Fast Inter Mode Decision Algorithm Based on Macroblock Tracking in H.264/AVC Video

Byung-Gyu Kim, Jong-Ho Kim, and Chang-Sik Cho

We propose a fast macroblock (MB) mode prediction and decision algorithm based on temporal correlation for P-slices in the H.264/AVC video standard. There are eight block types for temporal decorrelation, including SKIP mode based on rate-distortion (RD) optimization. This scheme gives rise to exhaustive computations (search) in the coding procedure. To overcome this problem, a thresholding method for fast inter mode decision using a MB tracking scheme to find the most correlated block and RD cost of the correlated block is suggested for early stop of the inter mode determination. We propose a two-step inter mode candidate selection method using statistical analysis. In the first step, a mode is selected based on the mode information of the co-located MB from the previous frame. Then, an adaptive thresholding scheme is applied using the RD cost of the most correlated MB. Secondly, additional candidate modes are considered to determine the best mode of the initial candidate modes that does not satisfy the designed thresholding rule. Comparative analysis shows that a speed-up factor of up to 70.59% is obtained when compared with the full mode search method with a negligible bit increment and a minimal loss of image quality.

Keywords: H.264/AVC, inter mode, macroblock (MB) tracking, adaptive thresholding, rate-distortion optimization.

Manuscript received Mar. 26, 2007; revised Sept. 10, 2007.

This work was supported by the IT R&D program of MIC/IITA, Rep. of Korea [2005-S-022-02, Embedded Software-based SmarTown Solution].

Byung-Gyu Kim (phone: + 82 42 860 5766, email: bgkim@etri.re.kr), Jong-Ho Kim (email: pooney@etri.re.kr), and Chang-Sik Cho (email: cscho@etri.re.kr) are with Embedded SW Research Division, ETRI, Daejeon, Rep. of Korea.

I. Introduction

The H.264/AVC video coding standard is the newest standard which has been defined by the Joint Video Team (JVT) [1], [2]. In this video coding standard, various techniques have been adopted to obtain high coding efficiency compared to previous standards.

Among the adopted techniques of H.264/AVC video coding, a motion estimation routine with seven variable blocks stands out from the previous video standards, such as MPEG-2 Video, MPEG-4 Visual, H.261, and H.263 [1], [2]. Generally, motion estimation for inter prediction is performed only on a 16×16 macroblock (MB). Then, each 16×16 MB is assigned one motion vector or two motion vectors for *B* frames which can lead to minimum block distortion.

Various block sizes for inter mode prediction are allowed in order to maximize coding efficiency based on rate-distortion (RD) optimization in the H.264/AVC coding standard [2]. They are SKIP, 16×16, 16×8, 8×16, 8×8, 8×4, 4×8, and 4×4 blocks as shown in Fig. 1. In addition, intra mode prediction

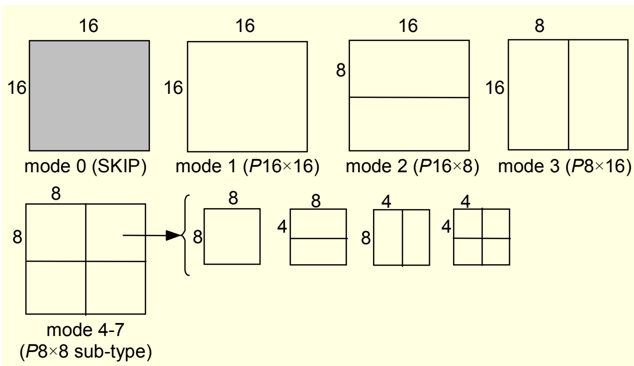


Fig. 1. Various block modes in H.264/AVC video standard.

(nine 4×4 and four 16×16 modes) follows the inter mode prediction to find the best residual image [1]-[4]. For any block mode, the motion vector estimation for a minimum residual image can also be performed by full search or fast motion search methods, such as the directional diamond search [5] and the fast-adaptive rood pattern search [6]. These structures for mode decision give rise to heavy complexity of the encoder when it is applied in real-time applications. Therefore, it is desirable to design a fast motion estimation scheme which can reduce the complexity of the H.264/AVC video encoder that results from full-mode decision.

Many kinds of fast inter/intra mode decision schemes have been reported with RD optimization (RDO) [7]-[17].

$$J_{RD} = SSD_{Mode} + \lambda \cdot \{R(Header) + R(Residual)\}, \quad (1)$$

where J_{RD} is a bitrate-distortion value as a cost function, SSD_{Mode} is the sum of the absolute difference between pixel values of the image for the given mode, λ denotes the Lagrangian multiplier, $R(x)$ is a bit amount for coding x, *Header* provides header information, and *Residual* is the residual data for the given MB.

Kim and others [7] proposed an algorithm using the property of an all-zero coefficient block which is produced by quantization and coefficient thresholding to effectively eliminate unnecessary inter modes. In this method, there are several threshold values which should be predefined. Also, transform coefficients are needed to make a fast decision. Jing and Chau [13] developed a fast inter mode decision scheme by using both the frame difference and the macroblock difference. This is a useful scheme, since only frame difference and macroblock difference images are required to assign a mode.

