A Novel Fast Inter Mode Decision Algorithm in H.264/AVC for Forest Fire Prevention Surveillance

Chen Chen[1], Ning Han[1], Chunlian Yao[2,*], and Yuan Li[1]

[1] Department of Automation, School of Technology, Beijing Forestry University, 100083 Beijing, P.R. China
[2] School of communication and art, Beijing Technology and Business University, 100048 Beijing, P.R. China
chenchen870713@gmail.com, hn217@bjfu.edu.cn, yaocl@btbu.edu.cn, bjfu.yuan.li@gmail.cn

Abstract. The new generation video coding standard H.264/AVC has been developed to achieve higher coding efficiency than previous video coding standards. Its variable block size motion compensation requires huge computation, which is a big problem for real time application. To reduce the complexity, we propose a fast inter mode decision algorithm with early mode prediction rules and early block size classification by using RD cost of the mode. Only parts of the inter prediction modes are chosen for motion estimation and rate distortion optimization. An adaptive threshold makes the proposed algorithm works smooth for different video sequences with a wide variety of activity. The experimental results indicate that the proposed algorithm can achieve 53.9% time saving in inter mode decision on average without any noticeable performance degradation, and the increased of bit rate is less than 1%.

Keywords: H.264/AVC, inter mode decision, spatial-temporal correlation, surveillance.

1 Introduction

The H.264/MPEG-4 Part-10 AVC [1] is the latest video coding standard developed by the Joint Video Team (JVT), which is an organization of the ISO Moving Picture Experts Group (MPEG) and the ITU-T Video Coding Experts Group (VCEG) [2]. It can achieve higher coding efficiency and less bit rate than other standards due to variable block size motion compensation, multiple reference frames, quarter-pixel motion vector accuracy, and so on. Compared with previous video coding standards such as H.263/MPEG-4, H.264/AVC incorporates seven block sizes from 16×16 to 4×4 for motion compensation [2,4]. It evaluates the distortion and rate of each candidate mode prior to selecting the mode for the current macroblock to achieve good rate-distortion performance. However, such exhaustively evaluating each mode entails high computational complexity which is a major problem for real-time application such as video surveillance. In forest video surveillance system, as in Fig. 1, video

* Corresponding author.

J. Liu et al. (Eds.): AMT 2009, LNCS 5820, pp. 397–408, 2009.

cameras are deployed in the forest, and each of them transmits the video to the terminal sever. Computational complexity reduction is desirable for low power consumption of the video sensor nodes as well as real-time video display on the screen in order to take measures as soon as a forest fire is found.

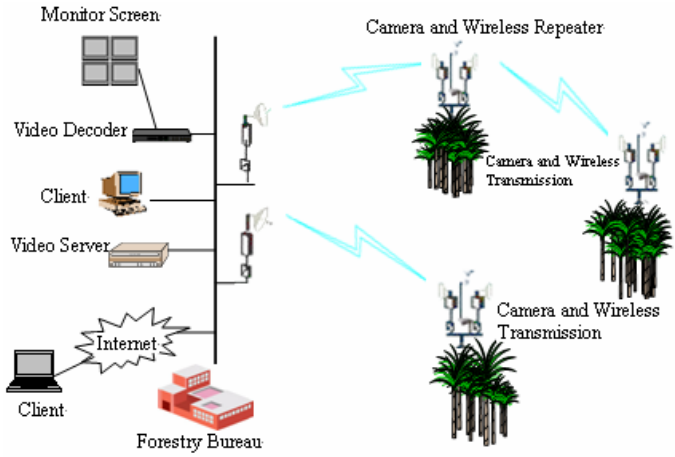

Fig. 1. Forest video surveillance system

To achieve this goal, many research have been done focusing on fast inter-frame mode decision. In [3], Kannangara et al. propose an algorithm for skip prediction in a H.264 encoder. However, this algorithm works well only for the low-motion sequences, and it still has to evaluate all the rest modes after a macroblock is chosen coding. Kim in [4] presents a fast macroblock mode prediction and decision algorithm based on contextual mode information for P-Slices and the probability characteristic of the inter mode. The concept of Region of Interest (ROI) is introduced in [5], and based on which different conditions for skip and large block mode decision are applied for the fast mode decision. Early skip mode detection, fast inter macroblock mode decision and intra prediction skip are three parts of the algorithm to reduce the computational complexity. In [7], a new fast mode decision algorithm is presented based on analysis of the histogram of the difference image between frames, which classifies the areas of each frame as active or non-active by means of an adaptive threshold technique, and it shows promising results with a time saving of over 43% for the test sequences. Common characteristic of these approaches is that the correlation among the modes of the macroblocks haven't been fully explored, which leaves a large place for further exploration.

