Fast Mode Decision for H.264/AVC using Mode and RD Cost Prediction

Ri Song-Hak, Joern Ostermann
Institut für Informationsverarbeitung
Leibniz Universität Hannover
Appelstr. 9A, D-30167, Hannover, Germany
{ri, ostermann}@tnt.uni-hannover.de

Abstract-**In an H.264/AVC encoder, each macroblock can be coded in one of a large number of coding modes, which requires a huge computational effort. In this paper, we present a new method to speed up the mode decision process using RD cost prediction in addition to mode prediction. In general, video coding exploits spatial and temporal redundancies between video blocks, in particular temporal redundancy is a crucial key to compress a video sequence with little loss of image quality. The proposed method determines the best coding mode of a given macroblock by predicting the mode and its rate-distortion (RD) cost from neighboring MBs in time and space. Compared to the H.264/AVC reference software, the simulation results show that the proposed method can save about 60% of the number of RD cost computations resulting in up to 57% total encoding time reduction with up to 3.5% bit rate increase at the same PSNR.**

Keywords-H.264/AVC encoder; mode decision; rate-distortion optimization; mode prediction; RD cost prediction

I. INTRODUCTION

Video coding plays an important role in multimedia communications and consumer electronics applications. The compression performance achieved by international video coding standard H.264/AVC which is jointly developed by the ITU-T Video Coding Experts Group and the ISO/IEC Moving Picture Experts Group enables high quality compressed video on a wide range of platforms [1].

However, it requires a huge amount of computational loads due to use of the variable block-size motion estimation, intra prediction in P slice coding, quarter-pixel motion compensation, multiple reference frames, etc. The complexity analysis described in [1] shows that examining all possible modes takes the most time out of the total encoding time. Hence, fast mode decision making becomes more and more important.

H.264/AVC Baseline profile employs seven different block sizes for inter frames. The size of a block can be 16x16, 16x8, 8x16, or 8x8, and each 8x8 can be further broken down to sub-macroblocks of size 8x8, 8x4, 4x8, or 4x4. To encode a given macroblock, H.264/AVC encoder tries all possible prediction modes in the following order; SKIP, Inter16x16, Inter16x8, Inter8x16, Inter8x8, Inter8x4, Inter4x8, Inter4x4, Intra4x4, Intra8x8, Intra16x16. The SKIP mode represents the case in which the block size is 16x16 but no motion and no residual information are coded. Except for SKIP and intra modes, each inter mode decision requires a motion estimation process.

In order to achieve the highest coding efficiency, H.264/AVC uses rate distortion optimization techniques to get the best coding results in terms of maximizing coding quality and minimizing coded data bits. The mode decision is made by comparing the rate distortion cost of each possible mode and by selecting the mode with the lowest rate distortion cost as the best one.

The existing fast mode decision algorithms can be classified into two categories:

The first class is to find the optimal mode by using some features, such as texture and edge information, which are computed from the raw video data. D. S. Turaga et al [5] and J. Chen et al [2] introduce the so-called mean removed mean absolute difference (mrMAD) and use the feature to make fast intra and inter mode decision. In [6], the 3x3 Sobel operater is used to get the edge map of a whole frame. The edge map and the gradient are both employed to find the best interpolation direction as the best intra mode. They also use the edge map to determine whether a macroblock is homogeneous in order to find the best inter mode. However, the algorithm has to evaluate all the pixels in the whole frame and it leads to high computational complexity.

The second class is trying to make full use of the relationship among the modes and predicts the best mode by using the already checked modes and their statistical data. A representative method [4] of such class divides all modes into 3 groups. Using one mode from each group, the best group is determined. All modes of the best group are evaluated to determine the best mode selection. Thus the number of candidate modes is greatly reduced. In [3], the most probable mode is predicted based on the observation that most modes are spatially correlated in a given frame. If the predicted mode satisfies some conditions which estimate if the predicted mode is the best mode, the encoder codes the macroblock with the predicted mode. Thus it can skip all of the calculations on other modes.