A fast algorithm using a ratio measurement of the RD costs of two classes defined as class 16 ({SKIP, *P*16×16, *P*16×8, *P*8×16}) and class 8 ({*P*8×8, *P*8×4, *P*4×8, *P*4×4}) has been proposed by Lee and Jeon [8]. In [11] and [12], Wu and others make use of the spatial homogeneity of a video object's textures and the temporal stationary characteristics inherent in the video sequence. However, the method suffers from a drawback in obtaining the edge image for the texture information and the difference image. These operations (the edge and difference images) are an additional load for this algorithm. Also, the defined threshold values (Thd_H and Thd_S) should be selected properly. The performance of this method is largely dependent on these parameters. Zhu and others [9], [10] proposed another approach for a fast inter mode decision that uses a pre-encoding, down-sampled image space. After obtaining candidate block modes, a refinement search is performed to find the best mode in the original image space. To detect an early SKIP block, Grecos and Yang [14], [15], also used the prediction and thresholding scheme. Salgado has also suggested a method using temporal correlation and sequential mode search [17]. This method illustrates a high speed-up ratio for the baseline profile in the encoding time. However, it incurs a considerable quality loss in terms of PSNR and bitrate.

To reduce the load of the mode search procedure effectively, we utilize the statistical properties for choosing candidate modes for the search procedure. Since there always is a high correlation between two successive frames or MBs, most of the current MBs do not vary widely from the corresponding MB in the previous frame. We propose an efficient thresholding scheme for early termination of the mode search based on the RD cost of the most correlated MB, which can be found by a simple MB tracking scheme using a *P*16×16 block type in the previous frame.

Section II describes our proposed algorithm and observations in detail. We verify the performance of the proposed algorithm in section III, and conclusions are presented in section IV.

II. The Proposed Fast Mode Decision Algorithm

There are several algorithms for fast inter mode decision based on a thresholding scheme using the value of the RD function. Most of them use the minimum RD cost value of the neighborhood regions of the current MB or pre-tuned value through extensive experiments. 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 [15]-[17]. If there is a stationary background in a 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 a 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 *P*16×16, in a sequential mode search approach is used based on the candidate mode selection method.

1. Sequential Candidate Modes Selection 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

Table 1. Statistical results for the mode change of MBs as temporal relationship (unit: %).

Mode transition ($MB^{t-1}_{ij} \rightarrow MB^{t}_{ij}$)		*QP*=24	*QP*=28	*QP*=32
SKIP (mode 0)	SKIP	60.86	65.21	69.94
	*P*16×16	20.03	17.49	16.54
	*P*16×8	5.82	5.44	4.69
	*P*8×16	5.62	6.00	5.33
	*P*8×8 sub-type	6.08	4.69	2.75
*P*16×16 (mode 1)	SKIP	22.04	24.59	30.03
	*P*16×16	35.83	36.04	35.87
	*P*16×8	10.39	10.28	10.51
	*P*8×16	11.08	11.31	11.61
	*P*8×8 sub-type	20.32	17.39	11.35
*P*16×8 (mode 2)	SKIP	17.59	20.82	25.97
	*P*16×16	26.55	27.06	27.66
	*P*16×8	18.60	19.36	19.55
	*P*8×16	9.02	9.85	10.56
	*P*8×8 sub-type	27.91	22.47	15.82
*P*8×16 (mode 3)	SKIP	15.14	19.14	24.63
	*P*16×16	24.68	25.53	27.43
	*P*16×8	10.43	9.36	9.89
	*P*8×16	17.66	20.48	21.10
	*P*8×8 sub-type	31.36	25.11	16.59
*P*8×8 (mode 4 (sub-block type))	SKIP	7.62	10.16	11.36
	*P*16×16	19.14	20.34	24.40
	*P*16×8	8.41	9.62	11.86
	*P*8×16	9.53	12.59	13.99
	*P*8×8 sub-type	54.96	46.81	38.08

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, especially focusing on the two kinds of MBs (SKIP and *P*16×16 block types) that can be thought of as part of the stationary region. The average probabilities of mode transition to the current MB from the available modes of the collocated MB are shown in this result. We have used six sequences (Mobile, Flower Garden, Carphone, Foreman, Stefan, and Salesman) with 100 frames, CIF and QCIF size, and 30 fps (frames per second) for experimental analysis.

With the SKIP mode of the colocated MB, the probability that the current MB will be SKIP or *P*16×16 mode is greater than 80%. The probability that the MB will be SKIP, *P*16×16, *P*16×8, or *P*8×16 is greater than 91%. There is a high probability for the MB to be SKIP, *P*16×16, *P*16×8, or *P*8×16 if the mode of the colocated MB is SKIP in the previous frame. When the colocated MB is *P*16×16 mode, the probability that it will be SKIP, *P*16×16, *P*16×8, or *P*8×16 is less than the SKIP colocated MB. This probability of about 80% is still valuable information for evaluation of a fast algorithm. When the mode of the colocated MB is a rectangular block (*P*16×8 or *P*8×16), the probability that it will be SKIP, *P*16×16, *P*16×8, and *P*8×16 is also over 70%.

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 from mode 0 to mode 4 as shown in Fig. 1. For example, candidate modes of the current MB will be {SKIP, *P*16×16} if the mode of the colocated MB is *P*16×16. Also, candidate modes of the current MB will be {SKIP, *P*16×16, *P*16×8} if the mode of the colocated MB is *P*16×8. When the mode of the colocated MB is a *P*8×8 sub-type, we consider all eight possible modes (SKIP, *P*16×16, *P*16×8, *P*8×16, *P*8×8 (4 sub-types)) as candidate modes.