In our proposed algorithm, we observe the modes of the macroblocks in a set of sequences encompass a wide variety of activity. After careful observation, we introduce five prediction rules by exploring the correlations among the modes of the macroblocks, and combined with an adaptive RD cost threshold TH_n to reduce the candidate modes.

The rest of the paper is organized as follows. In Section 2, we overview the rate-distortion optimized mode decision of H.264/AVC. Detailed statistical analysis is

presented in section 3. In Section 4, we propose our fast inter mode decision algorithm. In Section 5, we give our experimental results of the proposed algorithm. Finally, we draw some conclusions in Section 6.

2 Overview of Mode Decision in H.264/AVC

In H.264/AVC, in order to choose the best block size for a macroblock, RD cost of every candidate mode is calculated and the mode having the minimum RD value is chosen as the optimal mode. Fig. 2 depicts all available modes of a macroblock. The Lagrangian minimization cost function as follows,

$$J_{\mathrm{mod}e} = D + \lambda_{\mathrm{mod}e} * R$$

Where $J_{\mathrm{mod}e}$ is the rate-distortion cost and $\lambda_{\mathrm{mod}e}$ is the Lagrangian multiplier decided by QP value. D is the distortion measured as the sum of squared difference between original and reconstructed macroblock, and R reflects the total number of bits, including the macroblock header, motion vectors and the residual signal with regard to the coding mode [8].

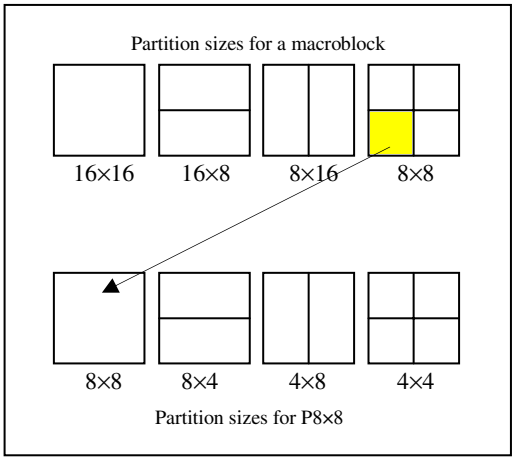

Fig. 2. Macroblock partition for inter-frame prediction

From Fig.2, we can see that the computation complexity for deciding the best inter mode is very high. Therefore, in the next section, we will develop a fast algorithm for inter mode decision to accelerate the coding process.

3 Observation and Statistical Analysis

In order to understand the percentage of each mode in the natural sequences, some experiments have been done on ten different sequences by using the H.264 reference software JM8.6[8]. Five test sequences are QCIF format and the other five sequences

are CIF format. All sequences are coding in IPPP structure with QP=28. We tested 250 frames for each sequence. The percentages of Skip mode and Inter16×16 mode are shown respectively in Fig. 3.

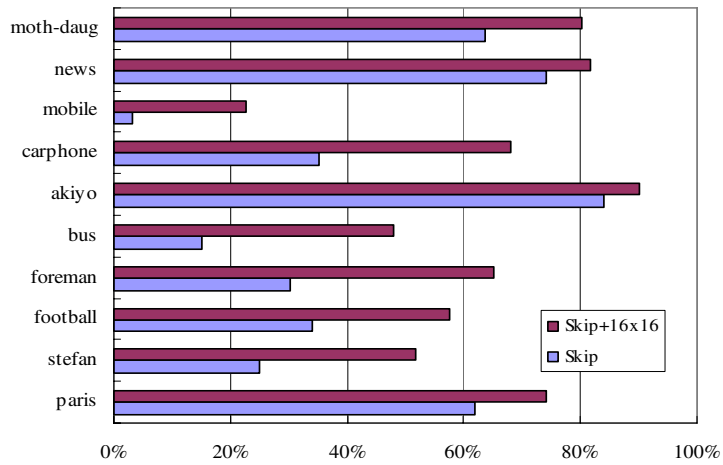

Fig. 3. Percentage of mode distribution

From Fig.3 we can easily know that Skip mode and Inter16×16 mode account for a large proportion of the mode distribution. Thus, if we can select the Skip mode and large partition size modes without going through all the candidate modes, a large quantity of coding time can be saved to meet the requirement of real time application.

