Transactions Letters

Fast Inter-Mode Decision in an H.264/AVC Encoder Using Mode and Lagrangian Cost Correlation

Song-Hak Ri, Yuri Vatis, *Student Member, IEEE*, and Joern Ostermann, *Fellow, IEEE*

***Abstract*—In this paper we present a novel algorithm to speed up the inter-mode decision process for the H.264/AVC encoding. The proposed inter-mode decision scheme determines the best coding mode of a given macroblock (MB) by predicting the best mode from neighboring MBs in time and in space and by estimating its rate-distortion (RD) cost from the MB in the previous frame. The performance of the proposed algorithm is evaluated in metrics such as the encoding time, the average peak signal-to-noise ratio and the coding bit-rate for test sequences. Simulation results demonstrate that the proposed algorithm can determine the best mode using only one or two rate-distortion cost computations for about half of the MBs resulting in up to 56% total encoding time reduction with on average 2.4% of bit rate increase at the same PSNR compared to H.264/AVC JM 12.1.**

***Index Terms*—H.264/AVC encoder, mode decision, mode prediction, rate-distortion (R-D) theory, video coding.**

I. Introduction

THE COMPRESSION performance achieved by the international video coding standard H.264/AVC [1] enables new video services, such as mobile video phones and multimedia streaming over mobile networks. Compared to previous video coding standard, the performance gains of H.264/AVC come at the expense of increased computational complexity [3]. There is a need to develop low-complexity implementations of the H.264/AVC coder that offer the performance and flexibility advantages of the standard without an excessive computational cost.

An H.264/AVC video encoder typically carries out a number of encoding processes including motion estimation, mode decision, transform, quantization, entropy coding. The computational complexity of the transform, quantization and entropy coding processes of an H.264/AVC encoder is relatively low, when compared to motion estimation and mode decision [3]. With about 85% of the total complexity, motion estimation is the most computationally expensive part. Since inter-mode decision requires the estimation of motion vectors for all possible block types for each macroblock (MB), the optimization of mode decisions will reduce the complexity significantly.

Up to now, low-complexity algorithms have been proposed for a number of aspects of H.264/AVC encoding process [6]–[17]. The existing low-complexity algorithms for inter-mode decision for the H.264/AVC can be classified into three classes: non-rate-distortion optimization (RDO) based class, RD estimation-based class and RDO-based class.

The non-RDO based class looks for the optimal mode by using some features, such as texture and edge information, which are computed from the raw video data. For instance, the fast mode decision algorithm proposed in [7] introduces the so-called mean removed mean absolute difference (mrMAD). In [5], [10], the 3×3 Sobel operator has been used to get the edge map of a whole frame. The edge map is employed to determine whether a MB is homogeneous. However, the algorithm has to evaluate all of the pixels in the whole frame which leads to additional computational complexity.

The RD estimation-based class estimates the rate and distortion values just after quantization of discrete cosine transform (DCT) coefficients in order to calculate the RD cost of the current prediction mode. The RD estimation proposed in [16] and [17] assumes that distortion is proportional to the quantization error [17] and the compression rate is related to the number and the sum of magnitudes of non-zero coefficients of quantized DCT coefficients [16], [17].

The RDO-based class reduces the number of RD cost computations by using statistical dependencies of the RD costs between modes. It predicts the best mode from already checked modes and their statistical relationships. The method [8] divides all modes into 3 groups. By evaluating three modes which are selected from each group, the most probably optimal group is determined. All modes of the most probable group are evaluated to determine the best prediction mode. Thus, the number of candidate modes is greatly reduced. In [9], the most probable mode is predicted based on the observation that most modes are spatially correlated in a given frame. The algorithm [13] and [14] compute a mode decision criteria r like $r = (J_{8\times 8} - J_{16\times 16}/J_{16\times 16})$ and $r = (J_{8\times 8} - J_{16\times 16}/J_{8\times 8} + J_{16\times 16})$, respectively. With this r they classify into three regions, $Class8 (= \{P8 \times 8\})$, $Class16 (= \{SKIP, 16 \times 16, 16 \times 8, 8 \times 16\})$ and $NoClass$ which means all possible modes. If r is less than th_low, $Class8$ is selected, and if r is greater than th_high, $Class16$ is selected. Otherwise, all coding modes are checked without

Manuscript received September 03, 2007; revised December 18, 2007. First published December 09, 2008; current version published January 30, 2009. This paper was recommended by Associate Editor K.-H. Tzou.