2. Simple MB Tracking

To track an object is to locate the current object region in an adjacent frame. To do this, the desired object region must be defined in the image plane. A search procedure for the

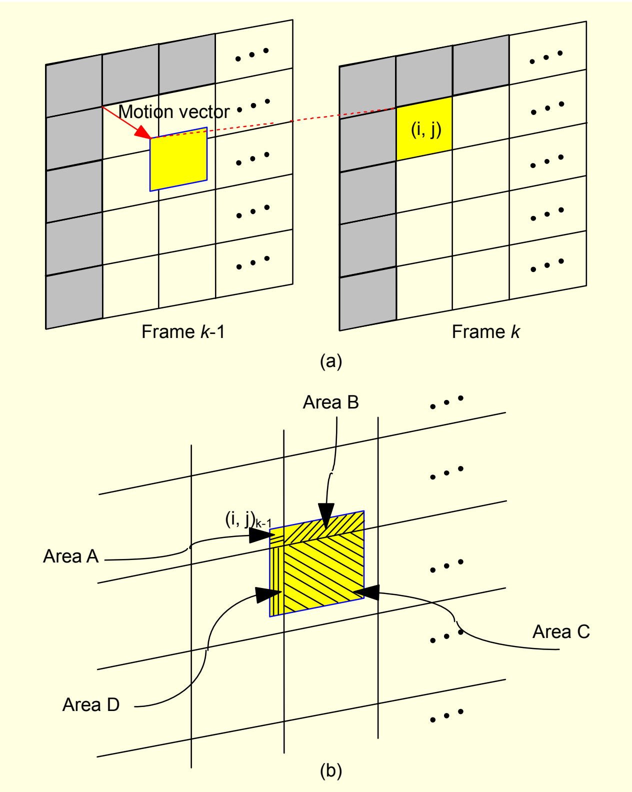


Fig. 2. MB tracking scheme using *P*16×16 block motion estimation.

desired object region with a predefined tracking criterion in the temporal domain is needed.

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 *P*16×16 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.

Once the best motion vector and most highly correlated region are obtained from the previous frame and decoded reference frame, we determine the most highly correlated MB from among the neighboring MBs that contains the A, B, C, and D regions (Fig. 2) as

$$\mathrm{Max}_{MB(k,l)}\{\text{Area A, Area B, Area C, Area D}\}, \qquad (2)$$

where (*k*, *l*) denotes an index of MBs which contains the overlapped (correlated) region of the current MB. We use (2) to determine which MB has the maximum correlation with the current MB. Also, it is more than some predefined threshold, τ. The proposed MB tracking scheme can be summarized in the following steps.

Step 1. Locate the correlated region (region *A*, *B*, *C*, and *D* in (2)) with a motion vector of *P*16×16 mode.

Step 2. Locate the most correlated MB with the highest correlation ratio.

Step 3. Use the predefined parameter τ to measure MB similarity.

Our algorithm is applied to the MB with the highest correlation ratio (Corr in Fig. 3) which is larger than the predefined value. In order to guarantee a sufficient degree of correlation between the two MBs, it is desirable to determine the MB when the correlation ratio of the acquired MB to the current MB is greater than the predefined threshold τ. We have chosen τ =0.85~0.9 based on experimental observations. After obtaining the most correlated MB in the previous frame, the RD cost value of this MB is used for an adaptive decision of the early termination for the current MB in the mode search procedure.

3. Overall Scheme of Proposed Algorithm

Figure 3 illustrates the overall scheme of the proposed inter mode decision algorithm. 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 an initial candidate mode (Mode_n in Fig. 3) as {SKIP} only. Then, compute the minimum RD cost (RDC_m in Fig. 3) of the candidate mode. If the minimum RD cost is less than the cost (RDC^{k-1}_{pq} in Fig. 3) of the colocated MB, the final mode (mode_m in Fig. 3) will be the one that has the minimum RD cost. Otherwise, {*P*16×16, *P*16×8, *P*8×16} will be added as candidate modes (Mode_q in Fig. 3). The mode that has the minimum RD cost will be the final mode (mode_r in Fig. 3) for the current MB.

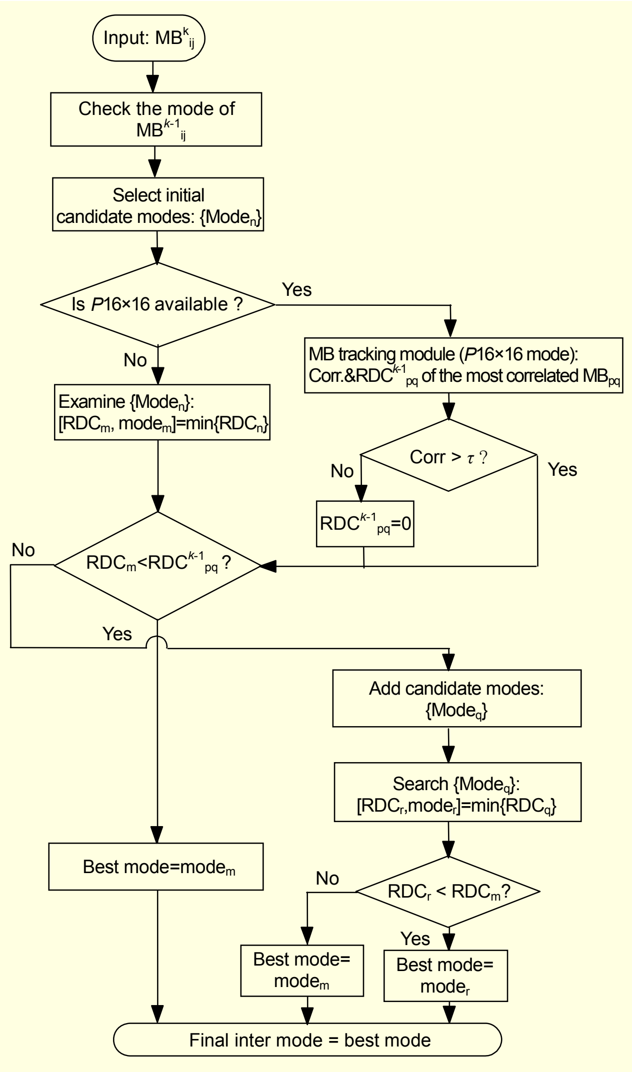


Fig. 3. Overall scheme of the proposed inter mode decision algorithm.

