An Efficient Inter-Frame Coding with Intra Skip Decision in H.264/AVC

Myounghoon Kim, Soonhong Jung, Chang-Su Kim, and Sanghoon Sull

Abstract —*In this paper, we propose a simple but effective intra-mode skip algorithm to reduce the computational cost for inter-frame coding. It makes use of motion, temporal, and spatial homogeneity characteristics of video sequences. Specifically, the motion homogeneity is defined by using the mean deviation of motion vectors (MV) of the 4×4 blocks. The temporal and spatial homogeneity are computed by the sum of absolute difference (SAD) between an original block and its prediction block of the best inter-mode and the intra-mode with large block size, respectively. Based on the three types of homogeneity, the proposed method skip the full intra-mode search for inter-frame coding to reduce the encoding time when a region has the motion homogeneity and temporal homogeneity is stronger than the spatial one. Experimental results demonstrate that the proposed algorithm significantly increases the skip decision accuracy by up to 20% and reduces the total encoding time by about 3~10% as compared to the existing methods with negligible loss in PSNR and small increment of bit rate.*[1]

Index Terms — H.264/AVC, inter-frame coding, intra-mode skip, rate-distortion optimization (RDO), mean deviation of motion vectors, motion homogeneity, temporal homogeneity, spatial homogeneity.

I. INTRODUCTION

The recently developed H.264/AVC video coding standard [1] significantly outperforms the previous standards such as MPEG-2 [2] in terms of coding efficiency by using the advanced features such as the intra-mode coding with various spatial prediction directions and motion compensation with variable block sizes. Under the variable block size motion compensation, a macroblock (MB) can be divided into 16×16, 16×8, 8×16, 8×8, 8×4, 4×8 and 4×4 to obtain better efficiency in inter-mode coding. In addition, the SKIP mode and DIRECT mode are also supported in H.264. To take full advantage of all modes, H.264 employs a nonnormative technique called rate-distortion optimization (RDO) to determine the optimal mode having the minimum RD cost from a set of coding modes. The RD cost is calculated using the Lagrangian function as follows:

$$J_{MODE} = SAD + \lambda_{MODE} R \, , \tag{1}$$

where λ_{MODE} is the Lagrangian multiplier for mode decision and R represents the number of coding bits associated with a given mode. In this technique, by exhaustively searching all combinations of coding-modes for each MB, H.264 can achieve the best coding quality while minimizing the bit rate. However, the RDO technique drastically increases complexity and computation load due to the large number of the mode combinations. Thus, there is a need for reducing complexity on the mode decisions for H.264/AVC video encoders with a minimum quality loss.

A number of efforts have been made to reduce the encoding complexity of H.264/AVC in intra-mode decision [3]-[5] and inter-mode decision [6], [7] by choosing the most probable modes while trying to maintain a similar coding efficiency compared with the full mode decision.

A variety of fast intra-mode decision algorithms have been developed to reduce the overall complexity of H.264/AVC. Chun *et al*. [3] suggested a method for efficient intra-mode decision by using the local edge direction obtained in transform domain to filter out majority of intra prediction modes. A directional field based approach was proposed by Pan *et al*. [4] where several directions are selected by using the distribution of the edge direction histogram. Also, a method that uses a modified prediction routine and an edge direction histogram was developed in which the optimal modes for the neighboring MBs are utilized for the intra-mode decision procedure [5].

The RDO procedure for inter-modes is more complex than that for intra-modes since the former involves motion search over a window of a reference position. Wu *et al*. [6] makes use of the spatial texture homogeneity of a video object and the temporal stationarity inherent in video sequences to decide the best mode in inter-coding. Zhu *et al*. [7] 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.

Recently, various intra-mode skip algorithms for inter-frame coding were proposed in [8-10]. They always perform the inter-mode decision first and then selectively the intra-mode search. The method in [8] compares the temporal correlation (average rate of prediction residuals) and spatial correlation (sum of boundary pixel errors between pixels at a boundary of the current and its adjacent upper and left

[1] Myounghoon Kim is with the Department of Electronics and Computer Engineering, Korea University, 1, 5-ka, Anam-dong, Sungbuk-ku, Seoul 136-701, Korea. (Telephone:+82-2-3290-3805,e-mail: mhkim@mpeg.korea.ac.kr).
Soonhong Jung is with the Department of Electronics and Computer Engineering, Korea University, 1, 5-ka, Anam-dong, Sungbuk-ku, Seoul 136-701, Korea. (Telephone:+82-2-3290-3805, e-mail: shjung@mpeg.korea.ac.kr).
Chang-Su Kim is with the Department of Electronics and Computer Engineering, Korea University, 1, 5-ka, Anam-dong, Sungbuk-ku, Seoul 136-701, Korea. (Telephone: +82-2-3290-3217, e-mail: changsukim@korea.ac.kr).
Sanghoon Sull is with the Department of Electronics and Computer Engineering, Korea University, 1, 5-ka, Anam-dong, Sungbuk-ku, Seoul 136-701, Korea. (Telephone: +82-2-3290-3805, e-mail: sull@mpeg.korea.ac.kr).
Corresponding Author: Sanghoon Sull.

Contributed Paper
Original manuscript received February 11, 2010
Revised manuscript received April 8, 2010
Current version published 06 29 2010;
Electronic version published 07 06 2010.

