Fast Mode Decision for H.264/AVC Based on Macroblock Motion Activity

Huanqiang Zeng, Canhui Cai, *Senior Member, IEEE*, and Kai-Kuang Ma, *Senior Member, IEEE*

***Abstract*—The intra-mode and inter-mode predictions are made available in H.264/AVC for effectively improving coding efficiency. However, exhaustively checking all the prediction modes for identifying the best one (commonly referred as *exhaustive mode decision*) greatly increases computational complexity. In this paper, a fast mode decision algorithm, called the *motion activity-based mode decision* (MAMD), is proposed to speed up the encoding process by reducing the number of modes required to be checked in a hierarchical manner, as follows. For each macroblock, the proposed MAMD algorithm always starts with checking the *rate-distortion* (RD) cost computed at the SKIP mode for a possible early termination, once the RD cost value is below a pre-determined 'low' threshold. On the other hand, if the RD cost value exceeds another 'high' threshold, then this indicates that only the intra modes are worthwhile to be checked. If the computed RD cost falls between the above-mentioned two thresholds, the remaining seven modes, which are classified into three *motion activity* classes in our work, will be examined, and only one of the three classes will be chosen for further mode checking. The above-mentioned motion activity can be quantitatively measured, which is equal to the maximum city-block length of the motion vector taken from a set of adjacent macroblocks (i.e., *region of support*, ROS). This measurement is then used to determine the most possible motion-activity class for the current macroblock. Experimental results have shown that, on average, the proposed MAMD algorithm reduces the computational complexity by 62.96%, while incurring only 0.059 dB loss in PSNR (peak signal-to-noise ratio) and 0.19% increment on the total bit rate compared with that of exhaustive mode decision, which is a default approach set in the JM reference software.**

***Index Terms*—H.264/AVC, mode decision, motion vector field adaptive search technique (MVFAST), motion activity, early termination, video coding.**

I. Introduction

THE H.264/AVC is the newest video coding standard developed by the Joint Video Team (JVT), which is formed by ITU-T VCEG and ISO/IEC MPEG standards committees [1]. Due to its high coding performance, H.264 is expected to be exploited in various video services and applications, such as IPTV, digital satellite broadcasting, to name a few. Similar to its predecessors, H.264 is also a block-based motion-compensated hybrid video coder. However, H.264 adopts multiple sophisticated prediction techniques such as multiple reference frames, variable block sizes, quarter-pixel accuracy, enhanced intra-prediction, and so on, to achieve much higher coding efficiency [2], but at the expense of heavy computational complexity [3]. Our experimental results have shown that the exhaustive mode decision process usually takes about 90% of the total computational load of the JM reference software. Therefore, how to reduce the computational complexity while maintaining almost the same coding gain and the total bit rate is the main objective on designing a *fast* mode decision algorithm for H.264 [4].

Since the inception of H.264, many fast mode decision methods have been proposed in the literatures [5]–[11]. In our view, they can be classified into two categories—one is to optimize the Lagrangian *rate distortion optimization* (RDO) function which has been adopted as the criterion for conducting the exhaustive mode decision and implemented in the JM reference software [5], [6], while the other is to predict the best mode without checking all the modes [7]–[11]. Our proposed method belongs to the latter. Some most recently introduced fast mode decision methods are highlighted as follows.

Considering the distortion as an equivalent measurement of the total energy of quantization error, the method proposed by Moon *et al.* [5] approximated the relationship between the bit rate and the distortion by exploiting a statistical model of quantization error to efficiently reduce the computational complexity. Tu *et al.* [6] estimated the bit rate and the distortion directly from the spatial-domain statistics. The proposed RD estimators achieved similar performance as that of exploiting the RDO function while its computational load is much lower. Ahmad *et al.* [7] presented a fast mode decision algorithm to determine the mode of the current *macroblock* (MB) based on the modes previously determined in the previous and current frames. Wu *et al.* [8] proposed a method based on the analysis of the edge map of the entire frame. The edge-map information is then used to decide the best edge direction and find the best *intra* mode. Furthermore, they also exploited the edge-map information to determine whether an MB is homogeneous for finding the best *inter* mode. Bu *et al.* [9] proposed a two-stage block size selection algorithm to effectively reduce the computational complexity. In the first stage, the *sum of absolute difference* (SAD) value of the current MB is calculated. If the SAD is small enough, further computation of the smaller block sizes 8×8, 8×4, 4×8, and 4×4 will be skipped; otherwise, an adaptive threshold according to the mode information of the surrounding MBs will be employed to skip some small block-sized modes. Choi *et al.* [10] proposed a fast mode decision algorithm involving two

Manuscript received August 29, 2007; revised March 20, 2008. .

Huanqiang Zeng was with the Institute of Information Science and Technology, Huaqiao University, Quanzhou, 362021, China. He is now with School of Electrical and Electronic Engineering, Nanyang Technological University, 639798, Singapore (e-mail: zeng0043@ntu.edu.sg).

Canhui Cai is with the Institute of Information Science and Technology, Huaqiao University, Quanzhou, 362021, China (e-mail: chcai@hqu.edu.cn).

Kai-Kuang Ma is with the School of Electrical and Electronic Engineering, Nanyang Technological University, 639798, Singapore (e-mail: ekkma@ntu.edu.sg).

Part of this work has been presented in [4].

Digital Object Identifier 10.1109/TCSVT.2009.2014014

stages of early terminations. In the first stage, the 16×16 block-sized mode checking is conducted for possible early termination. If such early termination condition is satisfied, then the SKIP mode is set; otherwise, the remaining inter modes are checked to find the best *inter* mode. The second early termination checking is then performed for possibly skipping all the intra modes by evaluating the average boundary errors, measured between the pixels located at the boundary of the current MB and that of its adjacent upper-left encoded MB. Kannangara *et al.* [11] introduced an adaptive model based on a Lagrangian cost function, in which the Lagrangian multiplier is estimated from a derived formula that involves some statistical measurements obtained from twelve video sequences, covering a wide range of motion activities. The computed value of the cost function will be used to determine whether the early termination condition is met, and, if so, the SKIP mode will be chosen as the best mode.

In this paper, a novel and more efficient mode decision algorithm, called the *motion activity-based mode decision* (MAMD), is proposed by evaluating the motion-activity status of spatially and temporally nearby MBs in a hierarchical manner. In our approach, the initial evaluation of the RD cost of the SKIP mode provides an early chance to possibly skip the checking process of the remaining modes for the current MB. If such early termination is not granted, then a further analysis based on the motion-vector lengths could provide a simple and yet reliable approach to avoid checking those less likely modes so as to further speed up the mode decision process. Experimental results have shown that the proposed MAMD algorithm significantly reduces the computational complexity while maintaining almost the same video coding quality and the total bit rate achieved by the exhaustive mode decision algorithm, which is the default approach made available in the JM reference software.

