Efficient Probability Based Macroblock Mode Selection In H.264/AVC

Yali Liu^{a} , Kun Tang^{b} and Huijuan Cui^{c}

State Key Lab on Microwave & Digital, Communications
Department of Electronic Engineering, Tsinghua University, 100084, Beijing, P.R. China

ABSTRACT

To get high compress efficiency, the latest H.264/AVC international video coding standard introduced many more advanced tools than previously used standards, such as using sophisticated prediction and rate-distortion (RD) mode selection. Although these new coding tools can significantly improve video coding efficiency, the improvement in performance comes at substantially higher computational complexity as a result of more complicated mode decision. To reduce the complexity of H.264 encoding, various fast mode decision algorithms have been proposed. Many of these algorithms take advantage of motion and video content classification so as to reduce the number of modes tried during mode decision. When used for video sequences with high motion or detailed visual information, video coding efficiency can be significantly compromised as a result of imperfect classification. In this paper, an efficient Probability Based Macroblock Mode Selection (PBMMS) algorithm is proposed. Each MB is predicted to different probability MB based on spatial and special correlation, which then leads to 50-60% computation savings and some compression performance improvements over existing H.264 fast mode decision algorithms.

Keywords: f ast mode selection, rate distortion, classify, computation savings

1. INTRODUCTION

Due to its high efficiency, H.264/AVC[1] video coding standard has gained more and more attentions recently. H.264 improves video coding efficiency by using sophisticated prediction and rate-distortion (RD) mode selection.[2] As a result, encoder complexity of H.264 is very high, which makes it difficult to use for practical applications on resource-constrained applications.

Many fast mode-selection algorithms[3–6] have been developed to reduce the encoder complexity of H.264. Some of them achieve this by making use of texture and edge information, such as the computation of the Sobel edge operators for every macroblock(MB)[7–9] to find whether a MB is homogeneous or the edge map of a whole frame. Even though average system performance can be improved with such algorithms, the worst-case performance for sequences for which such additional operations are not effective in predicting the proper encoding modes to use is actually lower than when the algorithms are not used.

The other typical approach classifies the block types into different groups. By predicting which group may contain the best mode, one can omit the other group. One state-of-the-art fast mode decision algorithm (FMMS) in[10] categorizes each MB into either complex motion or simple motion by the fuzzy classifier which exploits the spatial correlation among MBs. Then various early termination algorithms are used which lead to significant computation reduction as only a small subset of possible modes needs to be searched. The shortcoming of the algorithm is that when used for video sequences with high motion or detailed visual information, video coding efficiency can be significantly compromised as a result of imperfect classification.

Another new method (FMDP)[11] is to speed up the mode decision algorithm by exploring the inter and intra correlation between neighboring image blocks. Statistics show that MBs of the same mode tend to take place

---

Further author information: (Send correspondence to Yali Liu)
Yali Liu: E-mail: liuyl02@mails.tsinghua.edu.cn, Telephone: 86-10-62781420
Kun Tang: E-mail: tangkun@tsinghua.edu.cn
Huijuan Cui: cuihj@ee.tsinghua.edu.cn

at the neighboring place, so this method exploits the mode distribution and the mode similarity of neighboring MBs to reduce the computational load. Although it requires very little computation time to predict best mode, the probability that all the neighboring MBs have the same mode is very small, which leads to the limitation of the reduction of the computation and bad performance loss.

The insight behind the algorithm in this paper is that statistics from different sequences and resolutions show that although there are 7 different INTER and 2 INTRA modes to encode every Inter MB in H.264, the probabilistic distribution of different modes after RD optimized mode decision is not uniform. If one could correctly estimate the probabilities of different modes for the MB to be encoded, the performance of the encoder could be improved.

Motivated by the above observation, we propose an efficient Probability Based Macroblock Mode Selection (PBMMS) algorithm. By predicted to different probability subset based on spatial and special correlation, some MBs are then performed by motion content classification to a specific preferred mode search order coupled with distinct early termination scheme. Simulation shows that it can lead to 50-60% computation savings and some compression performance improvements over existing H.264 fast mode decision algorithms.