In [8], we combined temporal and spatial mode predictions which used the RD correlation in making a decision whether or not the predicted mode could be the best one.

Based on the analysis above, in addition to the mode prediction from [8], we present a new algorithm to make a progress to determine the best mode by using the prediction the optimal RD cost for a given macroblock.

This paper is organized as follows: Section 2 discusses mode prediction for fast mode decision. In Section 3, we propose a new mode decision scheme using mode and its RD cost prediction, and finally, experimental results and conclusions are presented in Section 4 and Section 5, respectively.

II. MODE PREDICTION

Video coding is achieved by reducing spatial and temporal redundancies between video frames. This implies indirectly that a mode of a given macroblock (MB hereafter) also might be correlated to that of MBs neighboring in space and time. It was noted that there was a spatial mode correlation between a given MB and its neighboring MBs and, therefore, it is possible to spatially predict a mode of the MB [3].

Since a video sequence contains, in general, more redundancy in time domain than in space domain, we stipulate that temporal mode correlation is higher than 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:

- How *high* is the correlation of spatial and temporal modes?
- Is it necessary to consider *all* modes for the mode prediction?

Let's mark the current MB as X, collocated MB of X in the previous frame as X_{-1} and neighboring MBs as A, B, C and D (see Fig. 1).

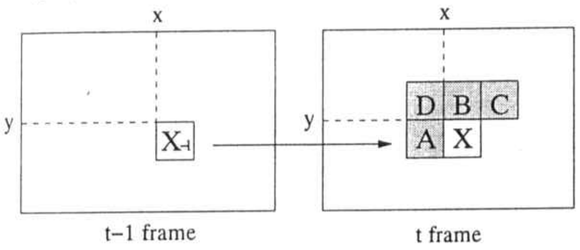

Figure 1. The current MB X, its collocated MB X_{-1} of the previous frame, and neighboring MBs, A, B, C and D.

To compare correlation of both mode predictions, let's define the following 3 events:

E_S : Modes of 2, 3 or 4 MBs out of A, B, C and D are the same as the RD-optimal mode of X.

E_T : The mode of X_{-1} is the same as the RD-optimal mode of X.

$E_C : E_S \cup E_T$

Here, E_S, E_T and E_C denote spatial, temporal and combined mode events. Table 1 shows the probabilities (**P_S**, **P_T** and **P_C**) of each event for the video sequences *container*, *mother & daughter*, *stefan*, *mobile*, *foreman* and *coastguard*.

In the Table 1, it is found that the probability of spatial mode prediction is lower than the probability of temporal mode prediction and, of course, combined mode prediction is also greater than spatial or temporal mode correlation. 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 *foreman* and *coastguard*, which are characterized by fast motion, 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 correlation, the encoder can predict the best mode of a given MB more frequently than by using spatial mode correlation. From now on, the combined mode prediction will be called mode prediction.

TABLE I. COMPARISON OF AN OCCURRENCE PROBABILITY OF SPATIAL, TEMPORAL AND COMBINED MODE EVENTS, QP (QUANTIZATION PARAMETER)=28, QCIF, 300 FRAMES.

Sequences	**P_S (%)**	**P_T (%)**	**P_C (%)**
Container	46.9	53.6	68.7
mother-daughter	33.8	40.3	56.5
Stefan	13.8	31.2	41.7
Mobile	10.2	27.4	35.5
Foreman	8.8	29.8	34.2
Coastguard	8.5	24.0	32.1

To answer to the second question, let's calculate the probability of an event where the RD optimized mode of a given MB is SKIP, Inter16x16, Inter16x8, Inter8x16, Sub8x8, Intra4x4 and Intra16x16, under the condition that X has the same RD optimized mode as X_{-1} (see Table 2). Let's mark the probability $P(SKIP|X = X_{-1})$ as **P_{SKIP}**, $P(Inter16x16|X = X_{-1})$ as **$P_{Inter16x16}$**, … , $P(Intra4x4|X = X_{-1})$ as **$P_{Intra4x4}$** and $P(Intra16x16|X = X_{-1})$ as **$P_{Intra16x16}$**.