The rest of this paper is organized as follows. An overview of the mode decision algorithm in H.264 is described in Section II. The proposed fast mode decision algorithm, MAMD, is presented in Section III. Extensive simulation results are provided in Section IV. Finally, conclusions are drawn in Section V.

II. Overview of H.264/AVC's Mode Decision

Unlike previously developed MPEG video coding standards, H.264 provides seven block sizes (or the so-called *modes*, as shown in Fig. 1) to conduct motion estimation for motion-compensated inter-frame coding in order to achieve higher coding efficiency. The above-mentioned seven block sizes are: 16×16, 16×8, 8×16, 8×8, 8×4, 4×8, and 4×4, where the last four block sizes are jointly denoted as P8×8 in H.264, as illustrated in Fig. 1(b).

For the *inter*-frame MB coding, there are 11 candidate modes: SKIP, 16×16, 16×8, 8×16, 8×8, 8×4, 4×8, 4×4, intra_4×4, intra_8×8 and intra_16×16, where the last three modes are denoted as I4MB, I8MB and I16MB, respectively. For the *intra*-frame MB coding, only I4MB, I8MB and I16MB are applicable. It should be pointed out that although I8MB has been introduced in H.264 *Fidelity Range Extension* (FRExt) profile, most H.264 profiles used do not support I8MB in fact [12]. Therefore, so far as the intra prediction is concerned, only I4MB and I16MB are considered in this paper.

Note that all the prediction modes, except the SKIP, I4MB and I16MB modes, require motion estimation operation. If an MB is coded with the SKIP mode, all the *discrete cosine transform* (DCT) coefficients of this MB are set to zero and its *motion vectors* (MVs) can be produced from the MVs of the neighboring encoded MBs; hence, there is no quantized prediction error information and MVs required to be transmitted. Therefore, the SKIP mode is thus highly beneficial to code large areas with no motion or with a fairly slow motion (e.g., slow camera panning). If an MB is coded with one of the intra modes (i.e., either I4MB or I16MB), the spatial correlation within each video frame will be exploited and only the residual block, which is the prediction error resulted between the current block and its prediction, will be encoded. Hence, the intra modes are very suitable for encoding a highly-textured region under fast motion.

In the reference software of H.264, the Lagrangian RDO function is adopted as the criterion for conducting the exhaustive mode decision. More specifically, the aforementioned modes are individually performed based on this criterion, and the results are compared to identify the best mode that corresponds to the minimum cost. Consequently, the computational complexity resulted from such exhaustive mode decision strategy is extremely high, and thus a *fast* mode decision algorithm would effectively reduce the computational complexity.

In summary, the exhaustive mode decision algorithm implemented in the H.264 reference software [13], [14] are described as follows:

1) Check whether the current MB can be considered as an *intra* MB, which is always the case if the MB comes from an intra frame. If the checking is positive, go to step 6; otherwise, proceed to the next step.
2) Compute the cost of the SKIP mode.
3) Perform the motion estimation process and compute the costs of three large block-sized modes (namely, 16×16, 16×8, and 8×16) for the current MB, respectively.
4) Select one of the four 8×8 blocks within the current MB, perform the motion estimation process and compute the costs of four smaller block-sized modes (namely, 8×8, 8×4, 4×8, and 4×4), respectively.
5) Repeat Step 4 likewise for the other three 8×8 blocks individually.
6) Perform the intra prediction procedure and compute the costs of I4MB and I16MB, respectively.
7) Among all the modes that have been checked in the previous steps, select the one that yields the minimum cost as the best mode.

III. The Proposed Motion Activity-Based Mode Decision (MAMD) Algorithm

A. Motivations

First, from the temporal point of view, motionless or slow-motion scenes are oftentimes encountered in real-world video sequences. From the spatial point of view, large homogeneous areas are also presented in the spatial domain quite frequently. In these cases, the SKIP mode is expected to be the optimal choice.

To verify this intuition, extensive simulation experiments are conducted by using different video sequences listed in Table I for investigating the distribution of the optimal mode resulted from exploiting the exhaustive mode decision of the H.264 reference software, and the results are documented in Table II. This study clearly indicates that the SKIP mode indeed dominates among all the available modes, especially for those sequences containing motionless or slow-motion content. This observation implies that the SKIP mode checking should be done at the inception of the mode decision process for each MB so that an early termination chance is given by evaluating whether the resulted RD cost is below a threshold for effectively decreasing computational complexity.

Second, one can easily perceive that a large block-sized mode is more suitable for a homogeneous region under slow motion, while a small block-sized mode, on the contrary, is more suitable for a region containing a fast moving object. In other words, the optimal mode of the current MB to be determined should be intimately related to the motion activities of its spatially and temporally adjacent MBs. The *motion activity* can be quantitatively measured as introduced in [15], which is defined as the maximum city-block length of the MVs taken from a *region of support* (ROS), consisting of spatially- and temporally-adjacent MBs. It is an effective measurement yardstick and has been utilized in the design of a fast motion estimation scheme, called the *motion vector field adaptive search technique* (MVFAST) [15], which has been adopted in MPEG-4 standard [16]. Motivated by this intuition, extensive simulation experiments for investigating the relationship between the optimal mode resulted from the exhaustive mode decision of the H.264 reference software and the motion activity of the current MB have been conducted by using different video sequences and documented in Table III. This study strongly suggests that MBs can be grouped into different motion-activity classes in terms of the type of motion activity. Furthermore, it has been observed from the Table III that among all the prediction modes in H.264, it would be quite difficult to pre-judge, for example, whether a 16×8 block size or an 8×16 block size is a better mode without calculating their RD costs; so does the case for 4×8 versus 8×4 as another example. Therefore, it would make sense to group such hard-to-differentiated modes under the same class. The classes resulted from such grouping are established and summarized in Table IV. Consequently, only the mode(s) involved in each class is (are) required to be checked.

B. The Proposed MAMD Algorithm

Further investigation indicates that the motion activity of an MB is intimately related to its RD cost, and this relation can be exploited on the design of our mode decision algorithm, as follows. If the RD cost computed at the SKIP mode is small enough (that is, below a pre-set threshold, say, T_{low}), then the current MB is probably motionless (i.e., Class 1 in Table IV); thus, the SKIP mode should be selected as the best mode, and the mode decision process proceeds to the next MB. In fact, this is equivalent to the so-called *early termination* method often being implemented in fast motion estimation process [15]. On the other extremal end, if the RD cost computed at the SKIP mode results in a fairly large value (that is, beyond another threshold, say, T_{high}), this situation signifies that the current MB quite likely involves a highly-textured region and even with a fast motion or at a scene cut. Therefore, the two intra modes from Class 5 in Table IV are the most likely candidate modes and should be further checked by computing their RD costs to see which one yields the least cost as the best mode to be assigned. Now, the question boils down to how to determine these two thresholds, T_{low} and T_{high}, which is investigated as follows.