The rest of paper is organized as follows: Section II give an overview of the mode decision algorithm specified in the H.264 standard. Section III contains a detailed description of the algorithm along with a flowchart. Simulation results are presented in Section IV and concluding remarks are given in Section V.

2. OVERVIEW OF MODE SELECTION IN H.264/AVC

2.1. Variable block sizes

To achieve the highest coding efficiency, H.264/AVC uses tree-structured hierarchical macroblock partitions. In the Inter mode, every macorblock can be divided into partitions of size 16×16, 16×8, 8×16 or 8×8 pixels. 8×8 partitions can be further partitioned into sub-partitions of sizes 8×8, 8×4, 4×8 or 4×4 pixels, as shown in Figure 1. In this paper, we use INTER16 × 16, INTER16 × 8, INTER8 × 16, INTER8 × 8, INTER8 × 4, INTER4 × 8, and INTER4 × 4 to denote these modes and partitions. For a P frame, each macroblock need to be tried all possible mode: SKIP, INTER16 × 16, INTER16 × 8, INTER8 × 16, INTER8 × 8, INTER8 × 4, INTER4 × 8, INTER4 × 4, INTRA16 × 16 and INTRA4 × 4 to get the optimal one. The SKIP mode represents the case where the block size is 16×16 and the macroblock has zero motion and residual data to be encoded. So there is no motion search required and is has the lowest computational complexity. Except the SKIP mode and the intra modes, all the other inter modes require motion estimation, which has most computation in all the encoder process.

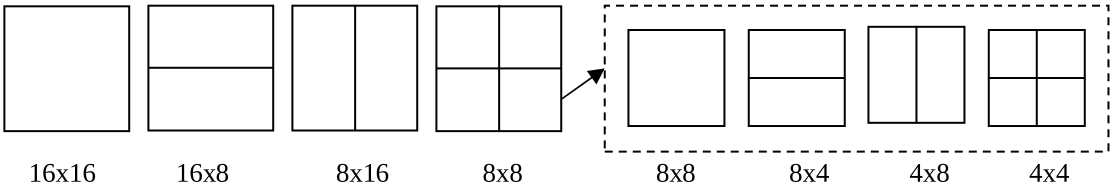

Figure 1. Variable block sizes defined in H.264

2.2. Rate Distortion Optimization (RDO) for H.264/AVC

A rate-distortion optimization technique[2] is adopted in H.264 to achieve an optimal tradeoff between picture quality and data rate. The mode decision is accomplished by comparing the rate distortion cost of each possible mode, and the mode with lowest cost is selected as the best mode.

According to Lagrange optimization technique, RD cost is typically computed as the following:

$$\mathbf{J} = D + \lambda \times R, \tag{1}$$

When(1)is used for mode decision with RDO, D is the SSD(Sum of Squared Differences) between the original frame and the reconstructed signal, R represents the bits for coding the MB header, the motion information and the transformed coefficients, and λ is a Lagrange multiplier λ_{mode}. Especially, forward and inverse integer transform, quantization, and entropy coding are performed repetitively on all the modes to achieve an optimal tradeoff between picture quality and data rate. So the computation load of the mode decision process is too high to be accepted, particularly in the practical application. Therefore fast mode decision algorithms attract more and more attentions.

When RDO is turned off, (1)is still used but with the different parameter. Here D is the SAD (Sum of absolute Differences) between the current frame and the reference signal, and R represents the bits for coding the motion vector and reverence frame. The reference signal is the best match one obtained by motion estimation, where the reference block will be searched on integer-pel positions as well as 1/2 and 1/4 pel position with different block sizes.

3. EFFICIENT PROBABILITY BASED MACROBLOCK MODE SELECTION

In this section, we describe the ideas behind our method for reducing the computational complexity of mode decision.