3.1 Prediction Rules for Macroblock's Partition Mode

X is a macroblock in the nth frame; X' is the collocated macroblock in the previous frame ($(n-1)^{th}$ frame); A is the up macroblock of X; B is the left macroblock of X. The positions of the three neighbor macroblocks are showed in Fig. 4. Statistical data suggest that the modes of A, B and X' have strong correlation with the mode of current macroblock X.

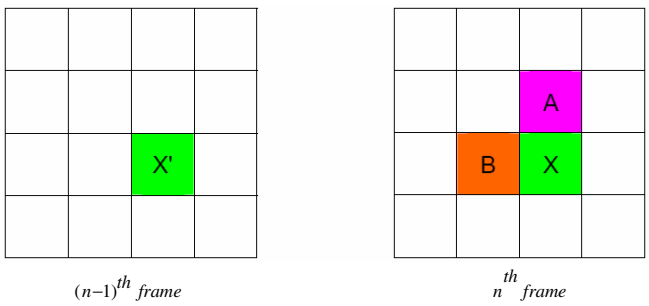

Fig. 4. Spatial and temporal neighbor macroblocks

According to the correlations between the modes of macroblock A, B and X' and the mode of current macroblock X, we develop five rules to predict the mode of macroblock X prior to motion estimation. The rules are illustrated as follows:

Case 1: If the modes of A, B and X' are all Skip mode, the mode of X is Skip mode.
Case 2: If the modes of A, B and X' all lie in subset {Skip, Inter16×16}, the mode of X lies in subset {Skip, Inter16×16, Inter16×8, Inter8×16}, but with one exception. That is when the modes of the three macroblocks are two Skip modes and an Inter16×16 mode, and then the mode of X lies in subset {Skip, Inter16×16}.
Case 3: If the modes of A, B and X' all lie in subset {Inter8×8, Inter8×4, Inter4×8, Inter4×4}, the mode of X lies in subset {Inter8×8, Inter8×4, Inter4×8, Inter4×4}.
Case 4: If at least one of A, B and X' has the mode in subset {Intra16×16, Intra4×4}, the mode of X lies in subset {Inter8×8, Inter8×4, Inter4×8, Inter4×4, Intra16×16, Intra4×4}.
Case 5: If the modes of A, B and X' don't match any conditions of the four cases above, then full search algorithm will be processed to find the best mode.

Since there are no top macroblock or left macroblock for the macroblocks in the first line or first column, the prediction rules for macroblocks in the first line and first column are different. We use the modes of macroblocks in the previous two frames, as shown in Fig. 5, to predict the mode of current macroblock.

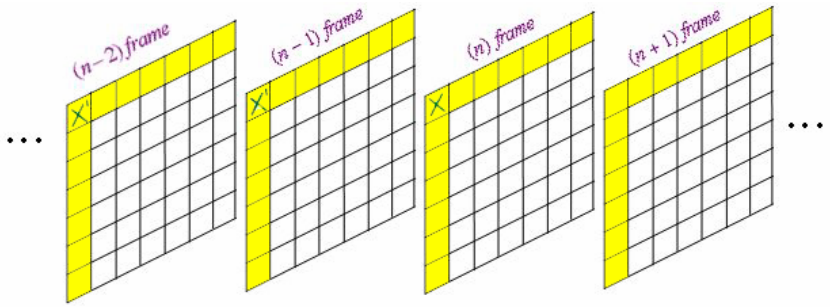

Fig. 5. The positions of macroblocks in the previous frames

The prediction rules for macroblocks in first line and first column are presented as follows:

Case 1: If the modes of the collocated macroblocks in previous two frames are Skip modes, the mode of the current macroblock is Skip mode.
Case 2: If the modes of the collocated macroblocks in previous two frames lie in subset {Skip, Inter16×16}, the mode of the current macroblock lies in subset {Skip, Inter16×16, Inter16×8, Inter8×16}.
Case 3: If the modes of the collocated macroblocks in previous two frames lie in subset {Inter8×8, Inter8×4, Inter4×8, Inter4×4, Intra16×16, Intra4×4}, the mode of the current lies in the same subset.