The authors are with the Institut für Informationsverarbeitung, Leibniz Universität Hannover, D-30167 Hannover, Germany (e-mail: ostermann@tnt.uni-hannover.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2008.2009257

early decision. The thresholds th_low and th_high are defined as $th\_low = mean - (std/2)$, $th\_high = mean + (std/2)$, where $mean$ is the average of r and std is standard deviation of r. The fast mode decision algorithms from [11], [12] use contextual prediction of the most probable mode group by using spatio-temporal context information between already coded MB modes. Based on high temporal correlation, the algorithm [11] utilize the correlation of RD cost between current and previous MB.

We propose a new RDO-based mode decision algorithm to raise the optimal mode selection hit-rate by incorporating spatial mode prediction into the prediction routine. Using temporal and spatial mode prediction increases the probability to select the optimal mode. The algorithm can significantly reduce computational complexity in the H.264/AVC encoder without any significant loss of rate-distortion performance as seen with a baseline optimized encoder [4], [5] or in [8]. The new algorithm is based on the well-known RDO method [2] and exploits the high correlation between the expected rate-distortion cost and the actual cost.

The rest of this paper is organized as follows. A new inter-mode decision scheme using mode and its RD cost prediction is discussed in Section II. The experimental results and conclusions are presented in Sections III and IV, respectively.

II. New Fast Inter-Mode Decision Algorithm

In Sections II-A–C, we introduce spatial-temporal mode prediction (STMP) and RD cost prediction. We use these concepts for a fast mode selection algorithm.

A. Mode Prediction

Video coding is achieved by reducing spatial and temporal redundancies between video frames. This implies indirectly that the best prediction mode of a MB might also be related to the best mode of MBs neighboring in space and in time. It was noted in [9] that there was a spatial mode-correlation between a MB and its neighboring MBs and therefore, it is possible to spatially predict the best mode of the MB. And it was found in [11] and [12] to select the most probable mode group from the optimal mode of the MBs neighboring to current MB in space and in time.

Since we can easily suppose that a video sequence generally contains more redundancies in the time domain than in the space domain, we stipulate that the temporal mode-correlation is higher than the spatial mode-correlation. Thus, we consider spatial, temporal and spatial-temporal prediction of the best mode for a given MB.

In order to do that, we must answer these two questions: 1) How high are the spatial and the temporal mode-correlation? 2) Is it necessary to consider all possible modes for the mode prediction?

Let's mark the current MB as X, the collocated MB of X in the previous frame as X_{-1} and neighboring MBs as A (left), B (upper), C (upper left) and D (upper right). Let's mark the coding mode of X, X_{-1}, A, B, C or D as m_X, $m_{X_{-1}}$, m_A, m_B, m_C or m_D, respectively. Before answering the two questions above, we define spatial- and temporal-mode prediction as follows:

TABLE I
Occurrence Probability of Spatial, Temporal and Combined Mode Events (in %) E_S, E_T and E_C for $\mathrm{QP} = 28$, 300 Frames

Sequences	QCIF			CIF		
	P_S	P_T	P_C	P_S	P_T	P_C
container	66.0	73.1	81.8	65.7	66.2	78.1
mo.&daughter	53.1	63.1	74.5	61.2	66.2	77.1
stefan	30.4	42.5	55.6	31.6	40.5	54.3
foreman	26.4	37.8	52.0	34.1	38.8	56.0
coastguard	25.2	42.1	55.4	30.8	36.9	53.7
mobile	28.3	34.0	49.7	31.6	34.2	52.8

TABLE II
Statistics of Modewise-Temporal Mode-Correlation in Case m_X and $m_{X_{-1}}$ are the Same, $\mathrm{Unit} = \%$, $\mathrm{QP} = 28$, QCIF, 300 fr