TABLE II. STATISTICS OF MODEWISE-TEMPORAL MODE CORRELATION IN CASE X AND X_{-1} HAVE THE SAME RD-OPTIMAL MODE (UNIT=%), QP=28, QCIF, 300 FRAMES

Sequences	**P_{SKIP}**	**$P_{Inter16x16}$**	**$P_{Inter16x8}$**	**$P_{Inter8x16}$**	**P_{Sub8x8}**	**$P_{Intra4x4}$**	**$P_{Intra16x16}$**
container	79.5	9.4	3.6	3.7	3.7	0.0	0.1
mother-da.	64.9	16.6	6.1	6.9	5.2	0.1	0.0
stefan	21.1	32.2	10.3	15.1	20.2	0.6	0.5
mobile	17.4	27.2	13.1	10.2	31.7	0.1	0.4
foreman	12.3	39.8	13.5	11.7	22.4	0.2	0.1
coastguard	3.4	19.7	16.0	14.2	46.6	0.0	0.0

As seen in Table 2, the Intra4x4 and the Intra16x16 modes in the MB X happen very rarely when X_{-1} has the modes Intra4x4 and Intra16x16, respectively. Therefore, we don't use the predicted modes, Intra4x4 and Intra16x16, as candidates for the best mode of a given MB, if the predicted mode is Intra4x4 or Intra16x16.

III. FAST MODE DECISION BY MODE AND RD COST PREDICTION

The most important thing for applying mode prediction to fast mode decision is to make sure that the predicted mode is the best mode for a given MB. So far, there have been several ways to decide whether the predicted mode is the best mode of the MB or not.

The most common method [3] adopts a threshold value derived from the RD cost which is already calculated. The threshold is set to an average of RD costs of neighboring MBs with the same mode and it is compared with the RD cost of the given MB *X* with the predicted mode to estimate if it is the best mode. Another method [7] adopts the square of the quantization parameter (QP) as a threshold to decide whether the predicted mode is to be used.

For the sequence foreman, the RD cost difference between the spatially predicted mode and the optimal mode is shown in Fig. 2. The size of this difference does not necessarily depend on the actual RD cost. Therefore, a threshold based on neighboring MBs or QP should not be used for evaluating the quality of the predicted mode.

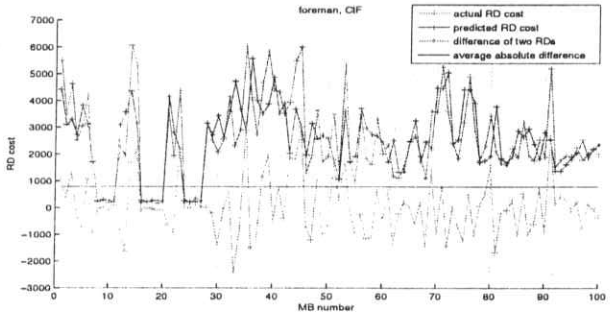

Figure 2. Relationship between RD cost of *X* and average RD cost of neighboring MBs with the same mode (spatial RD cost prediction).

In [8], we used the RD cost of X_{-1} as the threshold. Fig. 3 intuitively shows a relationship between the actually optimal RD cost of *X* and the optimal RD cost of X_{-1} when the optimal modes of *X* and X_{-1} are the same.

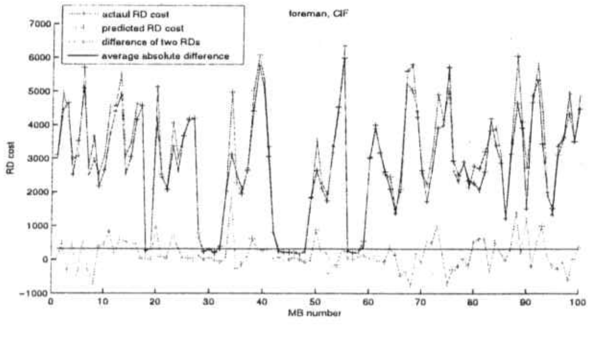

Figure 3. Relationship between RD cost of *X* and RD cost of X_{-1} when the best mode of *X* is the same as one of X_{-1} (from [8]).