Case 4: If the modes of the collocated macroblocks in previous two frames don't match any conditions of the three cases, then full search algorithm will be processed to find the best mode.

Once we predicted which subset the macroblock belongs to, we code the macroblock with all the modes in the subset, and choose the mode with minimum RD cost as the best mode.

With the above prediction rules, we test the classification accuracy by using a few sequences. Table 1 reflects some statistical results of 6 sequences with 200 frames. In the table, Err Ratio is the percentage of the optimal encoding modes of macroblocks which are different from that obtained from exhaustive mode selection in all the macroblocks. Full Search Ratio is the percentage of macroblocks should be coded by exhaustive mode selection in all the macroblocks. Direct Skip Ratio is the percentage of Skip mode macroblocks that can be directly predicted by Case 1 in all Skip mode macroblocks.

Table 1. Statistical results of some sequences by using prediction rules

Sequence	Format	Err Ratio %	Direct Skip Ratio %	Full Search Ratio %
Akiyo.qcif	IPPPP	0.8	88.9	6.2
Coastguard.qicf	IPPPP	2.8	10.5	59.6
Carphone.qcif	IPPPP	3.7	35.7	58.9
News.qicf	IPPPP	3.0	65.4	20.2
Foreman.cif	IPPPP	3.6	24.6	44.5
Stefan.cif	IPPPP	3.5	29.4	53.6

As we can see from the table that sequences with high motion and rich details, there are still a large proportion of macroblocks needed to be coded with full search after performing the prediction rules. In addition, although some macroblocks can be predicted to be in certain subset, all the modes in the subset are tried in order to find the optimal encoding mode. Thus for those sequences with high motion and rich details, the computational complexity is only reduced a little. In this case, fast mode decision algorithm should go further.

3.2 Mode Classification by RD Cost

While encoding different video sequences with various QP values, we find that the RD cost of different modes vary a lot. We test a few sequences with different QP values and Table 2 lists the means of RD cost for several modes in different sequences. From the table, it is found that in inter modes, the RD cost for Skip mode has minimum mean, the RD cost of the large size macroblock modes (16×16, 16×8 and 8×16) have approximate means, but they are very different from those of small size macroblock modes (8×8,8×4,4×8,4×4). So, the fact can also be used to decide a Skip mode marcoblock at early stage and decide the possible modes subset for current macroblock.

After performing the prediction rules, we calculate the RD cost of Skip mode $J(Skip)$ for current macroblock using formula (1) and means of RD cost for

each mode in the previous frame, then compares $J(Skip)$ with the mean of Skip macroblocks' RD cost $Mean(Skip)$ to decide whether the current macroblock is a Skip macroblock, which is shown in formula (2).

$$J(Skip) = SSD(S, C, Skip \mid QP) \ . \tag{1}$$

$$MODE \begin{cases} = Skip, & J(Skip) \le Mean(Skip) \\ \ne Skip, & J(Skip) > Mean(Skip) \end{cases} . \tag{2}$$

There still might be some Skip macroblocks haven't been decided after performing formula (2), so we compare $J(Skip)$ with $J(16\times16)$ in formula (3) as supplement for Skip mode detection. Since this step is performed after predicting possible modes and motion estimation for Inter16×16, it will only save the computation of mode 16×8 and 8×16, if the macroblock is decided as a Skip macroblock in this step.

$$MODE \begin{cases} = Skip, & J(Skip) \le J(16\times16) \\ \ne Skip, & J(Skip) > J(16\times16) \end{cases} . \tag{3}$$