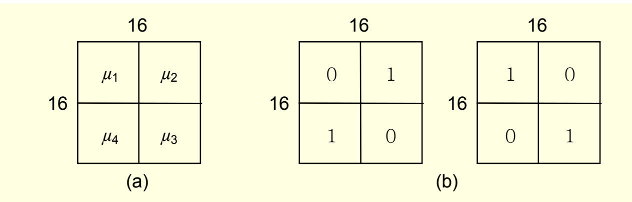


Fig. 4. Binary pattern for the refinement stage of the *P*16×16 mode: (a) 8×8 blocks and (b) binary patterns for searching all modes.

Step 2. If the colocated MB is $P16\times16$ mode, set initial 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. Otherwise, we compute a binary pattern for the current MB using spatial intensity as in Fig. 4(a).

In general, the selected mode of the given MB depends upon both image texture and motion texture (degree of motion). With more complex image and motion texture, there is a higher possibility for assignment as $P8\times8$ sub-partitions. Let μ_i be the average intensity of each 8×8 block, and let μ_T be the average value of intensity for the current MB (16×16) as

$$\mu_T = \frac{\sum_{x=0}^{15}\sum_{y=0}^{15} I(x,y)}{16^2}, \quad (3)$$

$$\mu_i = \frac{\sum_{x=0}^{7}\sum_{y=0}^{7} I(x,y)}{8^2}, \quad (4)$$

where $I(x,y)$ denotes a pixel value at the given position, and i denotes the index of an 8×8 block. If the average intensity of any 8×8 block is larger than μ_T, we set 1; otherwise, 0 will be set. After computation of a binary pattern for a refinement stage, if a texture occurs as described in Fig. 4(b), the remaining modes {$P16\times8$, $P8\times16$, $P8\times8$ sub-types} are added as candidate modes because this MB can be considered to have a complex texture. 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 $P16\times8$ mode, set initial 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. 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 $P8\times16$ mode, set initial 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. 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 $P8\times8$ mode or intra mode, set all eight possible modes as mentioned above. The mode that has the minimum cost will be the final mode for the current MB.

As shown in Fig. 3, if the correlation of the most correlated MB is less than τ, then our algorithm directly checks on the additional candidate modes. In this case, the final inter mode is determined after examining both of the initial candidate modes and additional modes. In this study, the proposed fast inter mode decision algorithm can yield good performance because of an adaptive RD thresholding scheme using the RD cost of the most correlated MB which is tracked in the previous frame for the current MB except for SKIP mode. Also, the technique for selecting partial candidate modes can improve the speed of the motion estimation procedure.

The proposed algorithm needs at least 796 (or 198) bytes memory resource for CIF (or QCIF) size sequences because it must save both neighbor RD costs and the modes. For the MB tracking scheme, there is no overhead because we can simply employ a motion vector estimation of the $P16\times16$ mode. After finishing computing the motion vector of this block mode, we must compute the correlation ratio with the obtained motion vector except for the SKIP mode in our algorithm. This is an additional overhead in our algorithm; however, this can be almost negligible when compared with other computations.

III. Results and Discussion

1. Experimental Environment

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 and IPPPPI in the main profile using CAVLC with a search range of MV = ±16; and the number of reference frames = 1. The Hadamard transform option was turned on.

The full mode search of the JM 11.0 reference software [19] of JVT was used for evaluation of the encoding performance and was compared with the proposed algorithm and other fast mode decision algorithms. We defined four measures for evaluating the encoding performance, including average $\Delta PSNR$, average $\Delta Bits$, mode number saving factor (ΔS), and an encoding-time saving factor, ΔT. The average $\Delta PSNR$ is the difference (in dB) between the average PSNR of a fast decision algorithm and the corresponding value of the full mode search method. As performance improves, this criterion becomes smaller. The average $\Delta Bits$ is the difference of bits as a percentage between the compared methods. Also, ΔS is defined as