The relationship between the RD costs of *X* and X_{-1} can be seen in the comparison of the following three correlation coefficients: correlation coefficient (ρ_S) between the spatially predicted RD cost and the optimal RD cost, correlation coefficient ($\rho_{T'}$) between the actually optimal RD cost of *X* and the optimal RD 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 optimal RD cost of *X* and the optimal RD cost of X_{-1}. Table 3 shows that the temporal correlation is greater than the spatial one.

TABLE III. COMPARISON OF THREE CORRELATION COEFFICIENTS FOR DIFFERENT METHODS OF PREDICTING THE RD COST OF A MB (FROM [8])

Sequences	QCIF			CIF		
	ρ_S	$\rho_{T'}$	ρ_T	P_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-da.	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

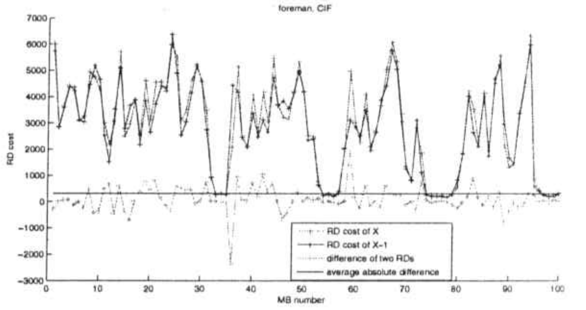

Figure 4. Relationship between RD cost of *X* and RD cost of X_{-1} (from [8]).

From Fig. 4 and Table 3, it should be noted that the correlation of RD costs is great even in the case that the optimal modes of *X* and X_{-1} are not the same, which means that the optimal RD cost of a MB can be predicted by RD cost of the optimal mode of the previous MB.

Unlike the case discussed in [8], in the case that the optimal mode of X_{-1} doesn't equal the optimal mode of *X*, we can use the correlation 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, Inter16x16, ... or P8x8.

Fig. 5 shows that the optimal RD cost difference for *X* is close to zero approximately when the 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 the range of the RD cost of X_{-1}.

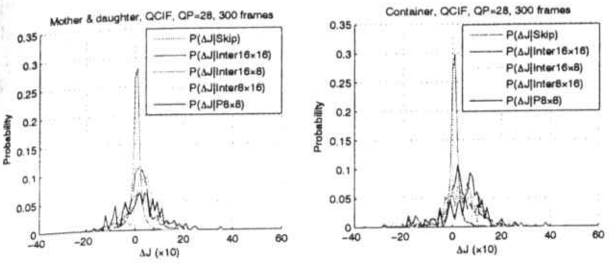

Figure 5. Measured pdfs of RD cost differences, where P(ΔJ|MODE) denotes the probability of RD cost difference (ΔJ) between X (irrespective of mode) and X_{-1} given MODE

Another problem in using mode prediction is error propagation, due to the misprediction of a mode. To prevent the propagation of mode prediction errors, an exhaustive mode decision will be carried out periodically. In the experiment, error propagation is likely to happen more frequently in video sequences with smooth and slow motion.

The last problem of mode prediction is that using only one predicted mode to decide upon the best mode could be unstable. It has been observed that sometimes temporal mode prediction shows better result than spatial one, and also vice versa. Therefore we apply two mode candidates, m_t and m_s, predicting the mode temporally and spatially and choose the mode with the lower RD cost. Fig. 6 shows the three probabilities: the red curve is the probability that the chosen candidate, m_t or m_s, is the best mode, the blue curve the probability that the other mode, m_s or m_t, is the best one and the green curve shows the probability that a different mode is the best mode. The probability that the chosen candidate is the best mode is far higher than the probability of a different mode.