Table 2. Mean values of RD cost for several modes

Sequence	*QP*	*Skip*	*Inter16×16*	*Inter16×8*	*Inter8×16*	*Inter8×8*
Stefan (CIF)	28	11948.5	14222.4	15281.8	14163.5	19849.3
	32	8820.4	12022.1	17047.1	17710.3	26407.8
	36	9173.2	18551.4	24578.6	27155.2	412371.7
Bus (CIF)	28	4914.9	6848.8	8043.6	7507.2	10890.3
	32	6143.2	8606.5	12302.9	12004.0	20109.8
	36	6954.4	22069.1	25448.4	29616.3	36184.6
Paris (CIF)	28	4408.6	7078.5	6753.3	7277.6	11673.8
	32	6199.3	12255.0	13847.2	18649.4	26335.2
	36	6843.6	14580.7	21023.8	24180.2	35779.0
Coastguard (QCIF)	28	5342	7446	8138	8214	9569
	32	7019.3	8150.2	11046.4	13282.0	15286.3
	36	9276.0	11885.3	15341.4	14766.9	22368.0
News (QICF)	28	2896.4	4141.5	7279.7	6516.5	10124.8
	32	3628.5	6323.0	8598.6	9121.2	15517.6
	36	3926.0	8519.2	12405.7	15614.2	24479.4
Mobile (QICF)	28	8561.6	11422.1	11797.0	13016.9	14561.7
	32	10642.7	14312.6	17253.1	17982.2	23939.7
	36	13123.7	18282.3	25501.2	24084.0	34415.8

Most of the Skip macroblocks can be predicted at early stage with the above two steps. For the remaining macroblocks in the frame, we compare the RD cost of Inter16×16 and Inter8×8 for the macroblock by formula (4) to decide which mode subset current macroblock belongs to.

$$MODE \begin{cases} \in SubsetA(Skip, Inter16\times16, Inter16\times8, Inter8\times16), J \le TH_n \\ \in SubsetB(Inter8\times8, Inter8\times4, Inter4\times8, Inter4\times4, Intra16, Intra4), J > TH_n \end{cases} \cdot \quad (4)$$

Due to the similarities between successive video frames, the coding cost of a macroblock in the temporally-preceding frame can be used to predict the coding cost of collocated macroblock in current frame. So we set the TH_n as an adaptive threshold, which can be calculated by using the RD cost of Inter16×16 and Inter8×8 modes in the previous frame. In formula (5), parameter α and β are set as 0.4 and 0.6 respectively according to our statistical analysis.

$$TH_n = \alpha \cdot Mean(Inter16\times16)_{n-1} + \beta \cdot Mean(Inter8\times8)_{n-1} \cdot \quad (5)$$

Furthermore, in the two mode subsets, the occurring probabilities of the modes are not uniform. Usually, Skip and Inter16×16 have a large proportion in Subset A, Inter8×8 is the higher probability mode in Subset B, Intra4 and Intra16 occupy a very small proportion in inter frame coding. Therefore, in the proposed algorithm, we prioritize the modes so that mode with the highest probability will be tried first, followed by the second highest probable mode, and so on. During this process, the computed RD cost will be checked against a content adaptive RD cost threshold, which is obtained from the mean of the already coded macroblock in the previous frame, to decide if the mode decision process should be terminated before trying the remaining modes in the subset. In this way we can avoid trying many unlikely coding modes.

4 Fast Inter Mode Decision

The proposed algorithm is summarized as follows:

Step 1: Calculate the mean of RDcost for every modes (Mean($mode_i$)) in the previous frame, and obtain the frequency of every modes (Freq($mode_i$)). Calculate TH_n by using formula (5).
Step 2: Using the prediction rules to predict which subset the macroblock belongs to. If match Case 1, go to Step 8. If match Case 2, go to Step 5. If match Case 3, go to Step 7. If match Case 4, go to Step 7. If match Case 5, go to Step 3.
Step 3: Perform motion estimation for Inter16×16 mode and calculate RD cost *J(16×16)*.
Step 4: If $J(16\times16) > TH_n$, set MODE∈Subset B, then go to Step 7; Otherwise, if $J(16\times16) \le TH_n$, set MODE∈Subset A, then go to Step 6.
Step 5: Calculate the RD cost of Skip mode *J(Skip)*. If *J(Skip)* ≤ *Mean(Skip)*, go to Step 8. Otherwise, try remaining modes in the subset until all the modes are tried, go to Step 9.
Step 6: Calculate the RD cost of Skip mode *J(Skip)*. If *J(Skip)* ≤ *Mean(Skip)*, go to Step 8. Otherwise, if *J(Skip)* ≤ *J(16×16)*, go to Step 8.
Step 7: Try the modes in the subset in descend order of $Freq(mode_i)$, if $J(mode_i) \le Mean(mode_i)$, set MODE= $mode_i$, go to Step 9; Otherwise, try remaining modes in the subset until all the modes are tried, go to Step 9.