Sequences	P_{SK}	$P_{16\times16}$	$P_{16\times8}$	$P_{8\times16}$	$P_{8\times8}$	P_{IR}
container	79.6	9.3	3.7	3.8	3.5	0.1
mo.&d.	64.3	16.8	6.4	6.7	5.7	0.1
stefan	22.3	32.7	11.1	14.6	18.3	1.0
mobile	16.7	28.1	14.9	11.5	28.0	0.8
foreman	11.6	39.1	14.2	12.1	22.6	0.4
coastguard	4.1	18.9	17.2	15.1	44.7	0.0

$$m_{XT} = m_{X_{-1}} \tag{1}$$

$$m_{XS} = \begin{cases} m_A : (m_A = m_B) \wedge (m_A = m_C) \wedge (m_A = m_D) \\ m_B : (m_B = m_C) \wedge (m_B = m_D) \\ m_C : m_C = m_D \end{cases} \tag{2}$$

where m_{XS} and m_{XT} denote spatially and temporarily predicted modes, respectively. To compare different correlations, let's define the following three events:

$$E_s : 2, 3 \text{ or } 4 \text{ values of } m_A,\ m_B,$$
$$m_c \text{ and } m_D \text{ are the same as } m_x.$$
$$E_r : m_{x-1} \text{ is the same as } m_x.$$
$$E_c : E_s \cup E_T.$$

Here, E_S, E_T and E_C denote spatial, temporal and combined mode events. Table I shows the probabilities (P_S, P_T, and P_C) of each event for some CIF video sequences.

Table I shows that the probability of E_S is lower than the probability of E_T. Obviously the probability of E_C is also greater than the probability of E_S or E_T. In the case of sequences such as *container* and *mother & daughter*, which are characterized by slow and smooth motion, the probability of a spatial mode event is similar to the temporal mode event. In the case of some sequences, such as *mobile* and *coastguard* , which are characterized by fine texture, the probability of a spatial mode event is far lower than that of the temporal mode event. The table tells us that by using combined mode prediction, the encoder can predict the best mode of a given MB more frequently than by using spatial mode-correlation.

We turn now to the second question. For all possible modes of X, let's calculate the probability of an event where the RD-optimal mode of X, m_X, is the same as $m_{X_{-1}}$ (see Table II).

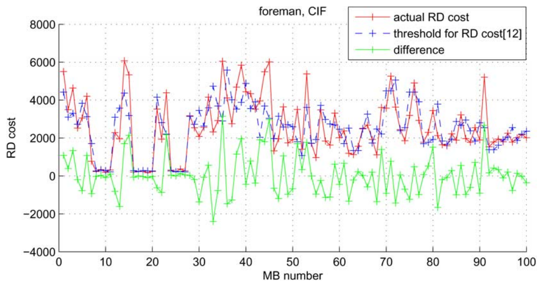

Fig. 1. Relationship between RD cost of X and average RD cost of neighboring MBs with the same mode (spatial RD cost prediction based on the threshold computed from [9]). The difference is computed by subtracting the threshold from the actual RD cost.

Let's mark the probability $P(SKIP|m_X = m_{X_{-1}})$ as P_{SK}, $P(16\times16|m_X = m_{X_{-1}})$ as $P_{16\times16}$, $P(16\times8|m_X = m_{X_{-1}})$ as $P_{16\times8}$, $P(8\times16|m_X = m_{X_{-1}})$ as $P_{8\times16}$, $P(P8\times8|m_X = m_{X_{-1}})$ as $P_{P8\times8}$, $P(Intra4\times4|m_X = m_{X_{-1}})$ as $P_{In4\times4}$ and $P(Intra16\times16|m_X = m_{X_{-1}})$ as $P_{In16\times16}$.

As seen in Table II, P_{IR} (both, 16×16 and 4×4) is very small. Therefore, we don't use the predicted modes, intra-4×4 and intra-16×16, as candidates for the mode prediction of a MB, if $m_{X_{-1}}$ is intra.

B. RD Cost Prediction

In the H.264/AVC encoder, the RD cost to decide the best inter-prediction mode is computed as follows:

$$J(s,c,MODE|\mathrm{QP}) = SSD(s,c,MODE|\mathrm{QP}) + \lambda_{MODE}\cdot R(s,c,MODE|\mathrm{QP}). \quad (3)$$

where QP is the quantization parameter, λ_{MODE} is the Lagrangian multiplier and SSD is the sum of the squared differences between the original block s_i and its reconstruction c_i. $R(s,c,MODE|\mathrm{QP})$ is the number of bits associated with the mode $MODE$ currently selected for the MB.

$$MODE = \{SKIP, 16\times16, 16\times8, 8\times16, P8\times8\}.$$

The $SKIP$ mode is a special 16×16 mode where no information for a MB is transmitted. Here, $R(s,c,SKIP|\mathrm{QP}) = 0$.

The crucial thing for applying mode prediction to fast mode decision is to make sure that the predicted mode has the smallest RD cost for a given MB. So far, there have been several ways [9], [15] to decide whether or not the predicted mode can be assumed to have the smallest RD cost.