The proposed algorithm is as follows:

1. *If the current frame is an exhaustive mode decision frame, check all modes and stop mode decision.*
2. *Get two predicted modes, m_t and m_s, from temporal and spatial mode predictions.*
3. *Get the actual RD cost, RD_{Pred}, of the collocated MB in the previous frame.*
4. *If both predicted modes are the same, apply it to the current MB, otherwise, compare the two RD costs by applying both modes and choose the better one.*
5. *If the chosen RD cost is lower than the threshold, $TH=\alpha \cdot RD_{Pred}$, set it to the best mode and stop, otherwise go to the next step (here, α is a positive, sequence-independent constant derived from experiments).*
6. *Check all other modes. If m_t is SKIP, stop checking remaining modes as soon as a mode with RD cost < TH is found.*

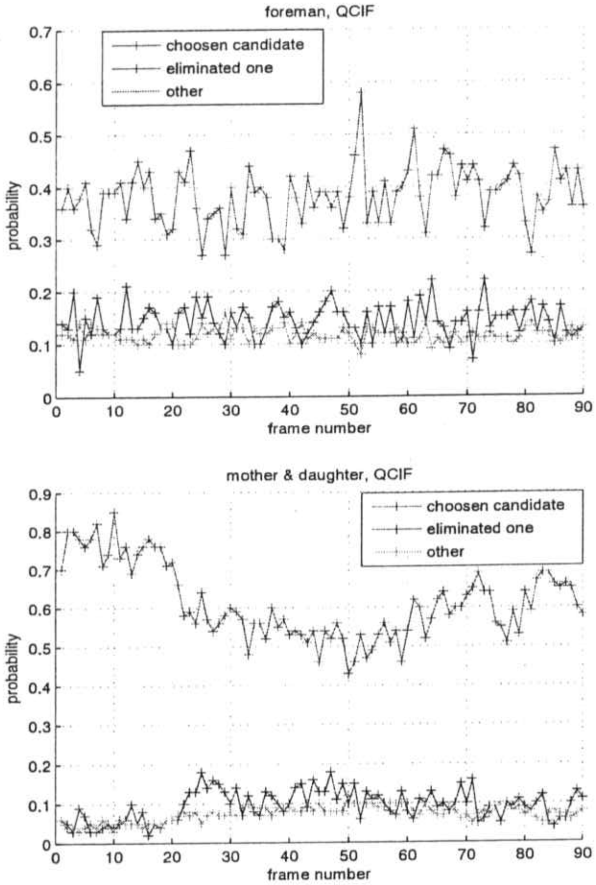

Figure 6. Probabilities at which the chosen, the eliminated candidate or other mode is the same as the RD-optimal mode (average probability in the case of other mode).

IV. EXPERIMENTAL RESULTS

The proposed fast mode decision scheme was implemented in H.264/AVC reference software JM 10.1 baseline profile for performance evaluation. The experimental conditions are as follows:

Software & Profile: H.264/AVC reference software JM 10.1 Base-line

Sequences : *container, coastguard, stefan, foreman, mobile, mother & daughter*

Video Format : QCIF, CIF

ME Strategy : Full Motion Estimation

The proposed algorithm was evaluated based on the exhaustive RDO mode decision of H.264/AVC in the following performance measures:

- Degradation of image quality in term of average Y-PSNR: ΔPSNR (dB)
- Increase of bit rate: +Bits (%)

- Prediction rate: PR (%)

$$\mathrm{PR} = \frac{N_{Pred}}{N_{Total}} \times 100(\%),$$

where, N_{Total} is total number of MBs and N_{Pred} is the number of MBs where the predicted mode equals the RD optimized mode.

- Encoding time saving: TS (%)

$$\mathrm{TS} = \frac{T_{REF} - T_{PROP}}{T_{REF}} \times 100(\%),$$

where, T_{REF} and T_{PROP} are the total encoding times of the REFerence and the PROPosed method, respectively.