Step 8: Set MODE= Skip. (MODE is the optimal coding mode)
Step 9: Encode current macroblock with MODE.

This algorithm is for macroblocks which not lie in first line or first column. The macroblocks in the first line or first column are coded as normal after performing prediction rules. The flowchart of the algorithm is as follows:

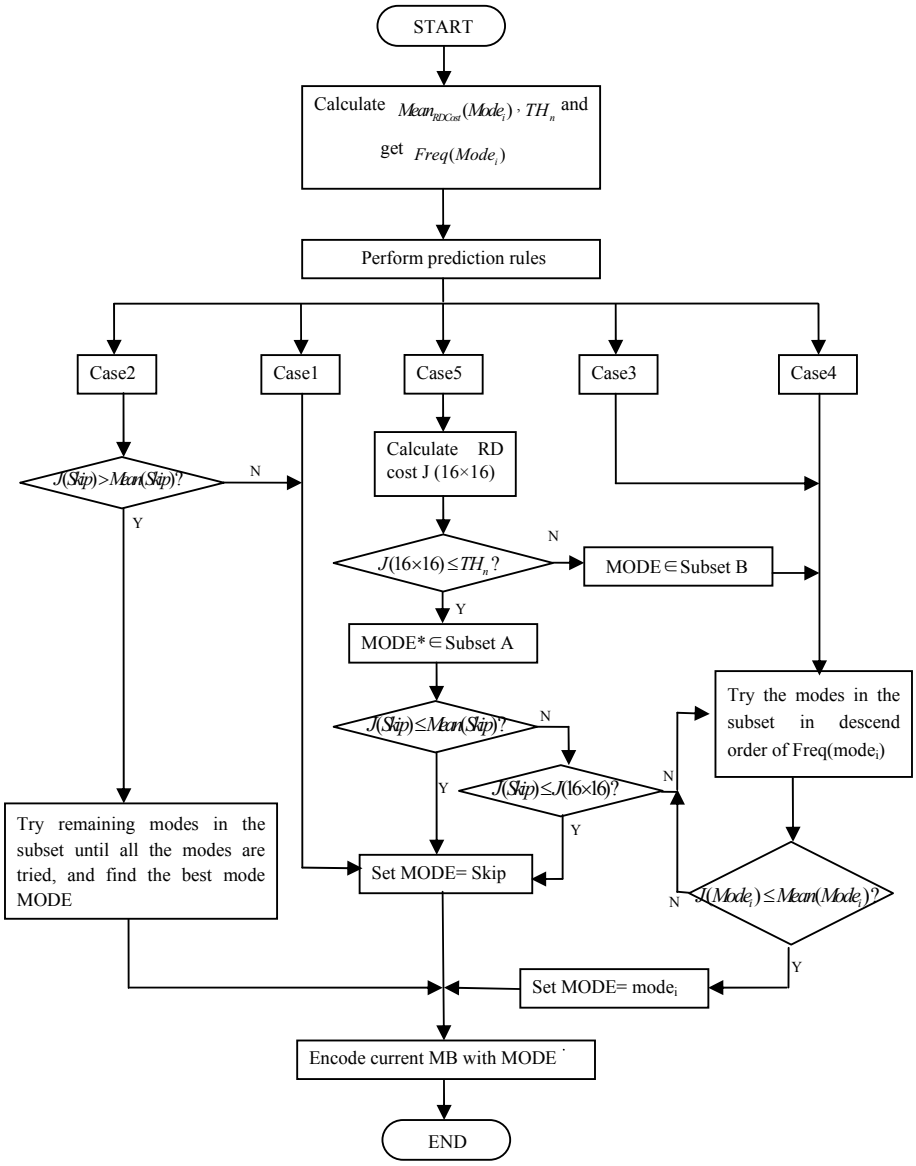

Fig. 6. Flowchart of the algorithm

In order to enhance the robustness of the algorithm, P frames with the original full search mode decision is conducted periodically. The user can set the period themselves according to their requirement.