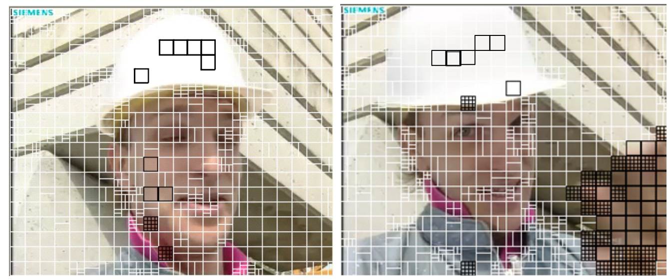
encoded blocks) to check whether one should perform the intra-mode search. The Kim's method [9] was developed under the assumption that the current MB is highly spatially correlated with neighboring MBs. It first selects the minimum RD cost among those of the neighboring MBs. Then, if the SAD of the selected inter-mode is larger than the minimum RD cost, it checks the intra-mode to obtain the better mode with smaller RD cost. However, the above two methods often becomes questionable for an object boundary where the spatial correlation is low between a current MB and its neighboring MBs. Lee *et al*. [10] uses both histogram difference to measure the similarity of two adjacent frames and rate part of the RDO. Fuzzy logic is then employed to determine whether the intra-mode search can be skipped for inter-frame coding. This algorithm works well in a video sequence with static background, while it is not clear that the intra-mode could be well skipped for moving background regions.

To reduce the load of the mode decision procedure, we propose a simple but effective method for skipping the full intra-mode search for inter-frame coding. It makes use of the motion, temporal, and spatial homogeneity where the motion homogeneity is evaluated from the mean deviation of the MVs of 4×4 blocks and temporal homogeneity is obtained by using the SAD of the selected inter-mode. Also, the spatial homogeneity can be efficiently approximated by the SAD of the intra mode with large block size (INTRA_16×16 in this paper) without performing the time-consuming full intra-search. It was experimentally found out that the mean deviation of the MVs, the SAD of the selected inter-mode, and the SAD of the intra mode with large block size well represent the degrees of motion, temporal, and spatial homogeneity. When a MB has the motion homogeneity and its temporal homogeneity is stronger than its spatial one, it is likely to be encoded using the inter-mode after the RDO computation. Thus, the full intra-mode search for this MB could be skipped with the small probability of skip decision error.

The remainder of this paper is organized as follows. Section II presents an observation and motivation. Section III describes the efficient skip algorithm of the full intra-mode search for inter-frame coding. Experiment results are shown in Section IV. Section V concludes the paper.

II. Observation and Motivation

It is observed that many natural video sequences contains a lot of regions having homogeneous MVs resulting from uniform motion of rigid objects, smooth motion of a moving background, and zero motion of a static background. An example of two frames is shown in Fig.1, in which the different block sizes of boxes overlaid on the images represent the optimal modes selected by the full mode decision. The white and black boxes represent the resulting optimal inter-mode and intra-mode, respectively. It can be seen that most of the regions having homogeneous MVs are coded using the inter-mode. On the other hand, when the frames contain a scene change, or suddenly appearing or fast moving objects such as the hand region having non-homogeneous MVs as in Fig. 1(b), the intra-mode could lead to a smaller SAD than the inter-mode and thus most MBs in this region are coded using the intra-mode.



(a) (b)

Fig. 1. Block sizes selected using the full mode decision for *Foreman* sequence. (a) 3th frame, (b) 153th frame.

TABLE I
FREQUENCY (%) OF INTER-MODES IN P-SLICE WHEN QP=28

Sequences	INTER 16×16	INTER 16×8	INTER 8×16	INTER 8×8	INTRA MODE
News	84.5	2.98	3.92	7.96	0.72
Mother &daughter	89.6	5.56	0.59	3.15	0.65
Coastguard	44.3	13.3	13.5	27.5	1.23
Foreman	59.5	10.8	12.2	14.8	2.45
Table tennis	52.4	8.50	7.51	27.1	4.35
Football	30.9	6.77	7.25	47.5	7.22
Average	**60.2**	**7.98**	**7.49**	**21.33**	**2.77**

We collected statistics of the MBs from six test sequences including low and simple motion (News, Mother and daughter), medium motion (Foreman, Coastguard), and high and complex motion (Football, Table tennis) with the same quantization parameter. Each sequence contains 100 frames encoded with IPPP structure. The frequency of different optimal modes selected using the full mode decision is listed in Table I. As shown in Table I, it can be seen that the frequency of MBs coded using the inter-mode (called inter-mode MB) is much higher than that of MBs coded using the intra-mode (called intra-mode MB) in inter-frame coding. Overall frequency of all inter-mode MBs is up to 97% on average. On the other hand, the frequency of the intra-mode MBs is less than 5%, except for the Football sequence whose ratio of intra-mode MBs is approximately 7.2%. Although there is only a small part of the intra-mode MBs in real video sequences, the full mode decision procedure in H.264 checks the intra-modes for all MBs of a given inter-frame coding, and thus a large amount of computation is wasted. Thus, if we can identify the inter-mode MBs, we can skip the time-consuming process of the full intra-mode search.

