Journal of Information & Computational Science 7: 10 (2010) 2071–2079
Available at http://www.joics.com

Hybrid Fast Mode Decision Algorithm for H.264/AVC High Profile ⋆

Meihua Gu [a,b], Ningmei Yu [a,*], Likang Kou [a], Chang Jiang [a]

[a] *School of Automation and Information Engineering, Xi'an University of Technology Xi'an 710048, China*

[b] *School of Electronic Information, Xi'an Polytechnic University, Xi'an 710048, China*

Abstract

In order to reduce the computational complexity of H.264/AVC high profile, a hybrid fast mode decision algorithm is proposed in this paper. For each macroblock (MB), rate-distortion (RD) cost at SKIP mode is computed firstly 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 only the intra modes are worthwhile to be checked. If the RD cost is between the above mentioned thresholds, the rest inter modes will be examined. At each stage of inter mode selection, temporal residual homogeneity and spatial residual homogeneity are checked for a possible early termination. In addition, the texture direction is predicted by exploiting the pixel direction error strengths for selecting the most possible candidate intra modes. Simulation results demonstrate that the proposed hybrid fast mode decision algorithm can achieve more than 60% timesaving than H.264 reference software JM16.1, with tiny degration of PSNR and less than 1% bit rate increase.

Keywords: H.264/AVC; Mode Decision; Temporal Homogeneity; Spatial Homogeneity; Texture Direction

1 Introduction

The High Profile of H.264 [1] was introduced in 2004 in response to the rapidly growing demand for high fidelity digital video in various applications such as professional film production and high definition TV/DVD [2]. This newly developed profile not only adopts multiple sophisticated techniques such as multiple reference frames, variable block sizes, quarter-pixel accuracy, enhanced intra-prediction, and so on, but also adopts other new coding tools to achieve much higher coding efficiency [2, 3]. One notable feature is that it adopts 8x8 intra prediction which, like the other coding tools adopted earlier, introduces significant computational overhead to the encoder. As

⋆Project supported by Special Scientific Research Program of Shaanxi Provincial Education Bureau (No. 2010JK558) and Special Funds of Shaanxi Major Subjects Construction (No. 107080903).

*Corresponding author.
Email address: `yunm@xaut.edu.cn` (Ningmei Yu).

October 2010

the trend of visual communications is moving towards high quality and high definition video, computational efficiency of video coding remains a challenging issue for many applications.

One important technique in H.264 high profile is using the Lagrangian rate-distortion optimization (RDO) performed for inter-mode prediction and intra-mode prediction. The RDO technique is used to check all possible coding modes and find the best coding result to obtain the highest coding efficiency. However, the selection of best coding modes is the most important part of a typical video encoder in computation. Experimental results have shown that the exhaustive mode decision process usually takes more than 90% of the total computational load of the JM reference software. To reduce the computation cost and maintain coding performance, many fast and efficient methods for mode decisions have been proposed in recent years [4-7]. Among them, reference [7] proposed a fast mode decision method, including two early termination stages, the first stage is to check the SKIP mode after 16x16 motion estimation, the other stage is after checking all the inter modes for the prediction of wether the rest intra mode decision is skipped. This algorithm can speed up mode selection process to some extent, SKIP mode is anticipated after motion estimation at inter 16x16, there are still calculating redundancy for those sequences with a large number of simple movement. In addition, whether intra mode selection should be checked is determined after the process of inter mode selection, this method is very effective for sequences with simple movement or moderate motion, because of the low probability of intra mode selection. But for those sequences with complex motion, large numbers of intra MB still need to check the RD cost for each inter mode, where exist vast computation redundancy.

To solve those problems mentioned above, this paper presents a novel hybrid fast mode selection algorithm for H.264/AVC high profile. The proposed method has three characteristics, one is checking RD cost of SKIP mode at the first time, two thresholds dependent on QP are set, SKIP mode and intra mode are pre determined before motion estimation according to the relationship of RD cost (SKIP) and the two thresholds. The second characteristic is the acceleration of inter mode selection by making use of the temporal and spatial features for the possible early termination during the process of inter mode selection. The third characteristic is the acceleration of intra mode decision by making use of predicting texture direction. Experimental results shows that the proposed hybrid algorithm based on the three aspects mentioned above can save approximately 60% computation time of JM16.1 with negligible effect on PSNR and less than 1% bit rate increase.