Since different QPs could yield different details of the picture, obviously, these two thresholds must be QP *dependent*. As specified in H.264, the range of QP values is from 0 to 51, and only the integer values from this range can be used in the JM reference software. In this paper, we focus on the most frequently-used range [23, 40].

For the delivery of robust fast mode decision performance, the threshold values of T_{low} and T_{high} play a crucial role on the entire mode decision process. For that, all the MBs from a set of commonly-used test sequences (Table I) are employed to empirically determine the reliable threshold values for T_{low} and T_{high}—with a goal of achieving 90% degree of confidence. In other words, if the RDcost(SKIP) of the current MB yielded is smaller than T_{low}, the SKIP mode is assigned to the current MB as its best mode, and the probability that the SKIP mode is indeed the best mode determined by the exhaustive mode decision is 0.9. On the contrary, if the RDcost(SKIP) of a MB is larger than T_{high}, either I4M or I16M, whichever yielding a smaller cost, will be assigned to the current MB as its best mode; in this case, the probability that either I4M or I16M is indeed the best mode is 0.9.

By plotting the threshold values of T_{low} and T_{high} versus the various QP values in Figs. 2(a) and (b) separately, it can be observed that each curve presents a smooth exponentially-shaped function. To mathematically model these functions, a Matlab function, `fminsearch`, which essentially performs the Simplex algorithm for curve fitting, is applied to approximate an exponential function with the format as: $f(x) = ae^{bx}$, based on the data points in Table V, where parameters a and b are to be determined. The resulted formulas are

$$T_{\text{low}} = 34\, e^{0.1759 \times QP} \quad (1)$$

$$T_{\text{high}} = 24215\, e^{0.0675 \times QP} \quad (2)$$

To appreciate how close of the approximation achieved in the exponential curve fitting, refer to Fig. 2.

A summary of the proposed MAMD algorithm is provided as follows and its flowchart is presented in Fig. 3.

1) Compute the cost of the SKIP mode, RDcost(SKIP).
2) If RDcost(SKIP) $< T_{\text{low}}$, then Class 1 (i.e., the SKIP mode) will be chosen as the best prediction mode, and the mode decision process proceeds to the next MB; otherwise, check the next step.
3) If RDcost(SKIP) $> T_{\text{high}}$, the RD costs of the two intra modes in Class 5 will be computed to see which one yields the least cost.
4) If $T_{\text{low}} \leq$ RDcost(SKIP) $\leq T_{\text{high}}$, the best prediction mode will be identified from one of the three Classes (i.e., Classes 2 to 4), according to the following algorithmic steps.

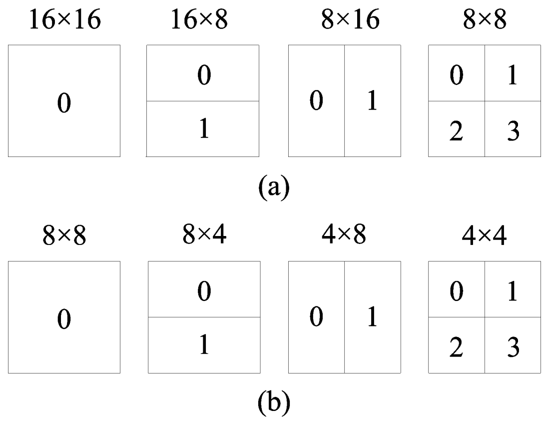

Fig. 1. The seven *modes* (or block sizes) used in H.264: (a) one 16×16 macroblock and its sub-blocks; (b) one 8×8 block and its further divided subblocks. These four block sizes are jointly denoted as "P8 × 8" in H.264.

4.1) Establish the MV set $\{\mathrm{mv}_1, \mathrm{mv}_2, \mathrm{mv}_3, \mathrm{mv}_4\}$ for the current macroblock MB_0 according to the ROS as shown in Fig. 4, where $\mathrm{mv}_i = (x_i, y_i)$ are the MVs of the corresponding MB_i, for $i = 1, 2, 3$, and 4, respectively. If an MB has multiple MVs for different partition sizes, a single MV of this MB is obtained using the bottom-up merging procedure as reported in [17].

4.2) Compute the city-block length of each MV:

$$l(\mathrm{mv}_i) = |x_i| + |y_i|, \text{for } i = 1, 2, 3, \text{and } 4. \quad (3)$$

4.3) Find the maximum city-block length, L, from the MV set:

$$L = \max\{l(\mathrm{mv}_1), l(\mathrm{mv}_2), l(\mathrm{mv}_3), l(\mathrm{mv}_4)\} \quad (4)$$

4.4) The motion activity of the current macroblock MB_0 can be determined according to

$$\text{Motion Activity} = \begin{cases} \text{Slow}, & \text{if } L \le L_1; \\ \text{Moderate}, & \text{if } L_1 < L \le L_2; \\ \text{Fast}, & \text{if } L_2 < L. \end{cases} \quad (5)$$

Based on the experimental results as shown in Table III, the thresholds $L_1 = 1$ and $L_2 = 2$ are adopted in this paper, which are in line with the recommendation made in [15]. According to Table IV, Classes 2, 3, and 4 correspond to "Slow," "Moderate," and "Fast" in (5), respectively.

4.5) Conduct experiments only to the modes that belong to the corresponding class according to Table IV.