3.1. Uneven Distributed Probability Of Different Modes

Firstly, we observe that there is some relationship between the mode and video content. For example, from some experiments, it is found that generally the SKIP mode or INTER16 × 16 is the best mode for MBs in the background or the smooth regions of the image. On the other hand, INTER8 × 8 or the further partition modes is the favorite choice for MBs which are in the moving regions or on the boundary of objects.

Furthermore, the probabilities for these modes and partition sizes are not uniform and it seems that the same kind of MBs tend to cluster together. Tables 1 contains the probabilities obtained for some standard test sequences in CIF size using the JVT JM61e reference software with baseline profile. Here it should be noticed that the "same kind" does not mean the "same", but that some modes in a subset. In other words, the chance that the neighboring MBs have the similar mode should be higher than the chance that the neighboring ones have the diverse mode. Here let A, B and C denote the left, top-left, top of the current MB in the current frame respectively, and let D represent the membership of the co-located MB in the previous frame, referring to Figure 2. Table 2 analyzes the mode correlation of neighboring MBs with the mode of current MB obtained for some standard test sequences in CIF size.

Table 1. Uneven Distribution of Different Modes

Sequence	*For Macroblock %*							*For Block of Inter*8×8 *%*			
	Skip	$Inter_{16x16}$	$Inter_{16x8}$	$Inter_{8x16}$	$Inter_{8x8}$	$Intra_{16x16}$	$Intra_{4x4}$	$Inter_{8x8}$	$Inter_{8x4}$	$Inter_{4x8}$	$Inter_{4x4}$
Tempete	13.02	36.74	9.29	8.51	30.37	1.33	0.735	48.63	20.19	18.88	12.30
Mobile	4.19	33.93	8.21	7.91	42.24	0.08	0.42	49.16	18.79	21.16	10.89
Mother	68.52	18.01	4.55	5.04	3.06	0.40	0.42	65.14	14.40	17.54	2.91
Paris	57.65	16.26	3.51	4.00	17.93	0.09	0.54	49.42	14.99	19.73	15.86
Container	70.97	20.95	2.40	2.42	2.18	0	1.06	62.53	14.65	16.44	6.38
Salesman	75.91	12.65	3.52	2.84	5.96	0.27	0.24	59.86	18.20	16.73	5.20

3.2. Probability Based Macroblock Mode Selection

Based on the observations of Tables 2, in the proposed algorithm, we classify MBs into two classes. MBs that are best encoded with SKIP or INTER16 × 16 or INTER8 × 8 modes are the High Probability MBs (HMBs), while MBs encoded with the other modes are the Low Probability MBs (LMBs). Here the subset of INTER8 × 8 which doesn't include the further partition modes and SKIP, INTER8 × 8 is called TRUE HMBs.

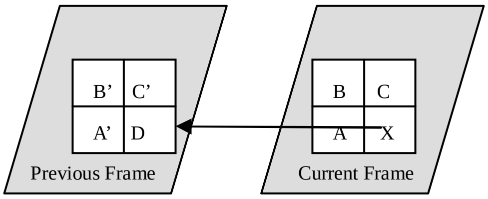

Figure 2. Neibouring MBs' Positions to the Current MB

Table 2. Mode Correlation of Neighboring MBs

Sequence	*Format*	*Case 1%*	*Case 2%*	*Case 3%*	*Case 4%*
Tempete	*IPPPP*	49.33	84.12	51.83	83.78
	IPBPB	56.28	88.09	57.94	87.48
Mobile	*IPPPP*	54.78	86.29	57.01	86.23
	IPBPB	58.20	88.68	59.68	88.31
Mother&Daughter	*IPPPP*	67.41	92.58	69.69	92.33
	IPBPB	72.29	94.90	73.63	94.60
Paris	*IPPPP*	70.08	93.68	72.13	93.60
	IPBPB	69.85	93.68	71.06	93.45
Container	*IPPPP*	76.50	96.17	78.41	96.15
	IPBPB	80.46	97.45	81.48	97.36
Salesman	*IPPPP*	73.63	95.66	75.62	95.57
	IPBPB	76.27	96.62	77.51	96.38

[*] Case1: MBs A, B, C and D all lie in the subset of $\{$SKIP, INTER16 × 16, INTER8 × 8$\}$, or none lies in it.