The most common method [9] is to adopt a threshold value derived from the RD costs which are already calculated. The threshold is set to the average of the RD costs of neighboring MBs with identical modes and it is compared with the RD cost of MB X with the predicted mode to evaluate if it is the best mode or not. Another method [15] adopts the square of the quantization parameter as a threshold for the RD cost to decide whether the predicted mode is to be used.

For the sequence *foreman*, the RD cost difference between RD costs of the spatially predicted mode computed by [9] and the RD-optimal mode is shown in Fig. 1. The size of this difference does not necessarily depend on the actual RD cost. Therefore, the threshold based on neighboring MBs or QP should not be used for evaluating the quality of the predicted mode. We choose the RD cost of X_{-1} as the threshold. Fig. 2 intuitively shows an advantage of the relationship between the actually RD-optimal cost of X and the RD-optimal cost of X_{-1} when the RD-optimal modes of X and X_{-1} are the same. This relationship between the RD costs of X and X_{-1} can also be seen in the comparison of the following three correlation coefficients: correlation coefficient (ρ_S) between the RD cost predicted spatially in [9] and the RD-optimal cost, correlation coefficient $(\rho_{T'})$ between the actually RD-optimal cost of X and the RD-optimal cost of X_{-1} when the optimal mode of X is the same as one of X_{-1}, and correlation coefficient (ρ_T) between the actually RD-optimal cost of X and the RD-optimal cost of X_{-1}. Table III shows that the temporal correlations ρ_T and $\rho_{T'}$ are greater than ρ_S. From Figs. 2, 3 and Table III, it should be noted that the correlation of RD costs is high, even in the case that the RD-optimal modes of X and X_{-1} are not the same (Fig. 3), which means that the RD-optimal cost of a MB can be predicted by the RD cost of the optimal mode of the previous MB.

TABLE III
COMPARISON OF THREE CORRELATION COEFFICIENTS IN QCIF AND CIF FORMAT

Sequences	QCIF			CIF		
	ρ_S	$\rho_{T'}$	ρ_T	ρ_S	$\rho_{T'}$	ρ_T
foreman	0.722	0.949	0.922	0.683	0.952	0.939
coastguard	0.772	0.942	0.933	0.560	0.934	0.921
stefan	0.870	0.969	0.957	0.779	0.975	0.972
mother&daughter	0.814	0.979	0.964	0.789	0.987	0.976
mobile	0.485	0.974	0.970	0.358	0.964	0.965
container	0.764	0.988	0.976	0.508	0.993	0.983
average	0.738	0.967	0.954	0.613	0.968	0.959

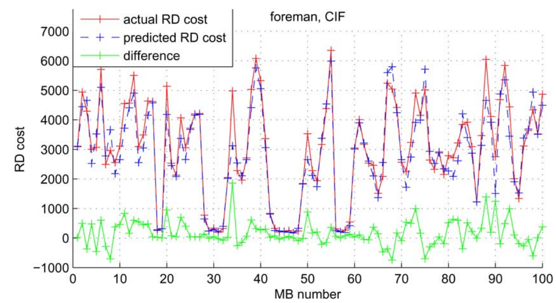

Fig. 2. Relationship between RD cost of X and RD cost of X_{-1} when the RD-optimal mode of X is the same as one of X_{-1}.

In the case that the RD-optimal mode of X_{-1} doesn't equal the RD-optimal mode of X, we can use the correlation coefficient (ρ_T) between RD costs of X and X_{-1} for making a decision of the best mode. For that, we evaluate the probability density function (pdf) of the RD cost difference of X and X_{-1} when the optimal mode of X_{-1} is $SKIP$, 16×16, ... or $P8\times8$, respectively.

Fig. 4 shows that the RD cost difference between X and X_{-1} is approximately close to zero when the RD-optimal mode of X_{-1} is the *SKIP* mode irrespective of whether or not X has the same mode as X_{-1}. This property allows the H.264/AVC encoder to skip checking other modes as soon as a mode is found with the RD cost in a certain range of the RD optimal cost ofX_{-1}.