We also tested with three test sequences to check the frequency of the inter- and intra-modes MBs selected using the full mode decision at MB level and show the frequency of each mode with respect to the mean deviation of MVs in Fig. 2, where the mean deviation (also called the mean absolute deviation) is defined as the mean of the absolute deviations of a set of data about the data's mean. Note that the motion homogeneity is inversely correlated with the mean deviation of MVs and thus the degree of motion homogeneity becomes smaller as the mean deviation of MVs is larger in Fig. 2. We see that there are a lot of MBs with homogeneous MVs and the most of them are the inter-mode MBs as shown in Fig. 2(a). On the other hand, the intra-mode MBs are widely distributed

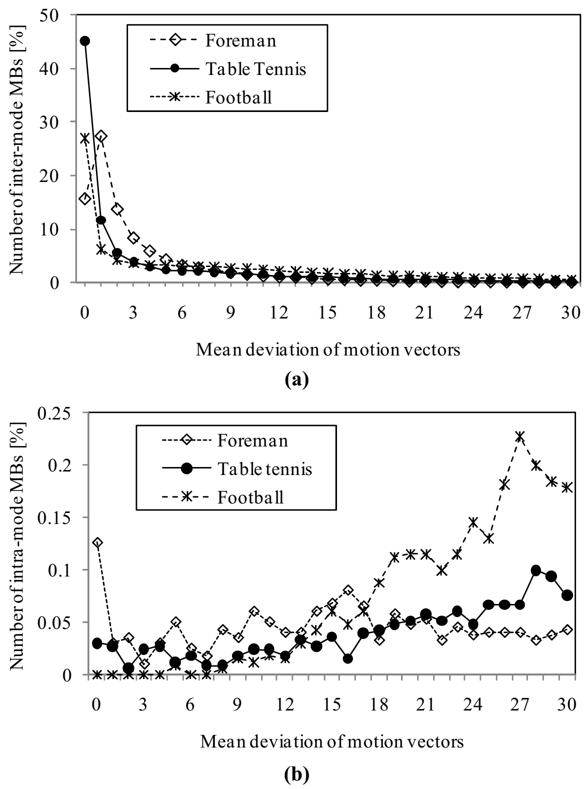


Fig.2. Distribution of the MBs according to the mean deviation of motion vectors which are coded by using (a) inter-mode and (b) intra-mode.

with respect to the degree of motion homogeneity and the chance to be intra-mode MB increases as the mean deviation of MVs is larger as shown in Fig. 2(b). Thus, for a region having non-homogeneous MVs, it seems reasonable to perform the full intra-mode search to determine the best coding mode since the coding of the prediction residual and MVs for the inter-mode could require more bits. On the other hand, for a region having homogeneous MVs, the full intra-mode search could be skipped to reduce the computational complexity of inter-frame coding.

However, note that there exist some intra-mode MBs having homogeneous MVs as shown in Fig. 2(b). Thus, for a region having homogeneous MVs, we also need to compare the spatial homogeneity with the temporal homogeneity to reduce the erroneous skip of the full intra-mode search. It was experimentally found out that the spatial homogeneity in this case is efficiently approximated by the SAD of the INTRA_16×16 without performing the full intra-search and the temporal homogeneity is obtained by the SAD of the selected inter-mode. Thus, if the SAD of the selected inter-mode is smaller than that of INTRA_16×16, the full intra-mode search is skipped. As a result, by using the motion, temporal, and spatial homogeneity, we can more accurately determine whether the full intra-mode search could be skipped for inter-frame coding.

III. Efficient Skip Algorithm of The Full Intra-Mode Search for Inter-Frame Coding

In this section, we first describe how to compute the motion, temporal, and spatial homogeneity. Then, the efficient skip algorithm of the full intra-mode search is presented for inter-frame coding.

A. Computation of motion, temporal, and spatial homogeneity

The MBs having the motion homogeneity and the temporal homogeneity stronger than the spatial homogeneity normally become the inter-mode MBs after RDO computations. The motion homogeneity is obtained by using the mean deviation of the normalized MVs (NMV) of INTER_4×4. The NMV is computed by normalizing the MV of each 4×4 block based on the temporal distance between the current frame n_c and the reference frame n_r. The NMV $\tilde{\mathbf{V}}_k = \{\tilde{v}x_k, \tilde{v}y_k\}$ for a 4×4 block B_k according to the slice type is defined as

$$\text{P-slices:} \quad \tilde{\mathbf{V}}_k = \frac{\mathbf{V}_k^0}{n_c - n_{r0}},$$
$$\text{B-slices:} \quad \tilde{\mathbf{V}}_k = \frac{1}{2}\cdot\left(\frac{\mathbf{V}_k^0}{n_c - n_{r0}} + \frac{\mathbf{V}_k^1}{n_c - n_{r1}}\right), \tag{2}$$

where n_{r0} and n_{r1} indicate the reference frame indices from *list*0 and *list*1, respectively, and V_k^0 and V_k^1 (for only B-slices) represent their corresponding MVs. The mean deviation of MVs, denoted by σ_{motion}, can be computed as

$$\sigma_{\text{motion}} = \frac{1}{16}\sum_{k=0}^{15}\left\{\left|\tilde{v}x_k - \frac{1}{16}\sum_{i=0}^{15}\tilde{v}x_i\right| + \left|\tilde{v}y_k - \frac{1}{16}\sum_{i=0}^{15}\tilde{v}y_i\right|\right\}, \tag{3}$$

where k is the 4×4 block index in a MB. The motion homogeneity is inversely correlated with the mean deviation of MVs. Therefore, small value of σ_{motion} for a given MB could imply that the MB is likely to be a region having strong motion homogeneity.