[†] Case2: MB X lies in the same subset as MBs A, B, C, and D when Case 1 is true.

[‡] Case3: MBs A, B and C all lie in the subset of $\{$SKIP, INTER16 × 16, INTER8 × 8$\}$, or none lies in it.

[§] Case4: MB X lies in the same subset as MBs A, B and C when Case 3 is true.

To calculate the current MB probability based on the neighboring ones(defined in the previous), then classify it to the different classes, we assume that the current MB mode is only connected with the neighboring A, B, C, D where mode probabilities are $P_i(x)$ $i = A, B, C, D$, and omit the influence of others. Our prediction of the probability of current MB is set to be the weighted sum of A, B, C and D. In our experiments we set $W_A = 0.3$, $W_B = 0.2$, $W_C = 0.2$, and $W_D = 0.3$. The weight for the MB to the left of the current MB is higher because motion in horizontal direction is usually more noticeable and the horizontal resolution of the video sequence is usually higher. In addition, usually the probability of co-located MB has the same mode with current MB is higher. Let P denote the probability of the current MB, then we have

$$P(x) = \sum_{i=A,B,C,D} W_i \cdot P_i(x) \tag{2}$$

A P_{mode} cut set with $P_{mode} \in [0, 1]$ is used to determine the category which current MB belongs.[12] Because we can get the probability of all modes, which are computed dynamically every frame, we let this cut set equal to the probability of LMB. So we have

$$P_{mode} = 1 - \{P_{SKIP} + P_{INTER16\times16} + P_{INTER8\times8}\} \tag{3}$$

Table 3 reflects the classification accuracy of 6 CIF sequences with 100 frames. In the tables, *Correct Ratio* is the probability of the MB classification to predict the same optimal encoding mode obtained from exhaustive mode selection, *HMBErrRatio* reflects the probability for HMBs to be mistakenly categorized as LMBs, while the *LMBErrRatio* reflects the probability for LMBs to be mistakenly categorized as HMBs. Compared with the classification accurate ratio of FMMS in,[12] our algorithm shows the robust over all kinds of sequences with different motion and other features.

Table 3. Macroblock Probability Classification Accuracy

Sequence	*Format*	*HMB Err Ratio %*	*LMB Err Ratio%*	*Correct Ratio %*
Tempete	*IPPPP*	17.47	0.18	82.33
	IPBPB	16.20	0.21	83.58
Mobile	*IPPPP*	13.51	0.66	86.42
	IPBPB	11.33	0.28	88.38
Mother&Daughter	*IPPPP*	10.05	0.58	89.89
	IPBPB	8.56	0.01	91.42
Paris	*IPPPP*	7.79	0.01	92.19
	IPBPB	6.95	0.27	92.78
Container	*IPPPP*	17.47	0.18	82.33
	IPBPB	16.20	0.21	83.58
Salesman	*IPPPP*	5.11	0.005	94.89
	IPBPB	4.50	0.08	95.42

A MB is determined to be an LMB when the weighted sum is lower than P_{mode} , and if the P_{mode} is higher than the minimum of $\{P_{SKIP}, P_{INTER16\times16}, P_{INTER8\times8}\}$, the MB is determined to be a true HMB. Otherwise, we need to further classify its motion character. Here a motion classifier[10] is continue used to determine if the MB contains complex motion information or simple motion information. By combining two types of classifiers, each MB can be efficiently categorized to different mode and motion search paths, which significantly reduces encoder complexity of H.264 for all types of content. Our fast mode decision algorithm consists of the following steps:

Step1: If the MB is in the first row or column of a frame, test all possible modes, select the best one, then exit.

Step2: Each MB is categorized by a probability classifier. If the predict mode is included in the HMBs, go to Step 4. Otherwise, go to Step 3.