5 Experimental Results

Our proposed algorithm was implemented into the reference software JM8.6 provided by JVT [6], tested on three standard QCIF sequences (Akiyo, Carphone and Mobile) and three standard CIF sequences (Foreman, Paris and Stefan). Each sequence has 150 frames. And all experimental conditions are described as follows: IPPP sequences, Fast_Motion_Estimation disabled, CAVLC, search range is ±16, RD optimization is off, frame rate is 30 and the platform is Intel Pentium Ⅳi1.73GHz PC with 1024MB memory. The test results are list in Table 3.

$$\Delta SNR = SNR_{proposed} - SNR_{JM\,8.6}$$

$$\Delta BitR = (BitRate_{proposed} - BitRate_{JM\,8.6}) / BitRate_{JM\,8.6} \times 100\%$$

$$\Delta Time = (Time_{proposed} - Time_{JM\,8.6}) / Time_{JM\,8.6} \times 100\%$$

Table 3. Experimental results with several different sequences

Sequence	*QP*	*SNR(Y) (dB)*	*△SNR(Y) (dB)*	*BitR (kbit/s)*	*△BitR (%)*	*Time (sec)*	*△Time (%)*
Foreman (cif)	28	36.91	-0.07	362.98	+0.21	2035.6	-45.2
	32	34.83	-0.04	216.45	-0.03	2053.3	-46.6
	36	32.81	-0.04	140.06	+0.36	2130.1	-43.7
	40	30.77	-0.03	95.28	+0.12	1864.5	-48.3
Paris (cif)	28	35.37	-0.06	529.31	+0.48	1554.3	-56.2
	32	33.06	-0.07	303.85	+0.83	1593.9	-57.1
	36	30.92	-0.05	176.44	-0.16	1424.5	-60.3
	40	28.84	-0.03	111.24	+0.67	1399.1	-58.9
Stefan (cif)	28	35.41	-0.09	1071.44	+0.53	2110.2	-44.5
	32	32.21	-0.13	541.62	-0.32	2008.5	-45.0
	36	29.43	-0.07	299.75	-0.24	2116.7	-47.4
	40	26.67	-0.11	190.28	+1.13	2089.3	-47.1
Akiyo (qcif)	28	38.25	-0.01	26.09	+0.23	198.6	-83.6
	32	36.57	-0.01	16.28	+0.30	207.3	-84.3
	36	35.08	-0.00	11.67	-0.45	173.6	-86.8
	40	33.73	+0.01	9.71	+0.36	192.1	-83.4
Carphone (qcif)	28	37.35	-0.03	85.65	+0.37	282.7	-68.6
	32	34.79	-0.04	50.09	+0.74	247.3	-72.2
	36	32.53	-0.03	32.61	+0.82	255.6	-71.5
	40	30.46	-0.05	21.63	+0.69	231.5	-73.0
Mobile (qcif)	28	33.35	-0.05	419.92	+1.08	638.6	-34.2
	32	30.18	-0.09	209.96	+2.48	706.4	-33.7
	36	27.69	-0.08	111.07	+2.32	660.2	-36.2
	40	25.30	-0.07	65.23	+0.25	562.5	-38.8

In addition, we apply the proposed algorithm to process part of the recorded video of a real forest fire scene from Linfen forest farm in Shanxi province of China. Fig. 7 shows a few images of the video sequence, and Fig. 8 compares SNR decoded by the

Fig. 7. Examples of original (top) and decoded (bottom) frames of the forest fire video sequences. Frame number 200, 250, 300 and 350 of a sequence are shown.

