MB is $P16 \times 16$ mode: Set candidate modes as $\{SKIP,\ P16\times16\}$, then compute the minimum RD cost of the candidate modes. If the minimum RD cost is less than the cost of the tracked MB (in the previous frame) using the proposed MB tracking scheme, the final mode will be one that has the minimum RD cost and no intra mode search. Otherwise, we compute a binary pattern for the current MB using spatial intensity as in Fig. 3 (a). Let μ_i be the average intensity of each 8×8 block and μ_T be the average value of intensity for the current MB (16×16). If the average intensity of any 8×8 block is larger than μ_T, then we set 1. Otherwise 0 will be set. After computation of a binary pattern for a refinement stage, if the texture occurs like described in Fig. 3 (b), the remaining modes $\{P16\times8,\ P8\times16,\ P8\times8$ sub-types$\}$ are added as candidate modes. Otherwise $\{P16\times8,\ P8\times16\}$ are added as candidate modes. The mode that has the minimum RD cost will be the final mode for the current MB.

Step 3) If the colocated MB is $P16 \times 8$ mode: Set candidate modes as $\{SKIP,\ P16\times16,\ P16\times8\}$, then compute the minimum RD cost of candidate modes. If the minimum RD cost is less than the cost of the tracked MB (in the previous frame) using the proposed MB tracking scheme, the final mode will be one that has the minimum RD cost and no intra mode search. Otherwise, the remaining modes $\{P8\times16,\ P8\times8$ sub-types$\}$ are added as candidate modes. The mode that has the minimum RD cost will be the final mode for the current MB.

Step 4) If the colocated MB is $P8 \times 16$ mode: Set candidate modes as $\{SKIP,\ P16\times16,\ P8\times16\}$, then compute the minimum RD cost of the candidate modes. If the minimum RD cost is less than the cost of the tracked MB (in the previous frame) using the proposed MB tracking scheme, the final mode will be one that has the minimum RD cost and no intra mode search. Otherwise, the remaining modes $\{P8\times16,\ P8\times8$ sub-types$\}$ are considered as other candidate modes. The mode that has the minimum RD cost will be the final mode for the current MB.

Step 5) If the colocated MB is $P8 \times 8$ mode: Set all possible eight modes as mentioned above, then the mode that has the minimum cost will be the final mode for the current MB.

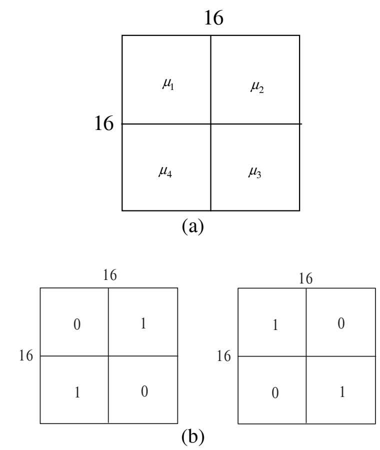

Fig. 3. A Binary pattern for the refinement stage of the $P16 \times 16$ mode: (a)8×8 blocks and (b)binary patterns for searching all modes.

3. RESULTS AND DISCUSSIONS

To verify the performance of the proposed fast mode decision algorithm, various MPEG standard sequences were used with CIF and QCIF sizes. Analyses were performed with encoding frames=100, RD optimization enabled, QP = 24, 28, and 32, sequence types of $IPPP$ in the Main profile, using CAVLC, with a search range of MV = ±16, and the number of the reference frames = 1. The Hadamard transform option was turned on also.

JM 11.0 reference software of the JVT was used as a reference code for evaluation of the encoding performance. We defined three measures for evaluating the encoding performance, including average $\Delta PSNR$, average $\Delta Bits$, and an encoding-time saving factor, ΔT. The average $\Delta PSNR$ is the difference in (dB) between the average $PSNR$ of the proposed method and the corresponding value of another method. As performance improves, this criterion becomes larger. The average $\Delta Bits$ is the bit-rate difference as a percentage (%) between the compared methods. Lastly, the encoding-time saving factor ΔT is defined for complexity comparison as:

$$\Delta T = \frac{T_{ref} - T_{proposed}}{T_{ref}} \times 100(\%), \qquad (2)$$

under the condition that the full mode search (FMS) is optimum (reference). As this value increases, the performance speed is increased. Also, it must be noted that positive values for the $\overline{PSNR}$ and $\Delta Bits$ indicate increments, and negative

Table 2. Performance comparison of the proposed algorithm on the JM 11.0 reference encoder for $IPPP$ sequences.