For the boundary MBs located in the first row, the first column and the last column, different treatments are required, since the above-mentioned ROS is not completely available. To be more specific, if the current MB_0 locates at the top-left corner of a frame, all the modes in Classes 2, 3, and 4 are required to be

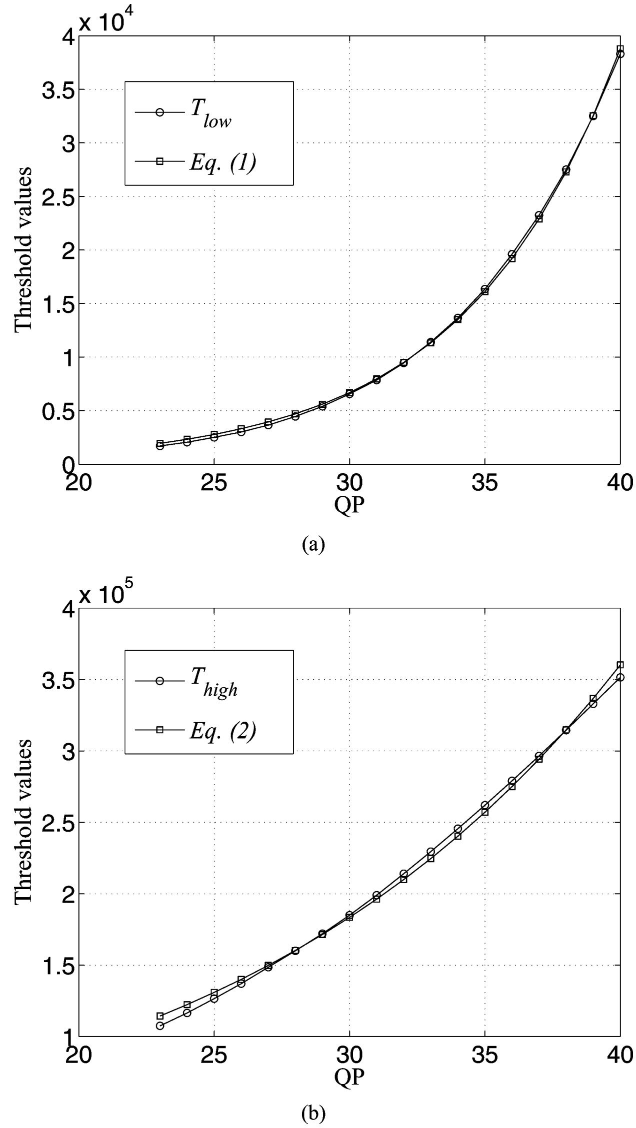

Fig. 2. The relationship between QP and two thresholds: (a) T_{low} versus QP and (b) T_{high} versus QP.

checked. For the rest of the boundary MBs locates at the top row, the left column and the right column, only one spatial MV is utilized for the evaluation of the motion activity of the current MB; that is, from the immediate-left MB, from the immediate-top MB, and from the immediate-left MB, respectively.

IV. Experimental Results and Discussion

A. Test Conditions

The proposed MAMD algorithm has been incorporated into the JVT Reference Software (Version JM10.2) [18] and experimented on multiple QCIF-format and CIF-format test sequences, covering a wide range of motion activities. All the simulation experiments are conducted on a PC with 1.6 GHz Intel Pentium processor and 512 MB memory. The test conditions are set as follows:

1) For each test sequence, 300 frames are encoded with the *group of picture* (GOP) structure, IPPP...

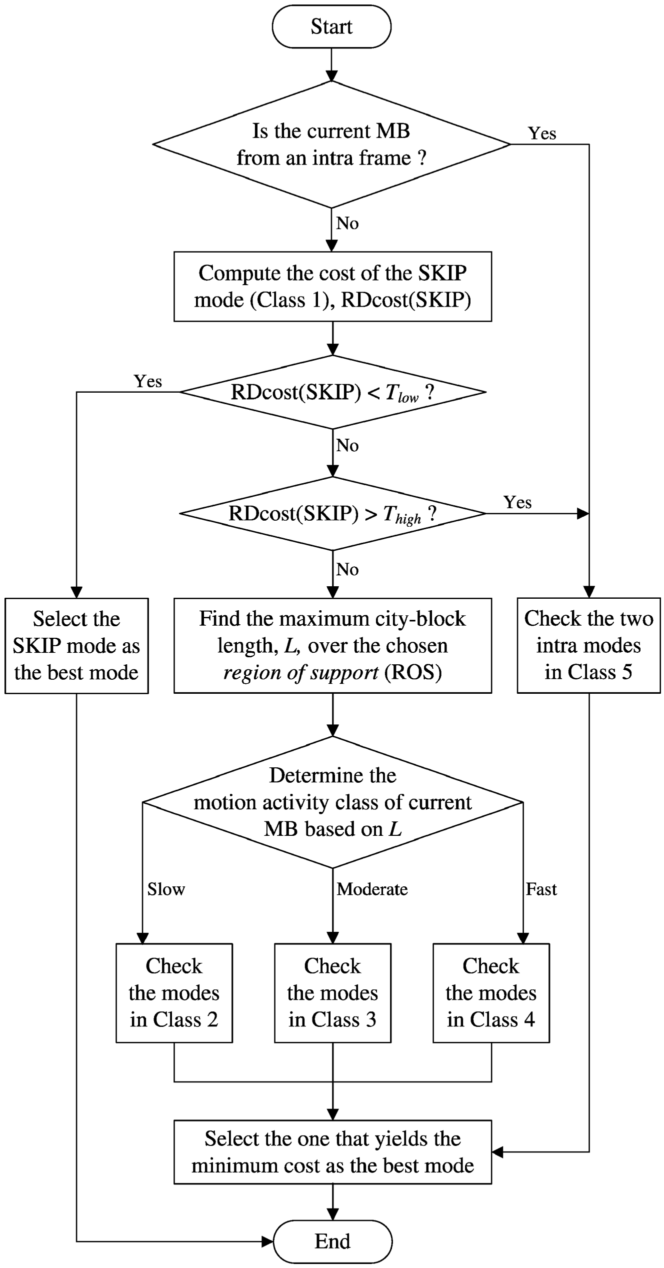

Fig. 3. The flowchart of the proposed fast mode decision method, MAMD.

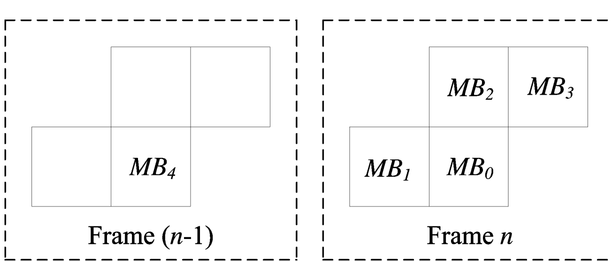

Fig. 4. The *region of support* (ROS) of the current macroblock, MB_0.

2) The quantization parameter QP is set at 24, 28, 32, 36 and 40, respectively.
3) The RDO is enabled.
4) The CABAC entropy coding is used.
5) Only one reference frame is used.