Step3: Check the mode of INTER8 $\times$ 16 and INTER16 $\times$ 8. Go to Step 9.

Step4: For B picture, calculate the RD cost of direct mode. If it is lower than the threshold, which is defined as the minimum of neighbouring MBs, skip all other modes and go to step 11. Otherwise, If the predict mode is included in the TRUE HMBs, go to Step 10, otherwise go to Step 5.

Step5: To categorize the MB with a motion classifier. If it has complex motion content, go to step 6. Otherwise, go to Step 8.

Step6: Check mode INTER8 $\times$ 8, INTER8 $\times$ 4, INTER4 $\times$ 8, INTER4 $\times$ 4. If there are more than two sub-macroblock modes are not INTER8 $\times$ 8, go to step 9. Otherwise, go to Step 7.

Step7: Check mode INTER16$\times$16, INTER16$\times$8 and INTER8$\times$16. If any mode cost is more than INTER8$\times$8 or the three modes have been tried, go Step 11.

Step8: Check mode INTER16 $\times$ 16 and INTER16 $\times$ 8, if $cost_{16\times16} < cost_{16\times8}$, go to Step 9. Otherwise, check all the other Inter modes.

Step9: Check INTRA16 $\times$ 16 and INTRA4 $\times$ 4.

Step10: Check INTER16 × 16 and INTER8 × 8.

Step11: Record the best MB mode and the minimum RD cost.

A flow chart of the algorithm is given in Figure 3.