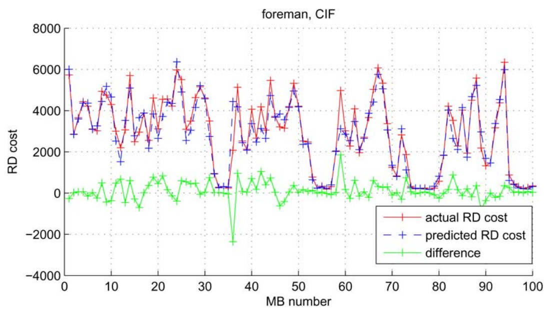

Fig. 3. Relationship between RD cost of X and RD cost of X_{-1}.

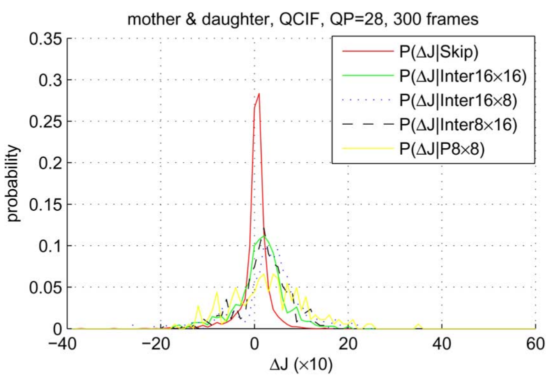

Fig. 4. Measured pdfs of RD cost differences, where $P(\Delta J|MODE)$ denotes the probability of RD cost difference (ΔJ, refer to (3)) between X (irrespective of mode) and X_{-1} given MODE.

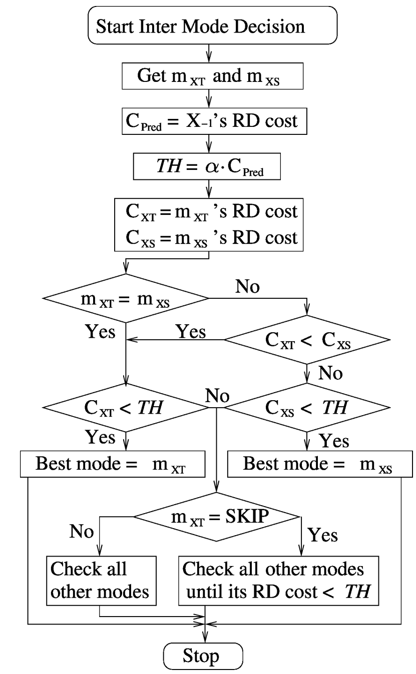

Fig. 5. Block diagram of the proposed $STMP$ algorithm.

C. New Fast Inter-Mode Decision Algorithm Stmp

It has been observed that sometimes temporal mode prediction shows better results than spatial one, and also vice versa. Therefore, we take two mode candidates, m_{XT} and m_{XS}, predicting the mode temporally and spatially, respectively, and choose the mode with the lower RD cost (see (1) and (2)). One problem which might happen in using mode prediction is error propagation, due to a mis-prediction of the best mode and further prediction from such a non-optimal mode. To prevent the propagation of mode prediction errors, an exhaustive mode decision will be carried out periodically. The proposed STMP algorithm is as Fig. 5.

In the diagram, m_{XT} and m_{XS} can be computed using (1) and (2), and α is a positive, sequence-independent constant derived from experiments. C_{pred} is the prediction RD costs of the temporarily collocated MB of the frame X_{-1}, TH is a particular threshold, C_{XT} and C_{XS} are the RD costs of the temporal and spatial candidates, respectively.

III. Experimental Results

The proposed algorithm was implemented in JM 12.1 [4] provided by JVT for the performance evaluation. JM 12.1 was run in the Baseline Profile. In the last sections, we proposed the fast inter-mode decision algorithm STMP which is evaluated based on JM12.1 High Complexity option (exhaustive RDO mode decision) [4] using the following performance measures: 1) Degradation of image quality in term of average Y-PSNR: $\triangle PS$ (dB). 2) Increase of bit rate: $+BT$ (%). 3) Prediction rate in P frames: $PR = (N_{Pred}/N_{Total}) \cdot 100(\%)$ (N_{Total} is the total number of MBs and N_{Pred} is the number of MBs where the predicted mode equals the actual RD-optimal mode). 4) Encoding time saving: $TS = (T_{REF} - T_{PROP}/T_{REF}) \cdot 100(\%)$ (T_{REF} and T_{PROP} are the total encoding times of the reference and the proposed method, respectively). In the experiment of the STMP, an exhaustive mode decision is implemented at an interval of 20 frames to prevent error propagation. Alternatively, an exhaustive mode desicion could be applied for 5% of the MBs of each frame resulting in the same computational complexity. The value α has been found empirically based on the trade-off between computational complexity and coding efficiency. If α is less than 1.0, only a few of MBs will be coded in the same mode, resulting in a high coding efficiency but also in a high computational complexity. A high value for α would give an opposite effect. In our experiments, we used $\alpha = 1.1$.