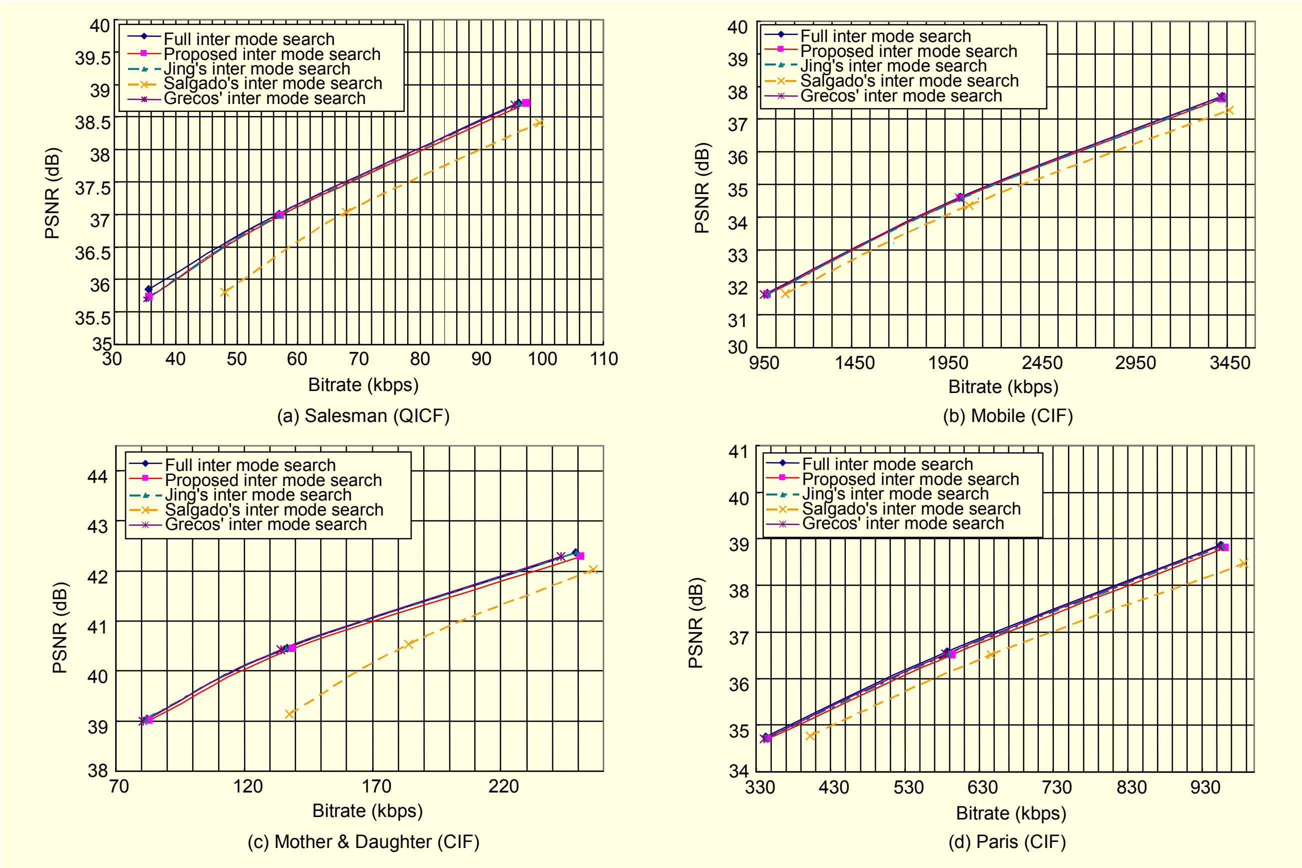


Fig. 5. RD performance for IPPP sequences: (a) $\Delta PSNR$ =-0.026 dB, $\Delta Bits$ = -0.037%, ΔS = 70.30%, ΔT = 64.91%; (b) $\Delta PSNR$ = -0.054 dB, $\Delta Bits$ = 0.016%, ΔS = 51.61%, ΔT = 39.54%; (c) $\Delta PSNR$ = -0.075 dB, $\Delta Bits$ =0.96%, ΔS = 62.49%, ΔT = 61.12%; and (d) $\Delta PSNR$ = -0.060 dB, $\Delta Bits$ = 0.84%, ΔS = 65.58%, ΔT = 55.25% with the proposed algorithm.

$$\Delta S = \frac{\#\text{ of mode search}_{ref} - \#\text{ of mode search}_{fast}}{\#\text{ of mode search}_{ref}}, \quad (5)$$

where the full mode search (FMS) is a reference and *fast* denotes the fast mode decision algorithm used. If this value becomes larger, performance speed is increased. Finally, the encoding-time saving factor ΔT is defined for complexity comparison as

$$\Delta T = \frac{T_{ref} - T_{proposed}}{T_{ref}} \times 100\ (\%), \quad (6)$$

under the condition that the FMS is also a reference. As this value increases, the performance speed is increased. Also, it must be noted that positive values for the $\Delta PSNR$ and $\Delta Bits$ indicate increments, and negative values indicate decrements.

We used three methods for an objective comparison of the encoding performance. These are Jing [13], Salgado [17], and Grecos' [14], [15] methods, which are well-known as fast and efficient algorithms. We used a fixed value of τ = 0.9 for all QPs and sequences.

2. Results for IPPP Sequences

Figure 5 illustrates the RD curves [18] for several sequences. From these results, we can see that the proposed algorithm has an RDO performance similar to the JM original encoder with the full inter mode search. Jing's method also gives similar performance to the full inter mode search but very low speed-up factor in terms of the encoding time and reduced number of modes searched. For Grecos' method, the RDO performance is similar to the full mode search with a speed-up ratio of 32.95%. This is better performance than Jing's method for average encoding time. Salgado's algorithm incurs quite large loss of PSNR at the same bitrate although it is very fast. The average PSNR loss of 0.5 dB is shown for the Salesman sequence in this result. Also, there is a degradation of 0.25 dB compared with the full mode search for the Mobile sequence. However, the proposed algorithm can provide a speed-up factor of up to 70% with negligible loss of the average PSNR and increment of bits.