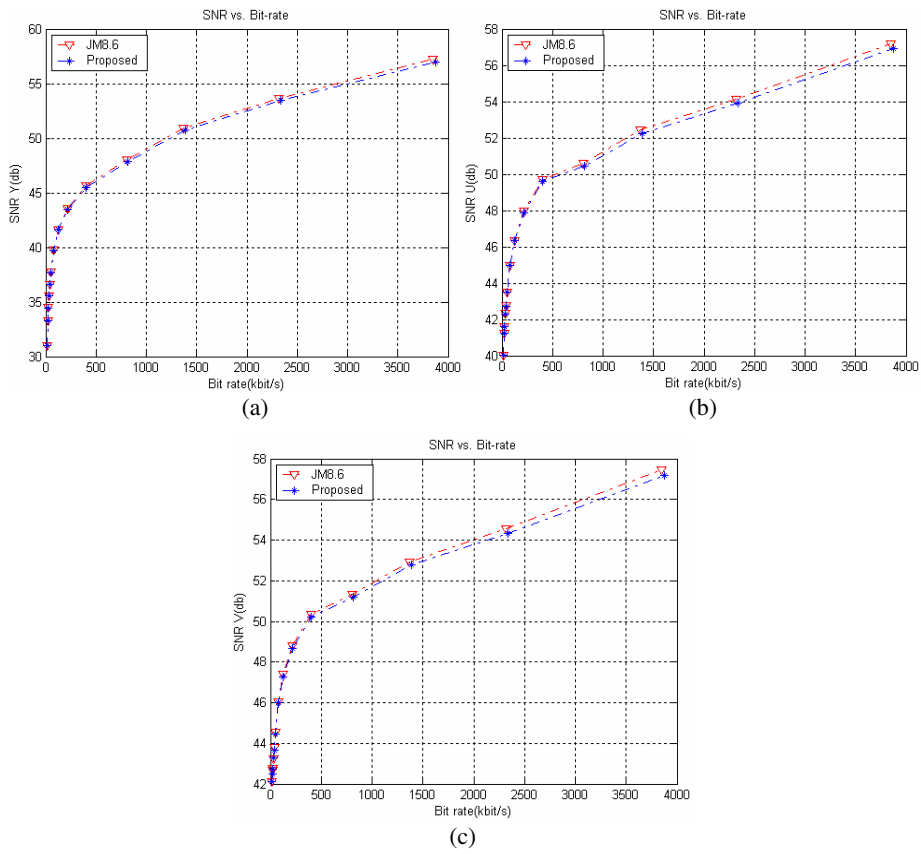

Fig. 8. SNR comparison between JM8.6 and the proposed algorithm (a) SNR Y (b) SNR U (c) SNR V

original JM8.6 baseline profile with that decoded by the proposed algorithm. From the images, we can see that the degradation of the images is hardly noticeable.

6 Conclusion

In this paper, a new fast Inter mode decision algorithm based on statistic and adaptive adjustment has been proposed. By fully exploring the correlation among the modes of the macroblocks and using RD cost to classify the mode category, the algorithm can reduce the computational time significantly with slight loss of PSNR and bit rate increase.

Acknowledgments

This research was supported by the Introduction of Advanced International Forestry Science and Technology Program from the State Forestry Administration, China (Grant No. 2005-4-04) and the National College Students Innovative Research Program from the Ministry of Education of China (Grant No. GCS08016).

References

1. Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264|ISO/IEC 14496-10 AVC), JVT-G050r1, Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG (May 2003)
2. La, B., Jeong, J., Choe, Y.: Selective intra-prediction Skip Algorithm for inter-frame coding in H.264/AVC. IEICE Electronics Express 6(2), 71–76 (2009)
3. Kannangara, C.S., Richardson, I.E.G., Bystrom, M., Solera, J., Zhao, Y., MacLennan, A., Cooney, R.: Low Complexity Skip Prediction for H.264 through Lagrangian Cost Estimation. IEEE Transactions on Circuits and Systems for Video Technology 16(2), 202–208 (2006)
4. Kim, B.-G., Song, S.-K.: Enhanced Inter Mode Decision Based on Contextual Prediction for P-Slices in H.264/AVC Video Coding. ETRI Journal 28(4), 425–434 (2006)
5. Na, T., Lee, Y., Lim, J., Joo, Y., Kim, K., Byun, J., Kim, M.: A Fast Macroblock Mode Decision Scheme Using ROI-Based Coding of H.264|MPEG-4 Part 10 AVC for Mobile Video Telephony Application. In: Proc. SPIE, vol. 7073, pp. 707301 (2008) DOI: 10.1117/12.795614
6. H. 264/AVC reference software JM8.6, `http://iphome.hi.de/suehring/tml/download`
7. Nieto, M., Salgado, L., Cabrera, J., García, N.: Fast mode decision on H.264/AVC baseline profile for real-time performance. Journal of Real-Time Image Processing 3(1-2), 61–75 (2008)
8. Grecos, C., Yang, M.: Fast Mode Prediction for the Baseline and Main Profiles in the H.264 Video Coding Standard. IEEE Trans. Multimedia 8(6), 1125–1134 (2006)