TABLE I
TEST VIDEO SEQUENCES

QCIF Sequences	CIF Sequences
Salesman	Akiyo
Table Tennis	News
Mobile Calendar	Coastguard
Bus	Flower Garden
Container	Foreman
Grandma	Silent

TABLE II
DISTRIBUTION OF THE OPTIMAL MODE RESULTED FROM EXPLOITING THE EXHAUSTIVE MODE DECISION USING THE VIDEO SEQUENCES LISTED IN TABLE I UNDER DIFFERENT QP VALUES

QP	SKIP	16×16	16×8	8×16	P8×8	I4MB	I16MB
24	45.77%	16.79%	7.00%	7.59%	21.28%	1.04%	0.53%
28	52.18%	16.33%	7.13%	7.73%	15.18%	0.88%	0.57%
32	59.60%	16.29%	6.37%	6.51%	9.81%	0.72%	0.70%
36	67.01%	16.11%	4.93%	4.38%	6.25%	0.48%	0.84%
40	75.35%	14.54%	3.37%	3.46%	1.97%	0.28%	1.03%

TABLE III
PERCENTAGES OF THE OPTIMAL BLOCK SIZES RESULTED FROM EXPLOITING THE EXHAUSTIVE MODE DECISION UNDER DIFFERENT VALUES OF CITY-BLOCK LENGTH BY USING THE VIDEO SEQUENCES LISTED IN TABLE I AT $QP = 28$

The City-block Length of the Prediction MV	16×16	16×8	8×16	P8×8
0	94.25%	2.49%	2.58%	0.68%
1	89.28%	4.67%	4.81%	1.24%
2	6.39%	42.12%	42.46%	9.03%
3	3.36%	6.13%	6.28%	84.23%
4	0.95%	3.88%	3.96%	91.21%

TABLE IV
THE MOTION-ACTIVITY CLASSES AND THEIR INVOLVED MODES

Class	Motion Activity	Involved Modes
1	Motionless	SKIP
2	Slow motion (homogeneous region)	SKIP 16×16
3	Moderate motion	16×8, 8×16
4	Fast motion	P8×8 (i.e., 8×8, 8×4, 4×8, 4×4)
5	Highly-textured region in fast motion or with scene changes	I4MB, I16MB

Note that fast motion estimation is not invoked in our simulation experiments in order to make fair comparison with the methods introduced in [9] and [10].

B. Performance Evaluation and Comparisons

Table VI documents some experimental results based on a set of QCIF-format test sequences by comparing the proposed fast mode decision algorithm, MAMD, with the exhaustive mode decision, respectively. Table VII compares the performance resulted from the proposed MAMD algorithm and the methods

TABLE V
THE EMPIRICALLY-DETERMINED THRESHOLD VALUES FOR $T_{\rm low}$ AND $T_{\rm high}$ WITH 90% DEGREE OF CONFIDENCE UNDER DIFFERENT QP VALUES. THESE VALUES ARE BASED ON RDCOST(SKIP) EXPERIMENTED ON ALL THE VIDEO SEQUENCES SHOWN IN TABLE I

QP	23	24	25	26	27	28
T_{low}	1700	2050	2500	3000	3650	4450
T_{high}	107500	116500	126500	137000	148500	160000
QP	29	30	31	32	33	34
T_{low}	5400	6550	7850	9450	11400	13650
T_{high}	172000	185000	199000	214000	229500	245500
QP	35	36	37	38	39	40
T_{low}	16350	19600	23250	27500	32500	38300
T_{high}	262000	279000	296500	314500	333000	351500

proposed in [9] and [10] over a set of CIF-format test sequences. In addition, the following performance indexes as suggested in [19] are used in both tables: △PSNR means the average PSNR changes (in dB); △B means the total bit rate changes (in percentage); △T means the average time saving (in percentage); "+" means increase; and "-" means decrease. All these measurements are conducted based on 300 frames for each test sequence.

From the results shown in both tables, it can be seen that the proposed MAMD algorithm constantly achieves—on average, 62.96% time saving with only 0.059 dB loss in PSNR and 0.19% increment in the total bit rate, comparing with the outcomes resulted by applying the exhaustive mode decision. Figs. 5(a) and (b) present the RD curves of two test sequences "Salesman" and "Foreman," showing that the proposed fast mode decision algorithm has a similar RD performance as that of the exhaustive mode decision of the H.264 reference software. Experimental results shown in Table VII further indicate that the proposed MAMD algorithm consistently outperforms a recently proposed approach by Bu *et al.* [9] in all aspects—with about 34% encoding time saving, 0.012 dB PSNR improvement, and 2.59% total bit rate reduction and the other recently proposed approach by Choi *et al.* [10]—with about 22% encoding time saving, 0.046 dB PSNR loss, and 0.40% total bit rate increment.

From the experimental results as demonstrated in Tables VI and VII, one can see that both thresholds $T_{\rm low}$ and $T_{\rm high}$ consistently work quite well at various QP values. It can be further observed that the amount of time saving (△T) is increased with QP value for almost all the test sequences except for the case of "Mobile Calendar," which deserves some insightful comments as follows.

According to (1) and (2), when the QP value is increased, both thresholds $T_{\rm low}$ and $T_{\rm high}$ are increased too. However, a larger $T_{\rm low}$ leads to more encoding time reduction (i.e., △T increased), but a reverse trend for the case of larger $T_{\rm high}$ (i.e., △T decreased). This is because when the $T_{\rm low}$ value is increased, the number of MBs that are eventually assigned with the SKIP mode as the best mode will be increased; this obviously speeds up the mode decision process. On the contrary, when the $T_{\rm high}$ value is increased, the number of MBs that are assigned with the intra mode as the best mode will be decreased; this will slow down the mode decision process, since other modes are required to be checked. With these two opposite trends, whether the final net △T is increased or decreased will depend on the contents of the input video sequence. For most natural video sequences, more time reduction is going to obtain from the increment of $T_{\rm low}$, compared with that of the extra time needed to be spent due to the increment of $T_{\rm high}$; consequently, the net △T is increased when the QP value is increased. However, since "Mobile Calendar" is a highly-textured video sequence, its △T benefited

TABLE VI
EXPERIMENTAL RESULTS OF THE PROPOSED MOTION ACTIVITY-BASED MODE DECISION (MAMD) ALGORITHM COMPARED WITH THE EXHAUSTIVE MODE DECISION APPROACH ON THE QCIF TEST SEQUENCES