The temporal homogeneity is obtained by using the SAD of the selected inter-mode which is highly correlated with the number of bits to encode the MB. There exist several techniques for detecting the regions having the temporal homogeneity. For example, the absolute difference or the histogram difference between consecutive frames is used to compute the temporal homogeneity [6], [10]. These techniques are effective for the video sequences with stationary regions such as static backgrounds, but not for the video sequences captured by a moving camera, for example. The proposed temporal homogeneity, denoted by $\varepsilon_{\text{inter\_SAD}}$, can be calculated as follows:

$$\text{P-slices:} \quad \varepsilon_{\text{inter\_SAD}} = \sum_{k=0}^{15}\text{SAD}(vx_k^0, vy_k^0),$$
$$\text{B-slices:} \quad \varepsilon_{\text{inter\_SAD}} = \frac{1}{2}\left(\begin{array}{l}\sum_{k=0}^{15}\text{SAD}(vx_k^0, vy_k^0) + \\ \sum_{k=0}^{15}\text{SAD}(vx_k^1, vy_k^1)\end{array}\right), \tag{4}$$

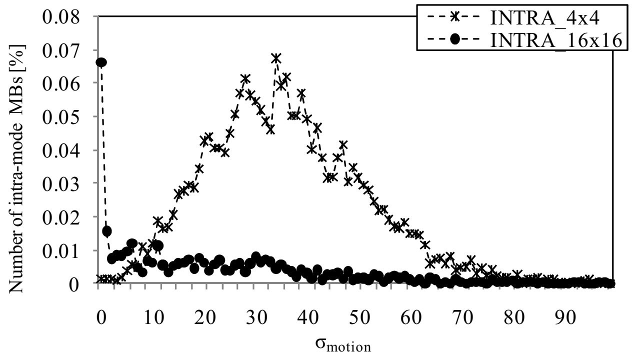


Fig. 3. Distribution of intra-mode MBs according to σ_{motion} for six test sequences.

$$\mathrm{SAD}(vx_k, vy_k) = \sum_{(x,y)\in B_k} \left| I(x,y) - P_m(x+vx_k, y+vy_k) \right|,$$

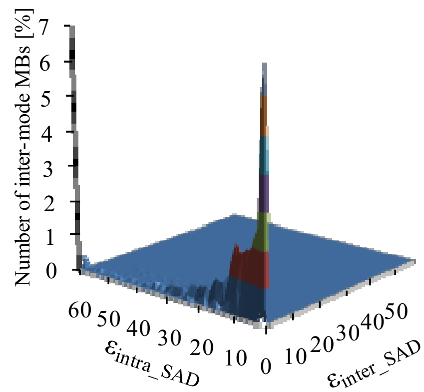
where k is the index of 4×4 block B_k in the current MB, $I(x, y)$ and $P(x+vx_k, y+vy_k)$ are pixel values of the current MB and its predictive MB, respectively, and m is the selected inter-mode.

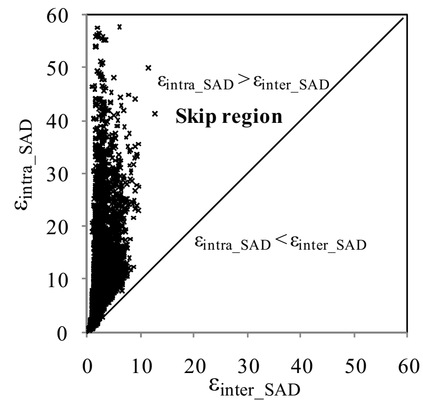
Whereas the temporal homogeneity refers to the similarity between consecutive frames in the temporal dimension, the spatial homogeneity refers to spatial similarity in a single video frame. There also exist several techniques for detecting the regions having the similar spatial property. For example, the simple statistical measurement such as variance and skewness is utilized to obtain the spatial homogeneity in [11] and texture is modeled using Gaussian Markov Random Field in [12]. The different textures are labeled separately using a hypothesis-and-test-based method on variable window sizes of the textures. This technique is effective but computationally expensive. An alternative way of determining the regions having the spatial homogeneity is to use the SAD of the INTRA_16×16.

In Fig. 3, we draw the average frequency of the INTRA_4×4 or INTRA_16×16 MBs with respect to σ_{motion} for six test sequences in Table I. Note that the frequency of the INTRA_16×16 MBs is relatively higher than that of INTRA_4×4 MBs for a region having small σ_{motion}. Thus, to estimate the spatial homogeneity for the regions having the strong motion homogeneity, it seems reasonable to use the SAD of the INTRA_16×16. To compute the spatial homogeneity, denoted by ε_{intra_SAD}, we calculate the SAD values for four modes of the INTRA_16×16 corresponding to the vertical, horizontal, DC, plane modes, and then pick the smallest one as the representative value. Thus, we have

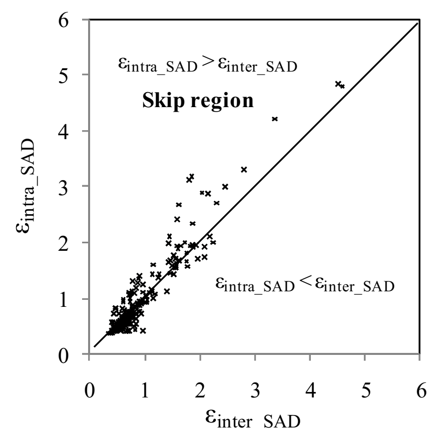
$$\begin{aligned} \varepsilon_{\mathrm{intra\_SAD}} &= \mathrm{SAD}(m_k)\,, \\ m_k &= \underset{m=\{m_0,\ldots,m_3\}}{\arg\min}\ \mathrm{SAD}(m)\ , \\ \mathrm{SAD}(m) &= \sum_{(x,y)\in MB} \left| I(x,y) - P_m(x,y) \right|, \end{aligned} \tag{5}$$

where m denotes one of four candidate modes, and $I(x, y)$ and $P(x, y)$ are pixel values of the current MB and the corresponding predictive INTRA_16×16, respectively.