We compared the performance of the proposed algorithm $STMP$ with two alternative methods: JM12.1 Fast High Complexity option (JM12.1 FHC) [4] and the spatial mode prediction based method (SP) [9], since SP [9] belongs to the same class of RDO-based methods as our proposal. In order to investigate the effects of individual techniques of $STMP$, we do experiment with three cases, using only mode prediction ($M+$), using mode and RD cost prediction ($M + RD+$), and *STMP* seen in Fig. 5. Note that in case of $M+$ and $M + RD+$, all modes are checked if the prediction does not deliver any mode (e.g., if the RD costs of the prediction mode are higher than $\alpha \cdot C_{Pred}$ in case of $M + RD+$).

For the CIF video format, Table IV shows about 45% of time saving with 0.04 dB PSNR degradation and 2.4% extra bits, while JM12.1 FHC shows the best RD efficiency and the worst time saving with 0.02 dB PSNR degradation, 0.2% additional bits and 14% of time saving. The SP algorithm demonstrates 35% of average time savings, 0.08 dB of PSNR degradation

TABLE IV
THE COMPARISON IN THE PERFORMANCE MEASURES, $\mathrm{QP} = 24$, CIF, 300 FRAMES

	JM12.1 FHC[4]			SP[9]			Proposal (M+)			Proposal (M+RD+)			Proposal (STMP)			
sequences	△PS	+BT	TS	△PS	+BT	TS	△PS	+BT	TS	△PS	+BT	TS	△PS	+BT	TS	PR
mobile	-0.01	0.0	8.5	-0.10	3.1	29.7	-0.27	4.1	45.1	-0.01	2.5	34.1	-0.01	2.9	38.6	43.8
coastguard	0.00	0.0	3.3	-0.10	2.3	26.1	-0.35	5.3	44.9	-0.07	1.9	33.9	-0.10	1.8	39.4	44.5
foreman	-0.02	0.1	11.3	-0.09	3.5	38.2	-0.28	4.7	51.8	-0.04	2.0	37.3	-0.05	2.1	40.9	49.3
stefan	-0.02	0.4	9.6	-0.08	3.7	28.6	-0.34	3.8	53.3	-0.01	2.3	38.5	-0.02	2.5	41.0	51.4
mother&daughter	-0.02	0.4	24.6	-0.03	3.9	46.3	-0.31	4.1	61.8	-0.03	1.5	47.6	-0.03	1.6	53.2	58.6
container	-0.04	0.2	23.8	-0.10	3.3	37.5	-0.25	4.6	63.5	-0.04	2.2	49.8	-0.05	2.6	56.5	62.6
average	-0.02	0.2	13.5	-0.08	3.3	34.4	-0.30	4.4	53.4	-0.03	2.1	40.2	-0.04	2.4	44.9	51.9

and 3.3% of extra bits. Moreover, it is found that, when considering the effects of individual parts of the *STMP* algorithm on speeding up and RD performance, using only mode prediction $(M+)$ results in similar time savings but poor RD performace. Combining mode prediction and RD cost prediction $(M+RD+)$ decreases speed up by 5–10% compared to *STMP* . This leads back to the fact that *STMP* does not always require to check all modes in case the $M + RD+$ does not provide a prediction. The experimental results shown for $\mathrm{QP} = 24$ are also valid for QCIF and other values of QP.

Table IV also shows that the prediction hit-rate of the best mode depends on the contents and resolutions of the video sequence, that is, how slow or fast motion is, and how fine the spatial resolution is. In terms of the tradeoff of complexity, quality and compression rate, the *STMP* algorithm shows the most promising results among other methods from the above comparisons. Compared to [11] and [12], the *STMP* is about 3%–10% faster.