QCIF Sequences	QP	△PSNR (dB)	△B (%)	△T (%)
Akiyo	24	-0.07	+0.34	-71.75
	28	-0.07	+0.12	-72.24
	32	-0.06	-0.32	-73.41
	36	-0.04	-0.37	-73.98
	40	-0.02	-0.25	-74.46
News	24	-0.05	+0.61	-65.49
	28	-0.06	+0.22	-66.64
	32	-0.05	+0.01	-66.92
	36	-0.04	+0.00	-67.83
	40	-0.05	-0.06	-67.86
Mother & Daughter	24	-0.10	+0.02	-66.26
	28	-0.07	-0.16	-66.58
	32	-0.05	-0.32	-66.99
	36	-0.05	-0.40	-67.58
	40	-0.05	-0.71	-72.11
Salesman	24	-0.04	+0.69	-59.58
	28	-0.04	+0.32	-59.69
	32	-0.03	+0.04	-59.92
	36	-0.02	-0.04	-61.53
	40	-0.02	-0.12	-66.70
Foreman	24	-0.07	+0.30	-55.12
	28	-0.09	+0.27	-56.05
	32	-0.10	+0.36	-57.37
	36	-0.11	+0.32	-59.17
	40	-0.14	-0.14	-61.02
Table Tennis	24	-0.04	+0.77	-56.26
	28	-0.04	+0.69	-58.97
	32	-0.05	+0.60	-61.42
	36	-0.06	+0.56	-63.53
	40	-0.08	+0.46	-65.24
Mobile Calendar	24	-0.11	+0.66	-62.13
	28	-0.10	+0.46	-61.28
	32	-0.09	+0.09	-59.49
	36	-0.06	+0.10	-57.83
	40	-0.04	-0.15	-56.98
Bus	24	-0.02	+0.51	-46.20
	28	-0.02	+0.69	-46.35
	32	-0.02	+0.84	-46.91
	36	-0.04	+1.31	-48.26
	40	-0.04	+1.63	-50.90
Stefan	24	-0.05	+0.23	-56.25
	28	-0.05	+0.35	-56.54
	32	-0.05	+0.38	-56.87
	36	-0.06	+0.51	-57.55
	40	-0.07	+0.22	-58.30

from the increment of $T_{\rm low}$ turns out to be smaller than the extra time required to be spent due to the increment of $T_{\rm high}$; therefore, the $\triangle$T of "Mobile Calendar" is decreased when the value of QP is increased.

C. Study of Performance Sensitivity

Recall that the values of thresholds $T_{\rm low}$ and $T_{\rm high}$ under different QPs as documented in Table V were experimentally obtained based on 12 MPEG test video sequences (Table I). Therefore, it is expected that the threshold values in Table V should be fairly robust. In other words, even with some extended perturbations on the threshold values, the resultant performance variations should be still small. To study this performance sensitivity issue, let us take the case of "Foreman" QCIF sequence at $QP = 24$ as an example. The performance sensitivity of our proposed MAMD algorithm due to the different values of $T_{\rm low}$ and $T_{\rm high}$ at this *fixed* QP under consideration is studied. For that, extensive simulation experiments have been conducted by varying the values of $T_{\rm low}$ and $T_{\rm high}$ in order to see how much changes incurred on the aspects of speed-up gain and coding efficiency under different threshold values. Some results are documented in Table VIII, from which one can see that the empirically-determined thresholds are quite robust in general, since there is only 0.14 dB loss in PSNR and 0.77% increment on the total bit rate even when the $T_{\rm low}$ is increased to 6050—which is nearly three times of the value 2050. Similar observations can be made in the case of $T_{\rm high}$. Similar sensitivity arguments as above-mentioned are also applied to other QP values.

Last but not least, it should be pointed out that since the number of MBs with their RDcost(SKIP) value lager than $T_{\rm high}$ is very small, the variation on the threshold $T_{\rm high}$ value could only result in small influence on the computational complexity, even smaller than that of the case of $T_{\rm low}$ value changes.

V. Conclusion

In this paper, a novel fast mode decision algorithm based on the evaluation of the *motion activity* from the spatially and temporally adjacent macroblocks is proposed. The novelty lies at the two key observations: (1) The RD cost of the SKIP mode can be utilized as the criterion for conducting early termination so as to greatly speed up the entire process of the mode decision, as the natural video sequences tend to have a higher probability of choosing the SKIP mode; (2) The city-block lengths of MVs over a good design of ROS can be exploited to predict the motion activity of the current MB so that the motion-activity class can be accurately identified. Experimental results have demonstrated that the proposed fast mode decision, MAMD, is consistently effective in speeding up the mode decision process by achieving 62.96% time-complexity reduction on average as compared with that of exhaustive mode decision approach based on the JVT reference software (version JM10.2), while only incurring a negligible loss of PSNR (about 0.059 dB on average) and slightly increasing the total bit rate (about 0.19% on average).

Acknowledgment

This work is partially supported by the National Natural Science Foundation of China under Grant 60772164 and the Fujian Province Natural Science Foundation under Grant A0710009.

References

[1] *Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification*, , Mar. 2003, ITU-T Rec. H.264 and ISO/IEC 14 496-10 AVC, Joint Video Team.
[2] T. Wiegand, G. J. Sullivan, G. Bjonte, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, July 2003.
[3] J. Ostermann, J. Bormans, P. List, D. Marpe, M. Narroschke, F. Pereira, T. Stockhammer, and T. Wedi, "Video coding with H.264/AVC: Tools, performance and complexity," *IEEE Circuits and Systems Magazine*, vol. 4, pp. 7–28, 2004.
[4] H. Q. Zeng, C. H. Cai, and K.-K. Ma, "A novel fast mode decision for the H.264/AVC based on local macroblock motion activity," in *The Fourth International Conference on Image and Graphics (ICIG)*, Aug. 2007, pp. 263–267.
[5] J. M. Moon, Y. H. Moon, and J. H. Kim, "A computation reduction method for RDO mode decision based on an approximation of the distortion," in *IEEE International Conference on Image Processing (ICIP)*, Oct. 2006, pp. 2481–2484.
[6] Y. K. Tu, J. F. Yang, and M. T. Sun, "Statistical rate-distortion estimation for H.264/AVC coders," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2006, pp. 365–368.
[7] A. Ahmad, N. Khan, S. Masud, and M. A. Maud, "Selection of variable block sizes in H.264," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2004, vol. 3, pp. 173–176.
[8] D. Wu, S. Wu, K. P. Lim, F. Pan, and X. Lin, "Block inter mode decision for fast encoding of H.264," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2004, vol. 3, pp. 181–184.
[9] J. Bu, S. Lou, C. Chen, and J. Zhu, "A predictive block-size mode selection for inter frame in H.264," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2006, vol. 2, pp. 917–920.
[10] I. Choi, J. Lee, and B. Jeon, "Fast coding mode selection with rate-distortion optimization for MPEG-4 part-10 AVC/H.264," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 12, pp. 1557–1561, Dec. 2006.
[11] C. S. Kannangara, I. E. G. Richardson, M. Bystrom, J. Solera, Y. Zhao, A. Maclennan, and R. Cooney, "Low complexity skip prediction for H.264 through Lagrangian cost estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 2, pp. 202–208, Feb. 2006.
[12] Q. Liu, R. M. Hu, L. Zhu, X. C. Zhang, and Z. Han, "Improved fast intra prediction algorithm of H.264/AVC," *Journal of Zhejiang University Science A.*, vol. 7, pp. 101–105, Jan. 2006.
[13] I. E. G. Richardson, H.264/MPEG-4 Part 10 White Paper: Intra Prediction Apr. 2003 [Online]. Available: http://www.vcodex.com/h264.html
[14] I. E. G. Richardson, H.264/MPEG-4 Part 10 White Paper: Inter Prediction Apr. 2003 [Online]. Available: http://www.vcodex.com/h264.html
[15] P. I. Hosur and K.-K. Ma, "Motion vector field adaptive fast motion estimation," in *Proceedings of the Second International Conference on Information, Communications and Signal Processing (ICICS)*, Singapore, Dec. 1999, pp. 234–246.
[16] *The MPEG-4 Optimization Model (OM), ISO/IEC JTC1/SC29/WG11/ N3325, Core Reference Technology Part, Noordwijkerout*, , Mar. 2000.
[17] K.-C. Hou, M.-J. Chen, and C.-T. Hsu, "Fast motion estimation by motion vector merging procedure for H.264," in *IEEE International Conference on Multimedia and Expo (ICME)*, July 2005, pp. 1444–1447.
[18] *Reference Software JM10.2*, [Online]. Available: http://iphome.hhi.de/ suehring/tml/download/old_jm/, Joint Video Team
[19] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," in *Document VCEG-M33, VCEG 13th Meeting*, Austin, Texas, USA, Apr. 2001.