(a)



(b)

Fig. 4. Distribution of (a) inter-mode and (b) intra-mode MBs having the motion homogeneity with respect to ε_{inter_SAD} and ε_{intra_SAD}.

B. Skip algorithm of the full intra-mode search

The proposed motion, temporal, and spatial homogeneity measures turned out to work well for detecting a MB having a high probability to be an inter-mode MB. Based on the motion homogeneity, the proposed method first checks whether the current MB has homogeneous MVs or not. Then we determine whether the full intra-mode search is skipped by using the temporal and spatial homogeneity. If σ_{motion} for a given MB is less than a threshold T_{motion}, the MB is classified as a region having homogeneous MVs. When the threshold T_{motion} is set to

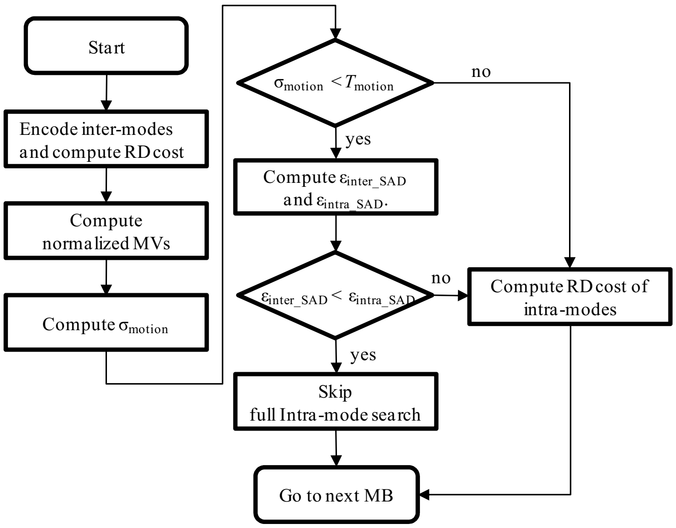


Fig. 5. Flowchart of the proposed algorithm.

5, about 55~75% out of the whole MBs are considered as having homogeneous MVs as shown in Fig. 2(a) and (b). However, there still exist about 0.11% intra-mode MBs having homogeneous MVs as shown in Fig. 2(b). For these MBs, the erroneous skip of the full intra-mode search could be made when the motion homogeneity is only considered. But this problem could be alleviated by checking the temporal and spatial homogeneity.

In Fig. 4, we plot the distribution of the inter-mode and intra-mode MBs having the motion homogeneity ($\sigma_{motion} < T_{motion}$) with respect to ε_{inter_SAD} and ε_{intra_SAD} for the six test sequences in Table I. When the motion homogeneity is only considered, the full intra-mode search is skipped for all MBs with $\sigma_{motion} < T_{motion}$ shown in Fig. 4(a) and (b), resulting in the erroneous skip of all intra-mode MBs in Fig. 4(b). Thus, we also utilize both ε_{inter_SAD} and ε_{intra_SAD}. Fig. 4(a) and (b) show that the most inter-mode MBs are included in the skip region where ε_{inter_SAD} of a MB is less than its ε_{intra_SAD} whereas the intra-mode MBs is widely distributed in the skip region and other region. The proposed method thus skips the full intra-mode search for all MBs within the skip region where ε_{inter_SAD} is less than ε_{intra_SAD}. Note that although the erroneous skips of the full intra-mode search are caused by about 50% of the intra-mode MBs within the skip region in Fig. 4(b), the overall number of successful skip ratio is not affected much since these intra-mode MBs are approximately 0.055% out of whole MBs. Thus, we see that the proposed method becomes more reliable when motion, temporal, and spatial homogeneity is considered. As a result, when the current MB has small σ_{motion} and ε_{inter_SAD} is less than ε_{intra_SAD}, our proposed method skips the full intra-mode search. Otherwise, it performs the full intra-mode search to obtain the better mode with the smaller RD cost. The flowchart for the proposed skip algorithm of the full intra-mode search is shown in Fig. 5.

IV. Experimental Results

We implemented the proposed method into the H.264/AVC reference software JM 12.2 [13] and evaluated performance with several representative test sequences. The test sequences