In the experiment, the exhaustive mode decision is implemented at an interval of 10 frames, to prevent error propagation. α is set to 1.1.

We compared the performance of the proposed method with that of two alternative methods which are based on spatial mode prediction [3] and combinational mode prediction [8]. For the QCIF video format, Table 4 shows that the proposed algorithm can achieve 45% of average time savings in total encoding time with 0.04dB PSNR degradation and 2.1% extra bits.

For the CIF video format, Table 5 shows about 53% of time saving with 0.07dB PSNR degradation and 2.7% extra bits. The proposed algorithm shows better performance than that of the alternative algorithm [3], which is achieving about 27% of average time savings, 0.03dB of PSNR degradation and 3.6% of extra bits for QCIF sequences, and 35% of average time savings, 0.08dB of PSNR degradation and 3.3% of extra bits for CIF sequence.

The proposed algorithm also shows that even though it has a marginally lower RD performance than our previous algorithm [8], it results in more time savings, in particular, showing a significant effect on low-to-mid motion sequences like *Mother & daughter*, *Container* and *Coastguard*.

TABLE IV. THE COMPARISON IN THE PERFORMANCE MEASURES, QP=24, QCIF, 300 FRAMES, WHERE THE ALTERNATIVE METHODS ARE SPATIAL AND TEMPORAL MODE PREDICTION BASED METHODS, RESPECTIVELY.

Sequences	**Alternative [3]**			**Alternative [8]**				**Proposal**			
	ΔPSNR (dB)	***+Bits (%)***	***TS (%)***	***ΔPSNR (dB)***	***+Bits (%)***	***TS (%)***	***PR (%)***	***ΔPSNR (dB)***	***+Bits (%)***	***TS (%)***	***PR (%)***
mobile	-0.05	3.6	24.6	0.00	1.4	42.5	45.6	-0.02	1.7	44.3	47.8
stefan	-0.06	4.7	29.6	0.04	1.9	43.2	49.6	-0.03	2.6	45.1	49.9
foreman	-0.01	3.5	26.5	-0.05	2.9	35.3	37.8	-0.05	2.8	39.3	41.6
mother-	-0.03	3.9	35.3	-0.02	1.2	46.4	51.5	-0.04	1.4	51.4	58.5
container	-0.02	3.7	26.2	0.01	1.7	35.1	40.7	-0.04	1.8	49.9	53.4
coastguard	-0.01	2.3	23.3	-0.02	2.0	35.3	38.9	-0.06	2.2	45.3	48.9
average	-0.03	3.6	27.6	-0.01	1.9	39.6	44.0	-0.04	2.1	45.7	51.6

TABLE V. THE COMPARISON IN THE PERFORMANCE MEASURES, QP=24, CIF, 300 FRAMES.

Sequences	**Alternative [3]**			**Alternative [8]**				**Proposal**			
	ΔPSNR (dB)	***+Bits (%)***	***TS (%)***	***ΔPSNR (dB)***	***+Bits (%)***	***TS (%)***	***PR (%)***	***ΔPSNR (dB)***	***+Bits (%)***	***TS (%)***	***PR (%)***
mobile	-0.10	3.1	29.7	-0.01	2.9	52.9	58.3	-0.05	2.9	53.9	59.2
stefan	-0.08	3.7	28.6	-0.02	2.5	47.4	55.5	-0.08	3.5	51.2	58.7
foreman	-0.09	3.5	38.2	-0.05	2.1	41.6	49.8	-0.09	3.1	47.3	52.6
mother-	-0.03	3.9	46.3	-0.03	1.6	46.0	52.4	-0.04	2.2	56.2	64.4
container	-0.10	3.3	37.5	-0.05	2.6	48.2	55.6	-0.06	2.4	57.2	59.1
coastguard	-0.10	2.3	26.1	-0.10	1.8	51.5	63.6	-0.10	2.2	54.9	61.6
average	-0.08	3.3	34.4	-0.04	2.4	47.9	55.9	-0.07	2.7	53.4	59.2