Contents		QP=24 $\Delta PSNR$	QP=24 $\Delta Bits$	QP=24 ΔT	QP=28 $\Delta PSNR$	QP=28 $\Delta Bits$	QP=28 ΔT	QP=32 $\Delta PSNR$	QP=32 $\Delta Bits$	QP=32 ΔT
Stefan(QCIF)	Alg-1	-0.029	0.902	4.42	-0.037	0.865	7.67	-0.035	0.898	9.27
	Alg-2	-0.496	0.040	49.09	-0.257	3.425	42.24	-0.028	8.316	39.36
	Proposed	**-0.059**	**0.258**	**44.82**	**-0.058**	**0.330**	**43.84**	**-0.065**	**0.988**	**43.95**
Salesman(QCIF)	Alg-1	-0.017	-0.148	15.72	-0.005	-0.176	13.38	-0.061	-0.397	10.76
	Alg-2	-0.148	2.404	76.81	0.024	19.05	59.38	-0.097	24.91	55.73
	Proposed	**-0.049**	**0.546**	**70.59**	**-0.025**	**0.645**	**61.41**	**-0.005**	**-0.709**	**62.74**
Mobile(CIF)	Alg-1	-0.041	0.219	8.80	-0.048	0.196	7.42	-0.044	0.288	7.17
	Alg-2	-0.426	0.935	46.08	-0.279	2.121	39.54	-0.044	9.431	35.57
	Proposed	**-0.056**	**0.100**	**37.42**	**-0.052**	**-0.015**	**39.58**	**-0.054**	**-0.035**	**41.63**
Paris(CIF)	Alg-1	-0.043	0.399	8.80	-0.045	0.264	9.24	-0.037	0.399	10.22
	Alg-2	-0.416	3.310	65.55	-0.068	9.96	52.12	0.017	17.61	49.11
	Proposed	**-0.065**	**0.729**	**54.67**	**-0.061**	**0.997**	**54.59**	**-0.056**	**0.794**	**56.47**
Foreman(CIF)	Alg-1	-0.025	0.409	8.90	-0.021	0.335	11.36	-0.033	0.694	10.93
	Alg-2	-0.415	9.158	48.77	-0.141	13.264	45.10	-0.011	17.794	45.65
	Proposed	**-0.061**	**1.037**	**43.26**	**-0.072**	**1.486**	**46.06**	**-0.091**	**1.574**	**50.55**
Mother& Daughter.(CIF)	Alg-1	-0.0073	0.199	14.18	-0.001	-0.249	18.50	-0.003	-0.350	30.474
	Alg-2	-0.343	2.874	71.82	0.070	15.605	59.97	0.099	20.205	58.83
	Proposed	**-0.076**	**0.849**	**60.12**	**-0.017**	**1.338**	**60.28**	**-0.030**	**0.917**	**61.71**

values indicate decrements.

We used two methods for an objective comparison of the encoding performance. These are Jing's [4] and Salgado's [5] methods, which are well-known as efficient and fast algorithms. In the table, Jing's method is denoted as Alg-1 and Salgado's method as Alg-2.

Table 2 shows the results of all algorithms for the IPPP sequence type. The proposed algorithm achieves a better speedup factor with a minimal loss of image quality and a minimal bit increment. In most sequences, Salgado's method yields a large bit increment and $PSNR$ loss. For QP=24 of the Salesman sequence, results show that the encoding speed can be improved up to 70.59% with an increase of 0.546% in the bits. The proposed scheme requires a negligible bit increment with a minimal loss of image quality, but provides a higher speed-up factor in most sequences.

Jing's (Alg-1) method achieves good $PSNR$ and bit performance with a lower speed-up factor than the other two methods. From this result, we can see that the proposed scheme can achieve an improvement of up to 40%$\sim$70% with the average loss of 0.052 (dB) and bit increment of 0.708%, in the total encoding time.

4. CONCLUSIONS

We have proposed a fast inter-mode determination algorithm based on the macro-block (MB) tracking scheme and rate-distortion (RD) cost for P-slices in H.264/AVC video coding. The proposed fast inter mode decision algorithm yields good performance because of an adaptive RD thresholding scheme using the RD cost of the most correlated MB that was tracked in the previous frame for the current MB, except for $SKIP$ mode. Also, the technique for the selection of the partial candidate modes provides improved speed for the motion estimation procedure. Through comparative analysis, a speed-up factor of 40%$\sim$70% was verified with a negligible bit increment and a minimal loss of image quality.

5. REFERENCES

[1] T. Wiegand, G. J. Sullivan, G. Bjontegard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," IEEE Trans. Circuit Syst. Video Technol., vol. 13, pp. 560 - 576., July 2003.

[2] D. Zhu, Q. Dai and R. Ding, "Fast Inter Prediction Mode Decision for H.264," Proc. of IEEE International Conference on Multimedia and Expo (ICME), vol. 1, pp. 1123-1126, 2004.

[3] D. Wu, S. Wu, K. P. Lim, F. Pan, Z. G. Li, and X. Lin, "Block Inter Mode Decision for fast Encoding of H.264," Proc. of IEEE International Conference on Acostics, Speech, and Signal Processing (ICASSP), vol. 3, pp. 181-184, 2004.

[4] X. Jing and L.-P. Chau, "Fast approach for H.264 inter mode decision," Electroincs Letters, Vol. 40, No. 17, September 2004.

[5] L. Salgado, and M. Nieto, "Sequence Independent Very Fast Mode Decision Algorithm on H.264/AVC Baseline Profile," Proc. of the International Conference on Image Processing 2006, Atlanta USA, pp. 41-44, Oct. 2006.

[6] C. Crecos and M. Y. Yang, "Fast Inter Mode Prediction for P Slices in the H264 Video Coding Standard," IEEE Trans. Broadcasing, vol. 51, no. 2, pp. 256 - 263, June 2005.

[7] Xin and Vetro, "Fast Mode Decision for Intra-only H.264/AVC Coding," Prof. of the 25-th Picture Coding Symposium (PCS), April 2006.