TABLE II
AVERAGE PERFORMANCE COMPARISON OF THE PROPOSED ALGORITHM FOR *IPPP* SEQUENCES

Contents		BDPSNR	BDBR	ΔT(%)	R_s(%)	P_e(%)
News (CIF)	**Proposed**	**0.003**	**-0.08**	**40.2**	**72.9**	**7.5**
	Kim's[9]	-0.013	0.29	37.9	60.5	24.0
	Lee's [10]	-0.009	0.21	32.7	58.8	33.6
Mother & Daughter (SIF)	**Proposed**	**0**	**0.007**	**36.6**	**72.7**	**7.9**
	Kim's[9]	-0.012	0.32	31.9	62.0	26.5
	Lee's [10]	-0.009	0.24	32.4	63.7	13.2
Coast guard (CIF)	**Proposed**	**0.001**	**-0.01**	**37.5**	**67.7**	**14.3**
	Kim's[9]	-0.008	0.2	28.4	52.5	51.4
	Lee's [10]	-0.002	0.06	24.9	46.8	25.9
Foreman (CIF)	**Proposed**	**0.001**	**-0.04**	**32.5**	**66.6**	**7.0**
	Kim's[9]	-0.013	0.359	28.3	59.2	29.7
	Lee's [10]	-0.015	0.409	28.9	60.8	20.8
Table tennis (SIF)	**Proposed**	**-0.011**	**0.288**	**38.9**	**66.6**	**5.9**
	Kim's[9]	-0.052	1.307	32.3	59.5	37.5
	Lee's [10]	-0.02	0.503	33.2	62.1	15.7
Football (SIF)	**Proposed**	**-0.005**	**0.059**	**28.9**	**52.7**	**0.3**
	Kim's[9]	-0.082	1.313	25.1	46.9	22.3
	Lee's [10]	-0.003	0.031	26.9	49.0	6.41
Drama (HD)	**Proposed**	**-0.01**	**-0.24**	**42.0**	**83.3**	**16.8**
	Kim's[9]	-0.01	0.1	31.0	73.1	20.9
	Lee's [10]	-0.02	-0.06	36.9	72.2	34.9
News (HD)	**Proposed**	**-0.03**	**0.07**	**30.6**	**61.4**	**15.1**
	Kim's[9]	-0.03	1.26	26.7	50.8	20.7
	Lee's [10]	-0.08	2.94	28.8	56.7	36.7
Talk Show (HD)	**Proposed**	**-0.01**	**-0.04**	**40.5**	**73.9**	**7.38**
	Kim's[9]	-0.02	0.72	28.6	55.6	28.6
	Lee's [10]	-0.02	0.29	42.9	78.7	26.8

include two SIF sequences (*Table tennis* and *Football*) and four CIF sequences (*News, Mother and daughter*, *Coastguard*, and *Foreman*). Each sequence consists of 100 frames with YUV 4:2:0 format with the quantization parameters set to QP 20, 24, 28, and 32. We also used three HD sequences (*Drama*, *News*, and *Talk show*) with each of which having 200 frames of 1920×1080. The test conditions are set as follows:

1) Motion vector search range is set to 16×16.
2) RD optimization is enabled.
3) The number of reference frames is set to 1.
4) CABAC is enabled.
5) Sequence types used are *IPPP* and *IBBP*.

To compare the average PSNR and bit rate of the proposed algorithm with that of the existing methods over a range of QPs (20, 24, 28, 32), we calculate Bjontegaard delta PSNR (BDPSNR) and Bjontegaard delta bit rate (BDBR) specified in [14]. Positive values of BDPSNR and BDBR mean increments whereas negative values represent decrements. We also defined an encoding time saving factor (ΔT), a ratio of the full intra-mode search skipped MBs to total number of MBs (R_s), and a skip decision error probability (P_e): The encoding time saving factor ΔT is defined as

$$\Delta T = \frac{T_r - T_p}{T_r} \times 100(\%), \tag{6}$$

TABLE III
AVERAGE PERFORMANCE COMPARISON OF THE PROPOSED ALGORITHM FOR *IBBP* SEQUENCES

Contents		BDPSNR	BDBR	ΔT(%)	R_s(%)	P_e(%)
News (CIF)	**Proposed**	**-0.003**	**0.088**	**29.5**	**70.8**	**5.0**
	Kim's[9]	-0.013	0.320	24.8	61.5	32.0
	Lee's [10]	-0.005	0.326	25.5	65.7	30.4
Mother& Daughter (SIF)	**Proposed**	**-0.006**	**0.143**	**27.4**	**70.6**	**4.9**
	Kim's[9]	-0.019	0.526	25.7	65.3	25.3
	Lee's [10]	-0.003	0.241	24.3	62.3	11.0
Coast guard (CIF)	**Proposed**	**0.004**	**-0.094**	**23.5**	**56.4**	**12.1**
	Kim's[9]	-0.023	0.617	21.3	52.3	24.3
	Lee's [10]	0.018	0.044	20.7	47.0	17.8
Foreman (CIF)	**Proposed**	**-0.008**	**0.232**	**25.5**	**61.7**	**10.2**
	Kim's[9]	-0.035	0.993	26.5	64.5	28.5
	Lee's [10]	-0.006	0.233	25.7	62.2	12.7
Table tennis (SIF)	**Proposed**	**0.004**	**-0.025**	**29.0**	**62.9**	**9.2**
	Kim's[9]	-0.055	0.45	16.3	42.6	35.4
	Lee's [10]	-0.016	0.463	25.8	58.1	22.4
Football (SIF)	**Proposed**	**-0.004**	**0.066**	**21.4**	**48.8**	**2.8**
	Kim's[9]	-0.058	0.955	22.6	51.9	14.0
	Lee's [10]	-0.013	0.166	23.3	52.1	5.3
Drama (HD)	**Proposed**	**-0.01**	**0.19**	**33.12**	**84.7**	**10.4**
	Kim's[9]	0	0.9	23.91	60.6	23.8
	Lee's [10]	-0.02	0.14	26.63	69.1	10.1
News (HD)	**Proposed**	**-0.04**	**1.02**	**27.77**	**72.8**	**21.2**
	Kim's[9]	-0.03	3.46	19.62	51.1	23.6
	Lee's [10]	-0.08	2.42	20.58	53.1	31.8
Talk Show (HD)	**Proposed**	**-0.01**	**0.05**	**31.05**	**72.5**	**9.55**
	Kim's[9]	-0.03	0.55	20.33	55.6	35.8

where T_r and T_p are average encoding times using the result under four QPs of the reference encoder with the full mode decision and our modified encoder, respectively. The skip decision error probability P_e can be defined as

$$P_e = P(skip \mid \text{intra}) \cdot P(\text{intra}), \qquad (7)$$

where P(intra) is the probability that the best mode is the intra-mode. We compared our results with Kim's method [9], Lee's method [10], and the reference software with the full mode decision for inter-frame coding.