Table 2 shows the results of all algorithms for the IPPP sequence type. The proposed algorithm achieves a better

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

Contents		*QP*=24				*QP*=28				*QP*=32				Average values			
		Δ *PSNR*	Δ *Bits*	Δ *S*	Δ *T*	Δ *PSNR*	Δ *Bits*	Δ*S*	Δ *T*	Δ *PSNR*	Δ *Bits*	Δ *S*	Δ *T*	Δ *PSNR*	Δ *Bits*	Δ *S*	Δ *T*
Stefan (QCIF)	Jing's	-0.029	0.902	21.34	4.42	-0.037	0.865	24.22	7.67	-0.035	0.898	26.95	9.27	-0.033	0.888	24.17	7.12
	Salgado's	-0.496	0.040	55.43	49.09	-0.257	3.425	47.83	42.24	-0.028	8.316	46.43	39.36	-0.260	3.927	49.89	43.56
	Grecos's	-0.013	0.125	16.51	3.51	-0.035	-0.052	15.14	10.49	-0.041	-0.909	14.76	18.55	-0.029	-0.278	15.47	10.85
	Proposed	-0.059	0.258	57.81	44.82	-0.058	0.330	57.45	43.84	-0.065	0.988	56.70	43.95	-0.060	0.525	57.32	43.53
Salesman (QCIF)	Jing's	-0.017	-0.148	24.74	15.72	-0.005	-0.176	32.49	13.38	-0.061	-0.397	38.13	10.76	-0.016	-0.112	31.78	10.07
	Salgado's	-0.148	2.404	76.89	76.81	0.024	19.05	68.11	59.38	-0.097	24.910	68.72	55.73	-0.009	18.79	71.27	63.97
	Grecos's	-0.036	-0.763	38.43	48.59	-0.038	-1.286	52.82	53.20	-0.025	-0.709	51.45	55.70	-0.033	-1.009	47.56	52.50
	Proposed	-0.049	0.546	71.58	70.59	-0.025	0.645	70.22	61.41	-0.005	-0.709	69.11	62.74	-0.026	0.633	70.30	64.91
Mobile (CIF)	Jing's	-0.041	0.219	17.02	8.80	-0.048	0.196	21.11	7.42	-0.044	0.288	26.63	7.17	-0.044	0.234	21.58	6.22
	Salgado's	-0.426	0.935	51.38	46.08	-0.279	2.121	44.37	39.54	-0.044	9.431	42.61	35.57	-0.249	4.162	46.12	40.64
	Grecos's	-0.013	-0.178	12.16	1.92	-0.030	-0.455	11.98	4.46	-0.066	-1.740	15.26	10.50	-0.036	-0.794	13.13	4.96
	Proposed	-0.056	0.100	55.40	37.42	-0.052	-0.015	51.43	39.58	-0.054	-0.035	48.01	41.63	-0.054	0.001	51.61	39.54
Paris (CIF)	Jing's	-0.043	0.399	35.00	8.80	-0.045	0.264	35.95	9.24	-0.037	0.399	37.15	10.22	-0.041	0.354	36.03	9.42
	Salgado's	-0.416	3.310	68.65	65.55	-0.068	9.96	61.12	52.12	0.017	17.610	60.90	49.11	-0.155	10.295	63.55	55.59
	Grecos's	-0.037	-0.248	19.41	33.37	-0.041	-0.416	21.41	38.84	-0.059	-0.578	27.85	43.02	-0.045	-0.414	22.89	38.41
	Proposed	-0.065	0.729	65.88	54.67	-0.061	0.997	65.74	54.59	-0.056	0.794	65.12	56.47	-0.060	0.840	65.58	55.24
Foreman (CIF)	Jing's	-0.025	0.409	24.81	809.	-0.021	0.335	27.80	11.36	-0.033	0.694	31.13	10.93	-0.026	0.479	27.91	10.39
	Salgado's	-0.415	9.158	55.56	48.77	-0.141	13.264	52.84	45.10	-0.011	17.794	54.53	45.65	-0.173	21.073	54.31	46.51
	Grecos's	-0.056	-1.160	12.82	12.90	-0.062	-1.137	13.32	21.35	-0.072	-1.065	21.89	28.84	-0.063	-1.121	16.01	20.80
	Proposed	-0.061	1.037	49.56	43.26	-0.072	1.486	49.84	46.06	-0.091	1.574	53.01	50.55	-0.076	1.366	50.47	46.63
Mother& Daughter (CIF)	Jing's	-0.007	0.199	23.42	14.18	-0.001	-0.249	29.42	18.50	-0.003	-0.350	35.68	30.47	-0.003	-0.133	29.50	10.72
	Salgado's	-0.343	2.874	75.20	71.82	0.070	15.605	68.19	59.97	0.099	20.205	68.65	58.83	-0.058	34.895	70.68	63.54
	Grecos's	-0.081	-2.221	32.07	37.50	-0.034	-1.935	40.77	43.59	-0.039	-1.883	48.07	49.71	-0.051	-2.013	40.30	43.60
	Proposed	-0.076	0.849	61.05	60.12	-0.017	1.338	62.64	60.28	-0.030	0.917	65.34	61.71	-0.041	1.035	63.01	60.70
Hall-Monitor (CIF)	Jing's	-0.007	-0.064	18.07	6.43	-0.008	-0.563	28.50	9.44	-0.001	-0.792	37.39	11.83	-0.001	-0.473	27.98	9.24
	Salgado's	-0.527	-8.726	68.99	65.25	-0.036	10.44	69.24	63.09	0.139	30.10	71.09	68.65	-0.141	15.606	69.77	62.84
	Grecos's	-0.042	-2.020	13.48	17.49	-0.062	-3.883	35.19	44.32	-0.013	-5.410	42.34	57.06	-0.039	-3.771	30.33	39.62
	Proposed	-0.096	-1.898	59.38	54.60	-0.090	-2.067	67.01	62.30	-0.007	-1.224	70.32	64.50	-0.067	-1.729	65.57	60.47
Akiyo (CIF)	Jing's	-0.003	0.041	30.13	9.27	-0.001	-0.012	31.29	8.72	-0.008	-0.410	33.12	11.24	-0.004	-0.127	31.51	9.74
	Salgado's	-0.139	4.473	79.38	77.08	0.101	25.56	72.70	61.73	0.040	30.80	71.14	56.46	-0.039	-3.771	30.33	39.62
	Grecos's	-0.054	-0.091	42.14	49.21	-0.047	-3.092	48.17	52.35	-0.038	-2.863	50.08	57.11	-0.046	-2.291	46.79	54.89
	Proposed	-0.035	0.244	69.15	63.53	-0.020	-1.396	69.75	63.76	-0.025	-0.550	70.73	66.65	-0.026	-0.567	69.87	64.64
Total average values	Jing's													-0.015	0.138	28.80	8.33
	Salgado's													-0.134	13.774	62.49	55.19
	Grecos's													-0.031	-1.370	29.06	32.95
	Proposed													-0.051	0.262	61.71	55.45