2 Mode Decision Algorithm for H.264/AVC

There are total of 11 types of candidate modes for the P-frame coding in H.264 high profile: SKIP, INTER16x16, INTER16x8, INTER8x16, P8x8 (8x8, 8x4, 4x8, 4x4), INTRA4x4 (I4MB), INTRA8x8 (I8MB), and INTRA16x16 (I16MB), both the inter prediction and intra prediction are supported. H.264 provides 7 kinds of block size for motion estimation in inter frame coding, and 3 kinds of block size for intra coding, as shown in Fig. 1.

To get better RD performance, H.264 uses RDO technology during the process of mode decision. Checking the RD cost for all the possible modes, the mode with least RD cost is selected as the best mode for current MB. RD cost function is defined as Eq. (1).

$$J(s, c, MODE|QP, \lambda_{MODE}) = SSD(s, c, MODE|QP) + \lambda_{MODE} \times R(s, c, MODE|QP) \quad (1)$$

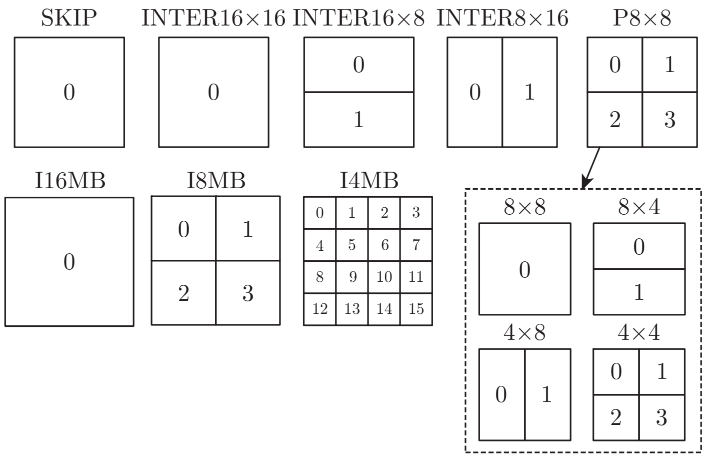

Fig. 1: The eleven modes (or block sizes) used in H.264/AVC high profile

where s and c is the source image and reconstructed image, $MODE$ represents one of MB mode, QP is quantization parameter, λ is lagrange multiplier, SSD is the sum of the squared differences between original signal s and its reconstruction c, $R(s, c, MODE|QP)$ is the number of bits associated with the chosen mode and QP, including the bits required for the chosen mode coding and the DCT coefficients for the given block.

3 Statistics and Analysis of Best Modes

Generally, selecting a larger block size usually leads to a larger motion compensated residual especially in regions containing different movements. On the contrary, selecting a smaller block size usually results in a smaller residual. For regions with homogenous motion, including uniform motion of rigid objects, smooth motion of a moving background, and zero motion of a static background, coding using larger block sizes will not result in a much larger residual. For regions with complex motion and regions containing motions from different objects, it is suitable to represent motion on a fine level, i.e., using smaller block sizes.

To count the distribution of best modes with full mode selection algorithm, seven sequences with different motion-activity are chosen, including News, Miss, and Paris with large still background, Football, Stefan and Mobile with fast motions in different directions, and Foreman with medium motions. The results are shown as Table 1.

As the video sequence has a strong correlation between time and space, the best modes are not in uniform distribution among the 6 modes. It has been observed that, many natural video sequence contains a lot of static background, and many regions with homogeneous motion or space, where most MBs tend to be coded as SKIP mode. As shown in Table 1, for the sequences News and Miss with low motion and simple texture, the best mode of most MBs is SKIP, only few MBs choose INTRA mode. For the sequences Football and Mobile with intense motions, the number of MBs choose large-block-size modes decreases, and more MBs choose P8x8 and INTRA mode. In addition, the distribution of best modes is not only in close relation with the features of motion and texture, but also changes with the varying value of QP. Small QP makes more MBs coded with small block-size, and large QP makes MBs tend to coded with large block-size, even SKIP mode. Therefore, QP value affects the results of mode decision closely.