Huanqiang Zeng received the B.S. degree and M.S. degree from Huaqiao University, Quanzhou, China in 2005 and 2008, respectively, both in electronic engineering.

He is currently pursuing the Ph.D. degree at the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. From May 2008 to July 2008, he was a research associate at the Temasek Laboratories, Nanyang Technological University, Singapore.

He worked on the algorithm optimization of video coding standards for his master degree. His current research interests include image processing, video-object segmentation, and multi-view video coding.

Canhui Cai (M'06-SM'08) received the B.S. degree from Xidian University, Xi'an, China in 1982, M.S. degree from Shanghai University, Shanghai, China in 1985, and Ph.D. degree from Tianjin University, Tianjin, China in 2003, all in electronic engineering.

Since 1984, he has been with the faculty of Huaqiao University, Quanzhou, China, and is currently a Professor of Electronic and Communication Engineering. He was a Visiting Professor at Delft University of Technology, Delft, the Netherlands from 1992 to 1993, and a Visiting Professor at University of California at Santa Barbara, Santa Barbara, CA from 1999 to 2000. His research areas include video communications, image and video signal processing, and computer vision. He has published over 60 papers in journals and conference proceedings.

Dr. Cai was a General Co-Chair of Intelligent Signal Processing and Communication Systems in 2007.

Kai-Kuang Ma (S'80-M'84-SM'95) received his B.E. degree (electronic engineering) from Chung Yuan Christian University, Chung-Li, Taiwan, Republic of China, M.S. degree (electrical engineering) from Duke University, Durham, NC, U.S.A., and the Ph.D. degree (electrical engineering) from North Carolina State University, Raleigh, NC, U.S.A.

He is now a full Professor at the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. From 1992 to 1995, he was a Member of Technical Staff at the Institute of Microelectronics, Singapore, working on digital video coding and the MPEG standards. From 1984 to 1992, he was with in IBM Corporation at Kingston, NY, and Research Triangle Park, NC, U.S.A., engaging on various DSP and VLSI advanced product development. His research interests are in the areas of digital signal, image and video processing. These include digital image/video coding and standards, wavelets and filter banks, image/video processing, denoising, super-resolution, error resilience and concealment, content-based image indexing and retrieval, and scene analysis and recognition. Prof. Ma has produced extensive publications in various international journals, conferences, and MPEG standardization meetings. He holds one USA patent on fast motion estimation algorithm.

He is a Distinguished Lecturer in the IEEE Circuits and Systems Society (2008–2009). He has been actively contributing various technical services for numerous international conferences; especially, serving as Technical Program Co-Chair of the IEEE International Conference on Image Processing (ICIP) 2004 and the International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS) 2007. Dr. Ma is currently serving as (Associate) Editor or Editorial Board Member of six international journals: the IEEE Transactions on Image Processing (2007–), IEEE Transactions on Communications (1997–), the IEEE Transactions on Multimedia (2002–), the International Journal of Image and Graphics (2002–), the Journal of Visual Communication and Image Representation (2005–), and Research Letters in Signal Processing (2008–). He is an elected member of three Technical Committees: Image and Multidimensional Signal Processing (IMDSP) Committee from the IEEE Signal Processing Society, Multimedia Communications Committee from the IEEE Communications Society, and Digital Signal Processing from the IEEE Circuits and Systems Society. He was serving as Singapore MPEG Chairman and Head of Delegation (1997–2001). He was Chairman of IEEE Signal Processing Singapore Chapter (2000–2002). He is a Senior Member of the IEEE and member of Eta Kappa Nu.

TABLE VII
EXPERIMENTAL RESULTS OF THREE METHODS: (A) BU *ET AL.* [9], (B) CHOI *ET AL.* [10] AND (C) OUR PROPOSED MOTION ACTIVITY-BASED MODE DECISION (MAMD) ALGORITHM. ALL THE INCREMENTAL DIFFERENCES ARE RESULTED FROM COMPARING EACH METHOD WITH RESPECT TO THE EXHAUSTIVE MODE DECISION APPROACH, INDIVIDUALLY