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 by the proposed algorithm can be improved up to 70.59% with an increase of 0.546% in the bits and PSNR loss of 0.049.

Grecos' method has a poor speed-up factor for the Stefan and Mobile sequences with large motion. However, the performance of Grecos' method is better (about 40%) for stationary sequences and for sequences with little motion, such as Paris, Hall-Monitor, Mother and Daughter, and Akiyo. 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 method achieves good PSNR and bit performance with a lower speed-up factor than the other three methods. From this result, we can see that the proposed scheme can achieve an improvement of up to 71.59% with the average loss

Table 3. Performance of the proposed algorithm on the JM 11.0 reference encoder for IPPPPI sequences.

Contents	*QP*=24			*QP*=28			*QP*=32			Average values		
	Δ *PSNR*	Δ *Bits*	Δ *T*	Δ *PSNR*	Δ *Bits*	Δ *T*	Δ *PSNR*	Δ *Bits*	Δ *T*	Δ *PSNR*	Δ *Bits*	Δ *T*
Salesman (QCIF)	-0.037	-0.551	42.18	-0.011	0.004	41.81	-0.015	-0.027	42.40	-0.021	-0.219	41.92
Mobile (CIF)	-0.034	0.045	26.76	-0.028	-0.031	27.83	-0.040	-0.039	31.11	-0.034	-0.008	28.56
Foreman (CIF)	-0.033	0.243	30.06	-0.026	0.414	34.19	-0.030	0.419	36.522	-0.029	0.358	33.70
Mother & Daughter (CIF)	-0.033	-0.397	42.14	-0.013	0.101	42.43	-0.013	0.266	44.36	-0.019	-0.004	42.97
News (CIF)	-0.036	0.120	45.61	-0.021	0.273	46.15	-0.030	0.109	46.29	-0.028	0.167	46.01
Total average values										**-0.026**	**0.058**	**38.63**

of 0.052 dB and bit increment of 0.708%, in the total encoding time.

In terms of the mode number saving factor (Δ*S*), the proposed algorithm provided good performance of between 65% and 70% for stationary sequences such as Paris, Akiyo, Salesman, and Hall-Monitor. In Mobile and Stefan, a mode number saving factor of between 48% and 58% was obtained with the proposed algorithm. Salgado's method gave better performance of 5% to 6% from the viewpoint of the mode number saving factor (Δ*S*) than others but it incurs serious quality loss. Jing's method obtained a mode number saving factor (Δ*S*) of 26% to 37%, but it has a very slow encoding time. Compared with Salgado's method, our algorithm is superior in all types of performance measured for most sequences.

Considering the average performance of each algorithm with all tested sequences, our proposed algorithm has a speed-up factor similar to that of Salgado's method. However, Salgado's method causes a large degradation in image quality, with an average loss of 0.132 dB and a bit increment of 13.77%. Jing's method has a slow average encoding time with a negligible bit increment of 0.138% and a loss of 0.015 dB. The speed-up factor of 32.95% for Grecos's algorithm was achieved with a small bit saving of 1.37% and a small quality loss of 0.031 dB. Compared with our method, this is of benefit from the viewpoint of the bit increment. However, our algorithm is faster by a factor of approximately 22% with a similar PSNR loss.

3. Results for IPPPPI Sequences

Table 3 shows the performance of our proposed algorithm for the IPPPPI sequence type. The speed-up factor is smaller than for the IPPP sequence type with higher image quality because of more inserted I-pictures. We achieved a speedup factor of more than 42% for the Mother and Daughter, Salesman, and News sequences. However, the speed-up factor was reduced to 26.76% for the Mobile sequence with large or global motion. Our algorithm yielded a speed-up factor of 38.63% for total average values with only a small quality loss of -0.026 dB and a bit increment of 0.058%. Thus, the proposed algorithm is still effective for this sequence type.

IV. Conclusion

We proposed a fast inter mode determination algorithm based on the MB tracking scheme and RD cost for P slices in H.264/AVC video coding. The proposed fast inter mode decision algorithm yielded good performance because of an adaptive RD thresholding scheme which uses the RD cost of the most correlated MB which is tracked in the previous frame for the current MB, except for the SKIP mode. Also, the technique for selecting partial candidate modes can improve the speed of the motion estimation procedure. Through comparative analysis, a speed-up factor of 40% to 70.59% was verified for IPPP sequences with a negligible bit increment and a minimal loss of image quality.

Acknowledgement

The authors wish to thank anonymous reviewers for their valuable comments and advices.

References

[1] ISO/IEC 14496-10, "Information Technology: Coding of Audio-Visual Objects--Part 10: Advanced Video Coding," Dec. 2003.