The performance comparison between the proposed method and the existing methods is presented in Tables II and III showing BDPSNR, BDBR, ΔT, R_s, and P_e for each sequence encoded with *IPPP* and *IBBP*, respectively. It can be seen that R_s and P_e of the proposed method is up to 66% and 10% on average for *IPPP* sequences, respectively whereas R_s and P_e of Kim's method are 58% and 29% and those of Lee's method are 60% and 23% on average, respectively. From these results, we see that the proposed method achieves better performance on the skip ratio R_s and the skip decision error P_e compared with the existing methods. Also, the proposed method reduces the total encoding time by 37% (ΔT) on average, while the coding efficiency loss in terms of BDBR and BDPSNR is 0.03% and -0.001dB, respectively. Kim's and Lee's methods can achieve up to 30% and 31% for ΔT, respectively. Kim's method also increases of 0.65% in BDBR and BDPSNR loss of 0.03dB, and Lee's method increases of 0.51% in BDBR and BDPSNR loss of 0.02dB on average.

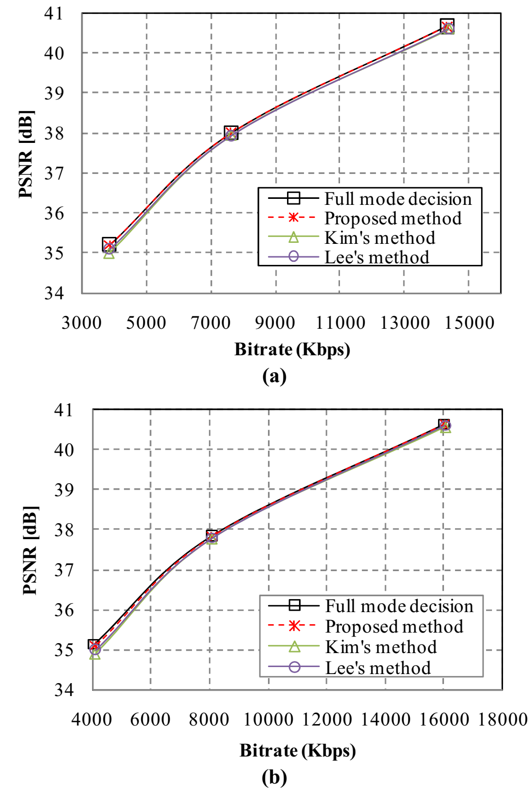


Fig. 6. RD curve for (a) *IPPP* and (b) *IBBP* sequences (*Talk show*).

Table III shows that our method reduces the total encoding time by 28% (ΔT) on average for *IBBP* sequences while the coding efficiency loss in terms of BDBR and BSPSNR is 0.18% and -0.008dB, respectively. Note that the value of ΔT is smaller compared with that for the *IPPP* sequences since the computational complexity of B-frame is higher than that of P-frame. The encoding time saving ΔT of Kim's method is up to 22% with increase of 0.94% in BDBR and BDPSNR loss of 0.03dB, and Lee's method's ΔT is up to 25% with increase of 0.46% in BDBR and BDPSNR loss of 0.018dB. From Tables II and III, we can observe that the proposed method can reduce the encoding time by additional 3~11%, increase the skip decision accuracy by up to 20%, and achieves a better bit rate saving of 0.62% with a negligible loss of image quality.

Fig. 6 shows the RD curves for *talk show* HD sequences corresponding to *IPPP* and *IBBP*. We can see that the proposed approach does not introduce any noticeable PSNR loss compared to the full mode decision, while Lee's and Kim's methods yield a small PSNR loss at the low bit rate for both *IPPP* and *IBBP* sequences. From these results, we can observe that the proposed method consistently achieves a higher encoding time saving and skip decision accuracy for each *IPPP* and *IBBP* sequences than the existing methods with the negligible loss of the bit rate and PSNR.

V. CONCLUSION

In this paper, we presented an efficient skip algorithm of the full intra-mode search for inter-frame coding based on the motion, temporal, and spatial homogeneity. The use of the proposed three homogeneity measures turned out to work well for detecting a MB having a high probability to be an inter-mode MB after RDO computations. For this MB, the full intra-mode search is skipped with the small probability of skip decision error. With the proposed algorithm, we significantly improve the skip decision accuracy by up to 20% and reduce the total encoding time by about 3~10% as compared to that the existing methods, with the negligible average PSNR loss of 0.009dB and the bit rate increments of 0.22%. Experiments with a variety of video sequences having different motion activities and resolutions demonstrate the feasibility of the proposed method for reducing the computational complexity without noticeable quality degradation.

REFERENCES

[1] ISO/IEC JTC1/SC29/WG11 (MPEG), "Coding of audio-visual objects – Part 10: Advanced Video Coding," *International Standard* 14496-10, ISO/IEC, 2004.