Table 1: Distribution of best MB mode with full search decision when QP=28(%)

Sequences	SKIP	16x16	16x8	8x16	P8x8	INTRA
News	74.68	6.25	2.98	3.91	11.87	0.31
Miss	72.39	17.03	3.67	3.17	1.3	2.44
Paris	60.15	11.81	4.21	4.83	18.51	0.49
Foreman	28.49	25.78	10.66	12.82	21.04	0.21
Stefan	28.11	27.83	8.84	8.00	25.14	2.06
Mobile	7.84	32.75	9.89	9.59	39.59	0.34
Football	12.37	15.35	7.99	7.82	30.41	26.05
Average	40.57	19.54	6.89	7.16	21.12	4.56

It is observed that many natural video sequences contain a lot of regions with homogenous motion, and most MBs in these motion-homogenous regions finally end up with larger block sizes such as INTER16x16 or SKIP. Two example frames are shown in Fig. 2, in which the block sizes of the optimal modes selected using full inter-mode decision are represented by different sized boxes overlaid on the corresponding MB. The left one of Fig. 2 is the 10th frame of QCIF sequence News, it can be seen that nearly all MBs in the static background are coded using 16x16 block size, and most MBs in the face and body regions with slight homogenous motion are coded using block sizes larger than 8x8. However, MBs located at the boundary between objects with different motion activities, such as the the dancers behind the announcers (fast motion), are more likely to be coded using small block sizes such as P8x8, and it is notable that there are on MBs coded by INTRA mode. The right one of Fig. 2 is the 10th frame of QCIF sequcence Football, it can be seen that there are more MBs coded by INTRA modes, and the intense motion regions such as the moving parts of the sportsmen are coded using smaller block sizes than 8x8.

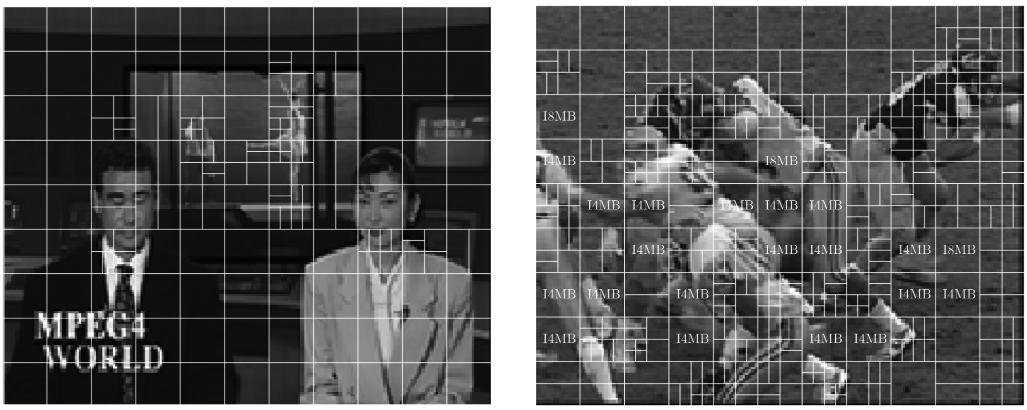

Fig. 2: Block sizes selected using full search mode decision for news and football