IV. CONCLUSION

In this paper, we presented the new RDO-based algorithm STMP to speed up the inter-mode decision process for H.264/AVC encoding. We determine the best inter-coding mode of a given MB by predicting the best mode from neighboring MBs in time and space and by estimating its RD cost from the MB in previous frame. Simulation results demonstrated that the proposed algorithm can save 45% of total encoding time of JM12.1 video coder with bit rate increase limited to 2.4% at the same PSNR, which outperforms SP algorithm [9] in all performance measures. On average, the proposed algorithm determines the best coding mode for 44%–52% of the MBs using only one or two RD cost computations. The coding modes of the remaining MBs are determined using exhaustive mode decisions or other fast mode decisions. Therefore, the results of this work can be combined with other fast methods [5], [8]. Compared to non-RDO and RD estimation based classes [10], [16] for fast inter-mode decision, the proposed STMP enables faster encoding without changing the encoder architecture with almost identical rate-distortion performance.

REFERENCES

[1] *Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification*, ITU-T Rec. H.264 and ISO/IEC 14496–10 AVC 2003.

[2] G. Sullivan and T. Wiegend, "Joint video team, rate-distortion optimization for video compression," *IEEE Signal Process. Mag.*, vol. 11, no. 1, pp. 74–90, Jan. 1998.

[3] Y. W. Huang, B. Y. Hsieh, S. Y. Chien, S. Y. Ma, and L. G. Chen, "Analysis and complexity reduction of multiple reference frames motion estimation in H.264/AVC," *IEEE Trans. Circuit Syst. Video Technol*, vol. 16, no. 4, pp. 507–522, Apr. 2006.

[4] *JM 12.1 Test Mode CODEC*, ISO/IECMPEG and ITU-T VCEG Joint Video Team [Online]. Available: http://iphome.hhi.de/suehring/tml

[5] *Text Description of Joint Model Reference Encoding Methods and Decoding Concealment Methods*, ISO/IEC MPEG and ITU-T VCEG Joint Video Team, JVT-O079, 2005.

[6] C. S. Kannangara, I. E. G. Richardson, M. Bystrom, J. R. Solera, Y. Zhao, A. MacLennan, and R. Cooney, "Low-complexity skip prediction for H.264 through Lagrangian cost estimation," *IEEE Trans. Circuit Syst. Video Technol*, vol. 16, no. 2, pp. 202–207, Feb. 2006.

[7] J. Chen, Y. Qu, and Y. He, "A fast mode decision algorithm in H.264," in *Proc. of the Picture Coding Symp*, San Francisco, CA, USA, 2004.

[8] P. Yin, H.-Y. C. Tourapis, A. M. Tourapis, and J. Boyce, "Fast mode decision and motion estimation for JVT/H.264," in *Proc. Int. Conf. Image Process.*, 2003, pp. 853–856.

[9] C. Y. Chang, C. H. Pan, and H. Chen, "Fast mode decision for P-frames in H.264," presented at the Picture Coding Symp., San Francisco, CA, 2004.

[10] 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. Circuit Syst. Video Technol.*, vol. 15, no. 6, pp. 953–958, Jun. 2005.

[11] B.-G. Kim, S.-K. Song, and C.-S. Cho, "Efficient inter-mode decision based on contextual prediction for the P-slice in H.264/AVC video coding," in *Proc. Int. Conf. Image Process.*, 2006, pp. 1333–1337.

[12] B.-G. Kim and C.-S. Cho, "A fast inter-mode decision algorithm based on macro-Block tracking for P slices in the H.264/AVC video standard," in *Proc. Int. Conf. Image Process.*, 2007, pp. 301–304.

[13] W. I. Choi, J. Lee, S. Yang, and B. Jeon, "Fast motion estimation and mode decision with variable motion block sizes," *Proc. SPIE*, vol. 5150, pp. 1561–1572, 2003.

[14] J. Lee and B. Jeon, "Pruned mode decision based on variable block sizes motion compensation for H.264," *Proc. MIPS*, vol. LNCS 2899, pp. 410–418, 2003.

[15] E. Arsura, G. Caccia, L. D. Vecchio, and R. Lancini, "JVT/H.264 rate-distortion optimization based on skipping mechanism and clustering mode selection using MPEG7 transcoding hints," presented at the Picture Coding Symp., San Fransisco, CA, 2004.

[16] Y. K. Tu, J.-F. Yang, and M.-T. Sun, "Efficient rate-distortion estimation for H.264/AVC coders," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 16, no. 5, pp. 600–611, May 2006.

[17] A. Ichigaya, M. Kurozumi, Y. Nishida, and E. Nakasu, "A method of estimating coding PSNR using quantized DCT coefficients," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 16, no. 2, pp. 251–259, Feb. 2006.