[2] ISO/IEC JTC1/SC29/WG11 (MPEG), "Information technology – Generic coding of moving pictures and associated audio information: Video," *International Standard 13818-2*, ISO/IEC, 2000.

[3] F. Pan, X. Lin, S. Rahardja, and K. P. Lim., "Fast mode decision algorithm for intraprediction in H.264/AVCvideo coding," *IEEE Trans. Circuits Syst. Video Technol*., vol. 15, no. 7, pp. 813–822, Jul. 2005.

[4] S. S. Chun, J.-C. Yoon, and S. Sull, "Efficient intra prediction mode decision for H.264 video," *PCM 2005*, Part I, LNCS 3767, pp. 168-178, 2005.

[5] D.-G. Sim and Y. Kim, "Context-adaptive mode decision for intra-block coding in H.264/MPEG-4 part 10," *Real-Time Imaging*, vol. 11, pp. 1–6, 2005.

[6] D. Zhu, Q. Dai, and R. Ding, "Fast inter-prediction mode decision for H.264," in Proc. *IEEE Int. Conf. Multimedia Expo (ICME)*, 2004, vol. 1, pp. 1123–1126.

[7] D. Wu, F. Pan, K. P. Lim, S. Wu, Z. G. Li, X. Lin, S. Rahardja, and C. C. Ko, "Fast intermode decision in H.264/AVC video coding," *IEEE Trans. Circuits Syst. Video Technol*., vol. 15, no. 6, pp. 953–958, Jul. 2005.

[8] I. Choi, J. Lee, and B. Jeon, "Fast coding mode selection with rate-distortion optimization for MPEG-4 Part-10," *IEEE Trans. Circuits Syst. Video Technol*., vol. 16, no. 12, pp. 1577-1561, Dec. 2006.

[9] B.-G. Kim, "Fast selective intra-mode search algorithm based on adaptive thresholding scheme for H.264/AVC encoding," *IEEE Trans. Circuits Syst. Video Technol,*. vol. 18, no. 1 pp. 127-133, Jan. 2008.

[10] P.-G. Lee, and Y.-J. Shih "Fast inter-frame coding with intra skip strategy in H.264 video coding," *IEEE Trans. on Consumer Electronics*, vol.55, no. 1 Feb. 2009. pp. 158-164.

[11] T. Uchiyama, N. Mukawa, and H. Kaneko, "Estimation of homogeneous regions for segmentation of textured images," in *Proc. IEEE ICPR*, 2000, pp. 1072–1075.

[12] X. W. Liu, D. L. Liang, and A. Srivastava, "Image segmentation using local spectral histograms," in *Proc. IEEE ICIP*, 2001, pp. 70–73.

[13] Joint Model (JM)—H.264/AVC Reference Software [Online]. Available: http://www.iphome.hhi.de/suehring/tml/download/

[14] T.Wiegand, H. Schwarz, A. Joch, F. Kossentini, G. Sullivan, "Rate constrained coder control and comparison of video coding standards," *IEEE Trans. On Circuits Syst. Video Technol*., vol. 13, pp. 688–703, July 2003.

BIOGRAPHIES

Myounghoon Kim received the B.S. and M.S. degree in computer science from the Kookmin University, Seoul, Korea, in 2003 and 2005, respectively. He is currently working toward the Ph.D. degree in the Department of Electronics and Computer Engineering of the Korea University. His research interests include video indexing for content-based retrieval, image processing, video signal processing, digital broadcasting and other issues on image and video technologies.

Soonhong Jung received the BS and MS degree in Electronic Engineering from Korea University, Seoul, Korea, in 2007 and 2009. He is currently working toward the PhD degree in electrical engineering at Korea University. His research interests are image and video signal processing, digital broadcasting, and other problems in image and video technologies.

Chang-Su Kim (S'95-M'01-SM'05) received the B.S. and M.S. degrees in control and instrumentation engineering from Seoul national University (SNU) in 1994 and 1996, respectively. In 2000, he received the Ph.D. degree in electrical engineering from SNU with a Distinguished Dissertation Award.

From 2000 to 2001, he was a Visiting Scholar with the Signal and Image Processing Institute, University of Southern California, Los Angeles, and a Consultant for InterVideo Inc., Los Angeles. From 2001 to 2003, he coordinated the 3D Data Compression Group in National Research Laboratory for 3D Visual Information Processing in SNU. From 2003 and 2005, he was an Assistant Professor in the Department of Information Engineering, Chinese University of Hong Kong. In Sept. 2005, he joined the School of Electrical Engineering, Korea University, where he is now an Associate Professor. His research topics include video and 3D graphics processing and multimedia communications. He has published more than 120 technical papers in international conferences and journals.

Sanghoon Sull (S'79-M'81) received the B.S. degree (with honors) in electronics engineering from the Seoul National University, Korea, in 1981, the M.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology in 1983, and Ph.D. degree in electrical and computer engineering from the University of Illinois, Urbana-Champaign, in 1993. In 1983-1986, he was with the Korea Broadcasting Systems, working on the development of the teletext system. In 1994-1996, he conducted research on motion analysis at the NASA Ames Research Center. In 1996-1997, he conducted research on video indexing/browsing and was involved in the development of the IBM DB2 Video Extender at the IBM Almaden Research Center. He joined the School of Electrical Engineering at the Korea University as an Assistant Professor in 1997 and is currently a Professor. His current research interests include 3D TV, video search/browsing, image processing, computer vision, and digital broadcasting.