4 The Proposed Algorithm

4.1 Pre-determination of SKIP Mode and INTRA Mode

For each MB, RD cost at SKIP mode (RDcost (SKIP)) is computed firstly for a possible early termination. If the best mode of the current MB is SKIP, the predicted MV difference (MVD) is zero, the residuals after DCT and quantization are all zero. Therefore, there is no need to code the MVD and residuals, moreover, the reconstruction MB is just the MB in the reference frame that the predicted MV pointed to. As SKIP mode is very suitable for still background and the regions with homogeneous motion and simple texture, the prediction accuracy is comparatively high. It can be inferred from Eq. (1) that the RD cost at SKIP mode is the smallest when the best mode is SKIP. On the other hand, INTRA mode is usually suitable for the regions with intense motions, complex texture and scene cut. If these kind of MBs choose SKIP mode, the MB in the last reference frame with the predicted MV points to is taken as the reconstructed MB, this will bring about tremendous errors. It can be inferred by Eq. (1) as well, the RD cost at SKIP mode will be comparatively large. Two thresholds named T_{low} and T_{high} are set in this paper, if RDcost (SKIP) is below T_{low}, SKIP mode is determined to be the best mode, and all the remaining RD cost computations can be eliminated. If RDcost (SKIP) is larger than T_{high}, then all the inter-mode selection can be skipped, and only INTRA mode selection is worthy of being checked. If RDcost (SKIP) is between the two thresholds, only the remaining inter modes should be checked. The prediction accuracy highly dependent on the two threshold values, this kind of idea is also confirmed by reference [8]. Experiments show that the two thresholds have a close relationship with QP, just as shown in Eq. (2). It is clearly that different QPs can yield different results of T_{low} and T_{high}, and then different distribution of the best modes.

$$T_{low} = 34e^{0.1759\times QP}, \qquad T_{high} = 24215e^{0.0675\times QP} \tag{2}$$

4.2 Early Termination Methods During Inter Mode Decision

4.2.1 Checking for Residual Homogeneity

The main idea of determining whether the residual MB is homogeneous is as follows: if the four predicted residual sub-blocks in one NxN block (N=16 or 8) satisfy Eq. (3), then the NxN block is determined to be residual homogeneous [9].

$$\sum_{i=1}^{N/2}\sum_{j=1}^{N/2} |r_k(i,j) - M_k| < \zeta \tag{3}$$

where ζ is threshold to be determined, $r_k(i,j)$ is the residue of the kth sub-block ($N/2$x$N/2$) after predicting NxN block. The definition of $r_k(i,j)$ is as Eq. (4).

$$r_k(i,j) = \hat{f}(i+v_x, j+v_y) - f(i,j) \tag{4}$$

where $\hat{f}(i+v_x, j+v_y)$ and $f(i,j)$ represent the reconstructed and original pixel values at the position of (i,j) in the kth sub-block respectively. M_k is the mean of predicted residues of the kth sub-block, the definition is as Eq. (5).

$$M_k = \frac{4}{N^2}\sum_{i=1}^{N/2}\sum_{j=1}^{N/2} r_k(i,j) \tag{5}$$

For the parameter ζ, large ζ value can bring about more residual homogeneous blocks, and can speed up the process of inter-mode selection, but may reduce the coding performance. On the contrary, small ζ can guarantee the coding performance, but can not reduce the encoding time. Therefore, the value of ζ should be compromised according to the actual application. On the other hand, for the residual block comes from the predicted error after motion estimation, so the value of ζ must be in close relation with Q_{step}. For simplicity, ζ and Q_{step} are regarded as satisfying linear relationship. Experiments show that the rational ζ value is set to be $8Q_{step}$ when $N = 16$, and $2Q_{step}$ when $N = 8$.

4.2.2 Checking for Spatial Homogeneity

If the MB has spatial similarity, the motion must be uniform and is not easy to split into small pieces. In this paper, the spatial homogeneity is checked by exploiting the edge information of MB [10], which comes from the edge map produced by sobel operator.

For a luma pixel $p_{i,j}$, the edge vector $\vec{D}_{i,j} = \{dx_{i,j}, dy_{i,j}\}$ is defined as Eq. (6, 7).

$$dx_{i,j} = p_{i-1,j+1} + 2 \times p_{i,j+1} + p_{i+1,j+1} - p_{i-1,j-1} - 2 \times p_{i,j-1} - p_{i+1,j-1} \tag{6}$$

$$dy_{i,j} = p_{i+1,j-1} + 2 \times p_{i+1,j} + p_{i+1,j+1} - p_{i-1,j-1} - 2 \times p_{i-1,j} - p_{i-1,j+1} \tag{7}$$

where $dx_{i,j}, dy_{i,j}$ represent the residual intensity of vertical and horizontal direction respectively. The amplitude of edge vector can be computed as Eq. (8).