Table 4 and Table 5 also show that the prediction rate of the best mode depends on the contents and resolutions of video sequence, that is, how slow or fast motion is, and how fine the spatial resolution is. Fig. 7 shows the rate-distortion performance of the four algorithms, the exhaustive method, two alternative methods (based on spatial mode prediction [3] and mode prediction [8]) and the proposed method. The curves show that the proposed algorithm has better RD efficiency than the spatial mode prediction based method [3], achieving similar efficiency to the exhaustive method.

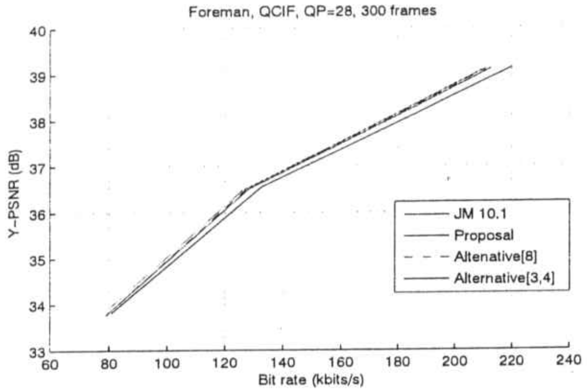

Figure 7. Comparison of several RD (PSNR vs. bitrate) plots, QP=24, 28 and 32.

CONCLUSIONS

In this paper, we proposed a new method to speed up the mode decision process using mode and RD cost prediction. The proposed method determines the best coding mode of a given MB by predicting the mode and its RD cost from neighboring MBs in time and space. This prediction reduces the number of RD cost computations by 60% compared to exhaustive RD cost calculations. Compared to the H.264/AVC reference software, the simulation result shows that the proposed method can save up to 57% of the total encoding time with up to 3.5% bit rate increase at the same PSNR.

REFERENCES

[1] J. Ostermann, J. Bormans, P. List, D. Marpe, M, Narroschke, F. Pereira, T. Stockhammer, T. Wedi, "Video Coding with H.264/AVC: Tools, Performance, and Complexity", IEEE Circuits and Systems Magazine, Vol. 4, No. 1, pp. 7-28, 2004.

[2] J. Chen, Y. Qu, Y. He, "A Fast Mode Decision Algorithm in H.264", 2004 Picture Coding Symposium (PCS2004), San Francisco, CA, USA, 2004.

[3] C. –Y. Chang, C.-H. Pan, H. Chen, "Fast Mode Decision for P-Frames in H.264", 2004 Picture Coding Symposium (PCS2004), San Francisco, CA, USA, 2004.

[4] P. Yin, H.-Y.C. Tourapis, A. M. Tourapis, J. Boyce, "Fast mode decision and motion estimation for JVT/H.264", in Proceedings of International Conference on Image Processing, pp. 853-856, 2003.

[5] D. S. Turaga, T. Chen, "Estimation and Mode Decision for Spatially correlated Motion Sequences", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 11, No. 10, pp. 1098-1107, 2001.

[6] K. P. Lim, S. Wu, D. J. Wu, S. Rahardja, X. Lin, F. Pan, Z. G. Li, "Fast Inter Mode Selection", JVT-I020, JVT 9th Meeting, san Diego, USA, 2003.

[7] E. Arsura, G. Caccia, L.D. Vecchio, R. Lancini, "JVT/H.264 Rate-Distortion Optimization based on Skipping Mechanism and Clustering Mode Selection using MPEG7 Transcoding Hints", 2004 Picture Coding Symposium (PCS2004), San Fransisco, CA, USA, 2004.

[8] Ri Song-Hak, Joern Ostermann, "Fast Mode Decision for H.264/AVC Using Mode Prediction", 2006 Advanced Concepts for Intelligent Vision System (ACIVS 2006), September 2006, Antwerp, Belgium (to be published)