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

[3] B.G. Kim, J. H. Kim, and C. S. Cho, "A Fast Intra SKIP Detection Algorithm for H.264/AVC Video Encoding," *ETRI Journal*, vol. 28, no. 6, Dec. 2006, pp. 721-731.

[4] K.B. Suh, S.M. Park, and H.J. Cho, "An Efficient Hardware

Architecture of Intra Prediction and TQ/IQIT Module for H.264 Encoder," *ETRI Journal*, vol. 27, no. 5, Oct. 2005, pp. 511-524.

[5] H. Jia and L. Zhang, "Directional Diamond Search Pattern for Fast Block Motion Estimation," *Electronics Letters*, vol. 39, no. 22, 2003, pp. 1581 - 1583.

[6] B.-G. Kim, S.-T. Kim, S.-K. Song, and P.-S. Mah, "`Fast-Adaptive Rood Pattern Search for Block Motion Estimation," *Electronics Letters*, vol. 41, no. 16, Aug. 2005, pp. 900-902.

[7] Y.-H. Kim, J.-W. Yoo, S.-W. Lee, J. Shin, J. Paik, and H.-K. Jung, "Adaptive Mode Decision for H.264 Encoder," *Electronics Letters*, vol. 40, no. 19, Sept. 2004, pp. 1172-1173.

[8] J. Lee and B.W. Jeon, "Fast Mode Decision for H.264 with Variable Motion Block Sizes," *Proc. of the Int'l Symp. Computer and Information Sciences*, pp. 723-730.

[9] D. Zhu, Q. Dai, and R. Ding, "Fast Inter Prediction Mode Decision for H.264," *Proc. of IEEE Int'l Conf. Multimedia and Expo* (ICME), vol. 1, 2004, pp. 1123-1126.

[10] C.-H. Kuo, M. Shen, and C.C.J. Kuo, "Fast Inter-Prediction Mode Decision and Motion Search for H.264," *Proc. of the IEEE Int'l Conf. Multimedia Expo* (ICME), 2004, pp. 663-666.

[11] 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 Int'l Conf. Acoustics, Speech, and Signal Processing* (ICASSP), vol. 3, 2004, pp. 181-184.

[12] D. Wu, and F. Pan et al., "Fast Intermode Decision in H.264.AVC Video Coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 6, July 2005, pp. 953-958.

[13] X. Jing and L.-P. Chau, "Fast Approach for H.264 Inter Mode Decision," *Electronics Letters*, vol. 40, no. 17, Sept. 2004, pp. 1050-1052.

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

[15] C. Grecos and M. Yang, "Fast Mode Prediction for Baseline and Main Profiles in H.264 Video Coding Standards," *IEEE Trans. Multimedia*, vol. 8, no. 6, Dec. 2006, pp. 1125- 1134.

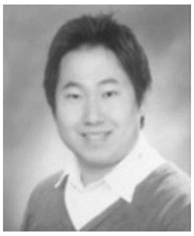
[16] B.G. Kim and S.K. Song, "`Enhanced Inter-Mode Decision Based on Contextual Prediction For P-Slice in H.264/AVC Video Coding," *ETRI Journal*, vol. 28, no. 4, Aug. 2006, pp. 425-434.

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

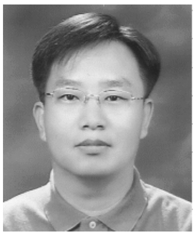
[18] G. Bjontegaard, "Calculation of Average PSNR Differences between RD-Curves," presented at the 13-th VCEG-M33 Meeting, Austin, TX, Apr. 2001.

[19] Joint Model (JM) - H.264/AVC Reference Software, http://iphome.hhi.de/suehring/tml/download/

Byung-Gyu Kim received his BS degree from Pusan National University, Rep. of Korea, in 1996 and MS degree from Korea Advanced Institute of Science and Technology (KAIST) in 1998. In 2004, he received the PhD degree in the Department of Electrical Engineering and Computer Science from Korea Advanced Institute of Science and Technology (KAIST). In March 2004, he joined in the Real-Time Multimedia Research Team at the Electronics and Telecommunications Research Institute (ETRI), Rep. of Korea, where he is currently a senior researcher. In 2007, he has been serving as an Editorial Board Member of the OEEE (Open Electrical and Electronics Engineers) Journals. He is also serving as an Editorial Board Member of the International Journal of Soft Computing and Journal of Engineering and Applied Sciences. He has published over 40 international journal and conference papers in his field. His research interests include image and video object segmentation for content-based image coding, wireless multimedia sensor networks, real-time multimedia communication, and intelligent information systems for image signal processing. He is a member of IEEE, ACM, and IEICE. He is also a member of the KMMS (Korea Multimedia Society).



Jong-Ho Kim received his BS degree from the Control and Computer Engineering Department, Korea Maritime University in 2005, and the MS degree from the Computer Software and Engineering Department, University of Science and Technology (UST), Daejeon, Rep. of Korea, Aug. 2007, Currently, he is studying at the University of Science and Technology (UST) for a PhD degree. In August 2005, he joined the Real-Time Multimedia Research Team at the Electronics and Telecommunications Research Institute (ETRI), Rep. of Korea, as a UST MS student. His research interests include video processing and video coding, especially motion estimation, and video codec mode decision methods.



Chang-Sik Cho received his BS and MS degrees from Kyungpook National University, Rep. of Korea, in 1993 and 1995. In February 1995, he joined in the Real-Time Multimedia Research Team at the Electronics and Telecommunications Research Institute (ETRI), Rep. of Korea, where he is currently a team leader and senior researcher. His research interests are real-time multimedia communication and embedded multimedia systems.