$$Amp(\vec{D}_{i,j}) = |dx_{i,j}| + |dy_{i,j}| \tag{8}$$

$Amp(\vec{D}_{i,j})$ can be exploited to determine whether a NxN block is spatial homogeneous, the determine condition is as Eq. (9).

$$H = \begin{cases} 1 \quad \text{if} \sum\limits_{i,j \in N \times N} Amp(\vec{D}_{i,j}) < TH \\ 0 \quad \text{if} \sum\limits_{i,j \in N \times N} Amp(\vec{D}_{i,j}) \geq TH \end{cases} \tag{9}$$

where TH is a pre-determined parameter. $H = 1$ indicates the spatial homogeneity for a NxN block, $H = 0$ indicates the complex texture of a NxN block. Experiments prove that excellent coding performance can be obtained with TH value set to be 20000 for 16x16 block and 5000 for 8x8 block.

4.3 Fast Intra Mode Selection Method

In each image block, if the texture has some direction, then this indicates that the sum of differences between the adjacent pixels in this direction is the smallest. By making use of this characteristic, the texture direction can be predicted, and the number of candidate modes can be reduced to some degree. Take intra 4x4 for example, there are total 8 different direction modes besides dc mode as shown in Fig. 3. For each 4x4 block, all the differences between two neighboring pixels in 8 directions can be computed corresponding to horizontal, vertical, diagonal-down-left, diagonal-down-right, vertical-right, horizontal-down, vertical-left, horizontal-up texture directions. Therefore, the pixel direction error strengths in the major directions can be expressed as Eq. (10).

$$\begin{array}{ll} d^{0^\circ} = |f(x, y+1) - f(x, y)| & d^{90^\circ} = |f(x+1, y) - f(x, y)| \\ d^{45^\circ} = |f(x+1, y-1) - f(x, y)| & d^{135^\circ} = |f(x+1, y+1) - f(x, y)| \end{array} \tag{10}$$

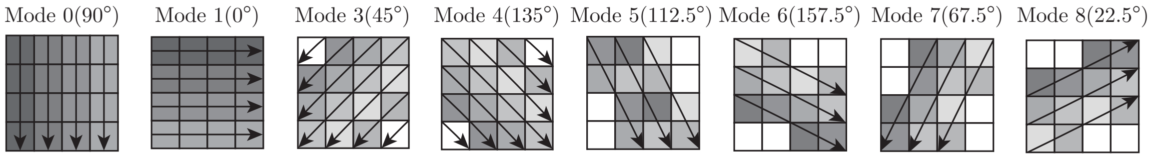

Fig. 3: Distribution of "equal pixels" in 4x4 luma intra-prediction modes besides dc mode

where x and y represent horizontal and vertical positions of pixel $f(x, y)$, respectively. Thus, the block direction error strengths by averaging all the possible pixel direction error strengths in each 4x4 block can be expressed by Eq. (11).

$$D^{0^\circ} = \frac{\sum_{i=0}^{11} di^{0^\circ}}{12} \qquad D^{90^\circ} = \frac{\sum_{i=0}^{11} di^{90^\circ}}{12} \qquad D^{45^\circ} = \frac{\sum_{i=0}^{8} di^{45^\circ}}{9} \qquad D^{135^\circ} = \frac{\sum_{i=0}^{8} di^{135^\circ}}{9} \tag{11}$$

Moreover, the remaining block direction error strengths for mode 5 to mode 8 can be obtained in a similar way according to the directions shown in Fig. 3.

According to the above computations, the mode with the smallest direction error strengths is chosen as the most possible candidate mode. To retain the prediction in smoother block, the dc mode is always chosen in the RDO. Therefore, by using the proposed method above, there are only 2 modes, instead of 9, will be chosen for 4x4 intra-prediction.