CIF Sequences	QP	Method	△PSNR (dB)	△B (%)	△T (%)
Akiyo	24	(A)	-0.11	+1.38	-45.88
		(B)	-0.01	-0.31	-41.55
		(C)	-0.06	-0.06	-73.03
	28	(A)	-0.06	+1.92	-42.67
		(B)	-0.02	-0.28	-46.73
		(C)	-0.07	-0.21	-74.28
	32	(A)	-0.04	+1.90	-41.56
		(B)	-0.01	-0.12	-54.27
		(C)	-0.06	-0.44	-75.35
	36	(A)	-0.05	+1.23	-39.13
		(B)	-0.01	+0.01	-54.74
		(C)	-0.08	-0.63	-76.37
	40	(A)	-0.06	+2.79	-36.46
		(B)	-0.00	-0.02	-55.09
		(C)	-0.05	-0.62	-77.09
News	24	(A)	-0.12	+2.72	-38.21
		(B)	-0.02	-0.15	-47.02
		(C)	-0.05	+0.42	-68.15
	28	(A)	-0.08	+3.11	-37.32
		(B)	-0.01	-0.13	-47.22
		(C)	-0.07	+0.26	-69.43
	32	(A)	-0.05	+3.52	-32.45
		(B)	-0.01	-0.09	-47.55
		(C)	-0.07	+0.07	-70.67
	36	(A)	-0.04	+4.34	-31.76
		(B)	-0.01	-0.11	-48.46
		(C)	-0.07	+0.06	-71.28
	40	(A)	-0.04	+4.45	-30.90
		(B)	-0.01	+0.01	-48.85
		(C)	-0.06	+0.20	-72.41
Coastguard	24	(A)	-0.07	+0.34	-19.66
		(B)	+0.00	-0.04	-20.01
		(C)	-0.02	+0.30	-50.28
	28	(A)	-0.06	+0.47	-18.98
		(B)	-0.01	-0.22	-20.39
		(C)	-0.02	+0.34	-50.62
	32	(A)	-0.05	+0.77	-19.55
		(B)	-0.02	-0.39	-26.49
		(C)	-0.03	+0.25	-52.52
	36	(A)	-0.06	+0.85	-20.47
		(B)	-0.03	-0.82	-30.25
		(C)	-0.06	+0.32	-56.54
	40	(A)	-0.08	+1.96	-22.68
		(B)	-0.02	-0.81	-43.63
		(C)	-0.06	+0.19	-61.42

CIF Sequences	QP	Method	△PSNR (dB)	△B (%)	△T (%)
Flower Garden	24	(A)	-0.08	+1.17	-20.00
		(B)	+0.00	-0.03	-38.21
		(C)	-0.02	+0.08	-56.82
	28	(A)	-0.07	+1.44	-20.82
		(B)	+0.00	-0.06	-38.36
		(C)	-0.02	+0.07	-56.95
	32	(A)	-0.06	+1.71	-21.46
		(B)	-0.01	-0.12	-38.89
		(C)	-0.02	+0.16	-57.01
	36	(A)	-0.04	+2.42	-22.57
		(B)	-0.03	-0.33	-40.57
		(C)	-0.02	+0.31	-57.06
	40	(A)	-0.06	+3.12	-24.09
		(B)	-0.02	-0.37	-46.54
		(C)	-0.02	+0.69	-57.09
Foreman	24	(A)	-0.13	+4.80	-23.45
		(B)	-0.01	-0.19	-36.53
		(C)	-0.07	+0.41	-54.57
	28	(A)	-0.11	+4.21	-25.88
		(B)	-0.01	-0.22	-37.74
		(C)	-0.09	+0.57	-56.98
	32	(A)	-0.10	+5.14	-26.68
		(B)	-0.02	-0.40	-40.25
		(C)	-0.13	+0.42	-60.31
	36	(A)	-0.09	+6.45	-27.96
		(B)	-0.02	-0.13	-41.15
		(C)	-0.16	+0.53	-63.15
	40	(A)	-0.07	+7.17	-27.25
		(B)	-0.02	-0.06	-43.28
		(C)	-0.19	+0.09	-65.81

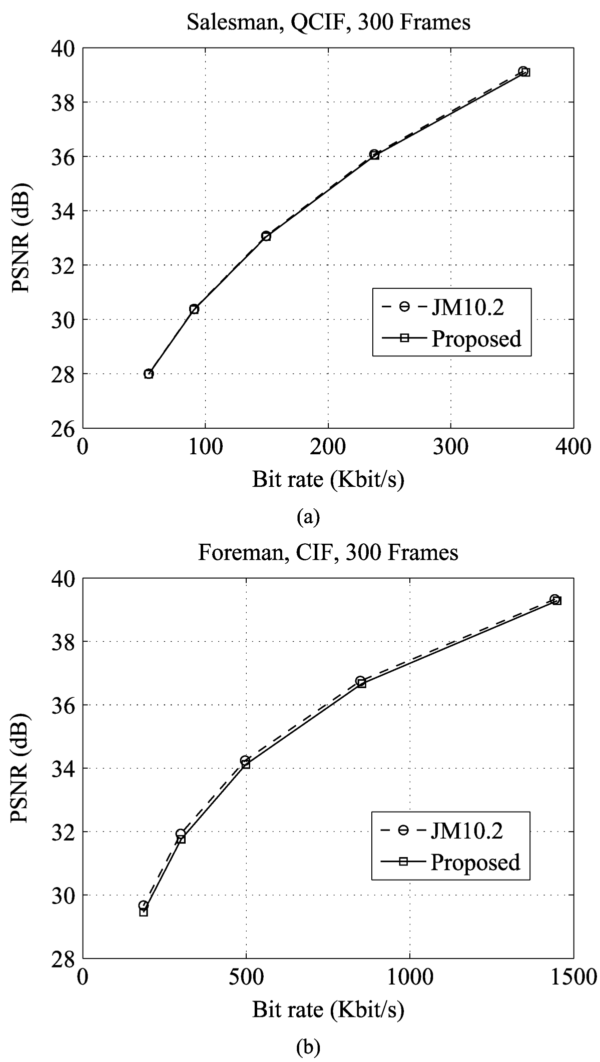

Fig. 5. The RD curves of two test sequences: (a) "Salesman;" (b) "Foreman."

TABLE VIII
EXPERIMENTAL RESULTS OF THE PROPOSED MOTION ACTIVITY-BASED MODE DECISION (MAMD) ALGORITHM COMPARED WITH THE EXHAUSTIVE MODE DECISION APPROACH ON STUDYING PERFORMANCE SENSITIVITY BY VARYING THE VALUES OF T_{low} AND T_{high} RESPECTIVELY UNDER THE FIXED $QP = 24$ AND BASED ON "FOREMAN" QCIF TEST SEQUENCE

	T_{low}	△PSNR (dB)	△B (%)	△T (%)
T_{high} = 116500	2050	-0.07	+0.30	-55.12
	3050	-0.10	+0.32	-57.52
	4050	-0.14	+0.35	-59.13
	5050	-0.18	+0.55	-60.09
	6050	-0.21	+1.07	-61.08
	T_{high}	△PSNR (dB)	△B (%)	△T (%)
T_{low} = 2050	116500	-0.07	+0.30	-55.12
	106500	-0.08	+0.41	-55.88
	96500	-0.10	+0.73	-56.62
	86500	-0.12	+1.05	-57.45
	76500	-0.14	+1.42	-58.39