As for intra 8x8 luma prediction, the the same method as intra 4x4 can be used to reduce the most possible candidate modes from 9 to 2, except that the block size is 8x8. For intra 16x16 luma prediction, the pixel direction error strengths is checked in 16x16 block at vertical mode, horizontal mode and plane mode, and the number of candidate modes is also reduced from 4 to 2.

5 Experimental Results

The proposed fast mode decision algorithm is implemented on JM16.1 [11] provided by JVT, and tested on various sequences with different motion-activity and texture. The simulation conditions are as follows: encode the first 100 frames of each sequence; run H.264 high profile; RD optimization is enabled; QP is equal to 28, 32, 36 and 40; reference frame number equals to 1; CABAC is enabled; GOP structure is IPPPP.

The comparison results are produced and tabulated based on the difference of coding time (ΔTIME), the PSNR difference (ΔPSNR) and the bit-rate difference (ΔBIT) compared with full search mode decision defaulted in JM16.1, and the RD curve is also plotted by fitting the testing results according to reference [12]. The test platform used is Intel Core-2.0 GHz, 1 G Mbytes RAM. Note that in the table positive values mean increments, and negative values mean decrements.

Table 2 and Table 3 list the simulation results of the proposed and the previous [7] algorithms compared to full search. The proposed algorithm obviously outperforms the previous algorithm [7] on the whole, especially in timesaving.

For the sequences with weak motions and large still background, such as News and Miss, most MBs are coded using SKIP mode. Therefore, the technique of pre-determining SKIP mode

Table 2: Comparison results for QCIF sequences

QCIF sequences	Previous algorithm [7]			Proposed algorithm		
	ΔPSNR[dB]	ΔBIT[%]	ΔTIME[%]	ΔPSNR[dB]	ΔBIT[%]	ΔTIME[%]
News	−0.023	0.77	−18.67	−0.017	0.98	−76.93
Foreman	−0.018	0.14	−5.53	−0.020	0.32	−54.61
Silent	0.007	0.40	−14.76	−0.003	0.55	−63.25
Container	−0.012	0.96	−18.91	−0.009	0.98	−60.47
Football	−0.006	0.09	−0.78	−0.004	0.17	−59.28
Average	−0.011	0.54	−11.73	−0.011	0.60	−62.91

Table 3: Comparison results for CIF sequences

CIF sequences	Previous algorithm [7]			Proposed algorithm		
	ΔPSNR[dB]	ΔBIT[%]	ΔTIME[%]	ΔPSNR[dB]	ΔBIT[%]	ΔTIME[%]
Miss	−0.012	0.18	−1.74	−0.008	0.20	−73.54
Paris	−0.011	−0.16	4.75	−0.010	0.38	−56.21
Mobile	−0.002	0.05	−0.56	−0.012	0.41	−54.32
Tempete	−0.004	−0.06	−0.37	−0.009	0.39	−55.68
Stefan	−0.058	1.14	−4.60	−0.055	1.16	−58.73
Average	−0.018	0.23	−0.51	−0.019	0.51	−59.70

can reduce much encoding time by avoiding the remaining RDcost computations. It can be seen that the proposed algorithm achieves remarkable timesaving about 76.93% and 73.54% for News and Miss respectively, with negligible loss in PSNR and increments in bit rate. For those sequences with intense motion activities such as Football, there are many INTRA MBs, pre-determining INTRA mode can skip the process of inter mode checking, and the fast intra mode selection method can reduce a great deal of intra coding time, it can be seen from Table 2 that the timesaving is remarkable as well. For those sequences with medium motion activities and complex texture, such as Paris, Mobile, Stefan, Foreman, Paris and so on, most of the best modes are inter-modes with small block sizes, the early termination checking at each inter-mode stage can guarantee the possible timesaving.

Fig. 4 shows the RD curves of QCIF sequence News, and CIF sequence Mobile. The figures

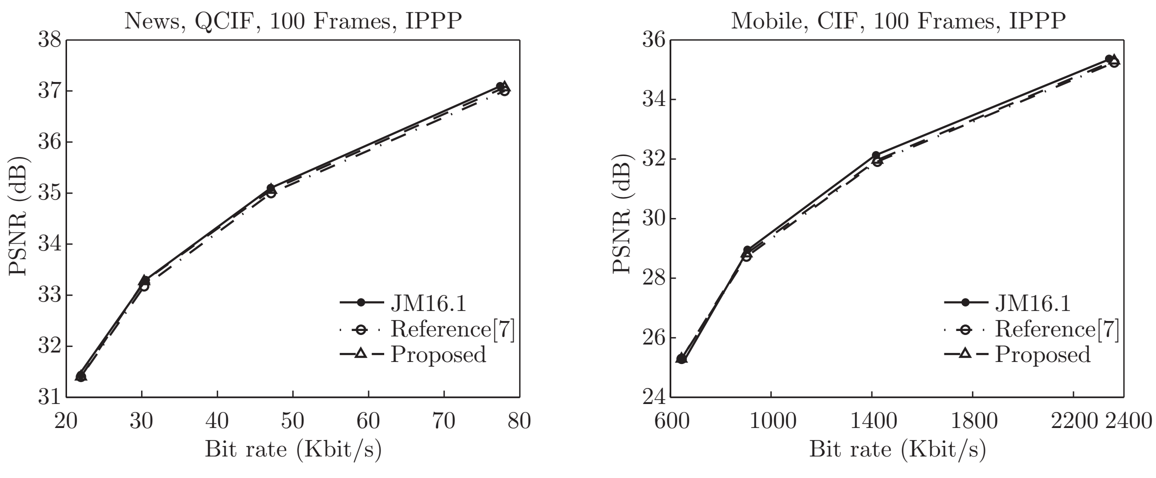

Fig. 4: RD curves of different sequences

have shown that the proposed algorithm has the similar RDO performance as that of JM16.1, and is superior than the previous algorithm [7].

6 Conclusion

This paper presents a fast mode decision algorithm for H.264/AVC high profile. By making use of pre-determination of SKIP and INTRA mode, a large amount of unnecessary computations are avoided. By making use of temporal and spatial homogeneity, the possible early termination can be guaranteed. By making use of predicting texture directions, the number of candidate intra modes has been largely reduced. Experimental results show that the proposed hybrid algorithm based on the three aspects mentioned above saved approximately 60% computation time of JM16.1 with negligible effect on PSNR and less than 1% bit rate increase.

References

[1] JVT, Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification, 2005

[2] D. Marpe, T. Wiegand, and S. Gordon, H.264/MPEG4-AVC fidelity range extensions: tools, profiles, performance, and application areas, in: Proc. IEEE International Conference on Image Processing, 2005, pp. 593-596

[3] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, Overview of H.264 video coding standard, IEEE Transactions on Circuits and Systems for Video Technology 13 (2003) 560-576

[4] 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: Proc. IEEE International Conference on Image Processing, 2006, pp. 2481-2484

[5] Y. K. Tu, J. F. Yang, and M. T. Sun, Statistical rate-distortion estimation for H.264/AVC coders, in: Proc. IEEE International Symposium on Circuits and Systems, 2006, pp. 365-368

[6] J. Bu, S. Lou, C. Chen, and J. Zhu, A predictive block-size mode selection for inter frame in H.264, in: Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, 2006, pp. 917-920

[7] 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 16 (2006) 1557-1561

[8] H. Q. Zeng, C. H. Cai, Fast mode decision for H.264/AVC based on macroblock motion activity, IEEE Transactions on Circuits and Systems for Video Technology 19 (2009) 491-499

[9] H. M. Wang, J. K. Lin, A successive termination and elimination method for fast H.264/AVC SATD-based inter mode decision, IET Signal Processing 3 (2009) 165-176

[10] F. Pan, X. Lin, and R. Susanto, Fast mode decision algorithm for intra prediction in JVT, in: The 7th JVT-G013 Meeting, March 2003

[11] JVT, Reference Software JM16.1, http://iphome.hhi.de/suehring/tml/download, 2009

[12] G. Bjontegaard, Calculation of average PSNR differences between RD-curves, in: The 13th VCEG-M33 Meeting, April 2001