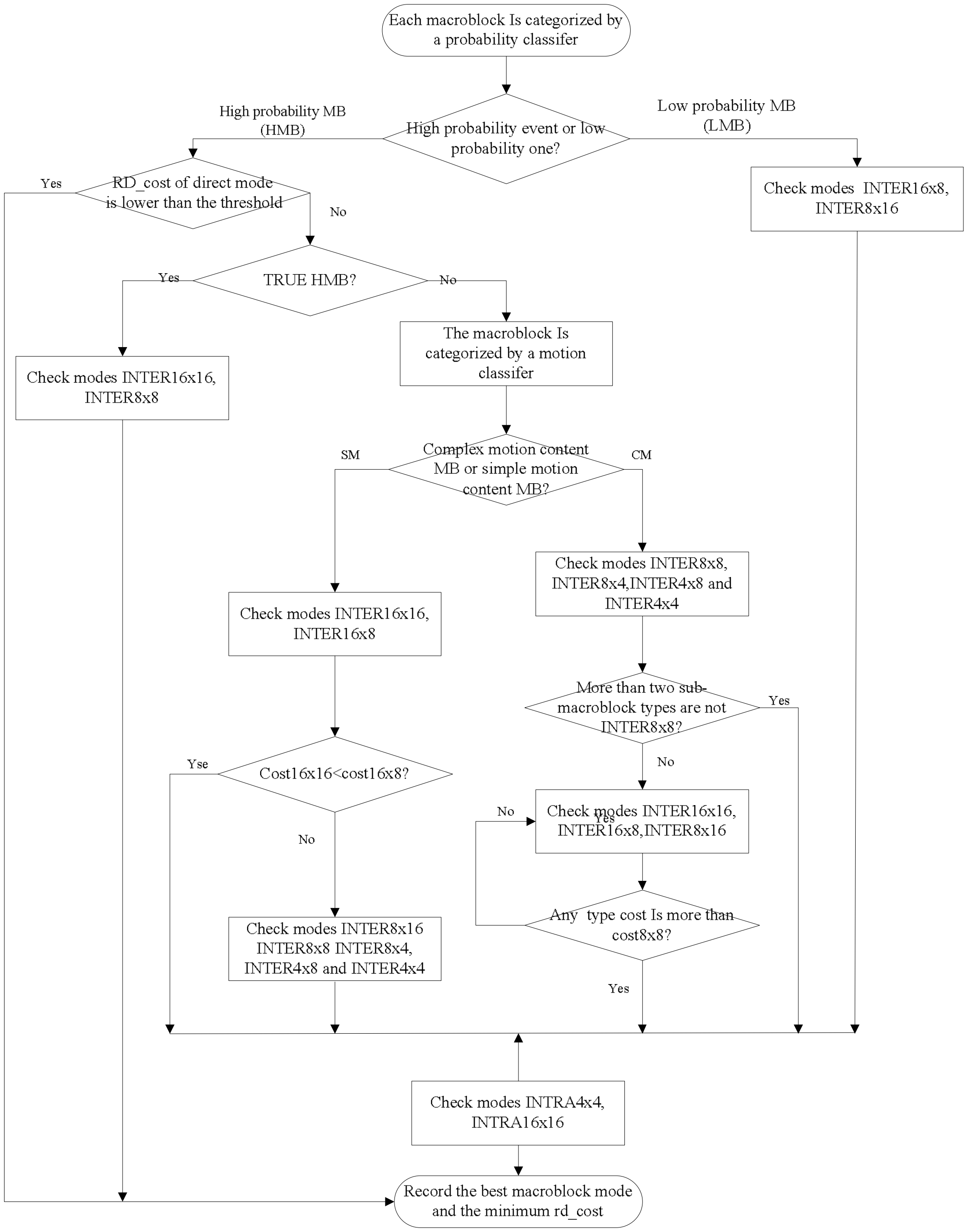

Figure 3. Probability Based Macroblock Mode Selection(PBMMS) process

4. SIMULATION RESULTS

CIF (352 × 288) and QCIF (176 × 144) format video sequences with different motion content were employed in simulation within the H.264/AVC JM61e framework. One CIF format and one QCIF format *Stefan* were encoded to show the frame size influence.

For each sequence, 101 frames were encoded. A single reference frame was used in the experiments. As our algorithm was designed for mode selection for P and B frames, only the frame at the beginning of each sequence was encoded as an Intra frame. The encoding parameters were as follows: The quantization QP factor was set to 28 for I and P frames, and 30 for B frame, search window [-16, +16].

We first compared the proposed algorithm with exhaustive macroblock mode selection (EMMS) and two typical algorithms referred previously with baseline profile. The rate distortion performance was measured by average peak signal-to-noise ratio (PSNR) of the luminance component and the average bit rate. The complexity was measured by the total number of modes tried for Inter motion compensation and Intra prediction for all 4 × 4 blocks in the entire sequence.

The simulation results are given in Table 4 and Fig 4, from which we can easily see that the proposed algorithm significantly reduced the number of modes tried during the RD optimized encoding with virtually no loss of RD compression performance. As a matter of fact, with less than 0.04dB maximum PSNR loss and a bitrate increase of at most 1%, the algorithm reduced the Inter motion compensation by 60-70% and Intra prediction by 40-50% for the test sequences.

Then we compared the proposed algorithm with the fast mode decision algorithm of[10] over a wide range of coding conditions. Figure 5(a) and Figure 5(b) plot the rate distortion curve of *Mobile* and the corresponding search saving against the PSNR in the main profile. From these we can find that the proposed algorithm can further reduce 50% encoding complexity in a broad range of bit rates and image qualities.

Table 4. Comparison of overall rate distortion performance of 101 frames of different sequences

Sequence	*Algorithm*	*PSNRY(dB)*	*Bitrate(kbit/s)*	*Inter Motion Search times*	*Intra Motion Search times*
Mother&Daughter	*EMMS*	38.90	143.02	4390848	1267200
	FMMS	38.88(-0.02)	144.40(+1.0%)	2278512(-48.11%)	1226720 (-3.19%)
	FMDP	38.72(-0.18)	143.13(+0.1%)	2804608(-36.13%)	760992(-37.97%)
	PBMMS	38.86(-0.04)	145.12(+1.4%)	1536432(-65.01%)	607456(-52.06%)
Mobile	*EMMS*	33.93	2056.27	4390848	1267200
	FMMS	33.92(-0.01)	2074.80(+0.9%)	3454960(-21.31%)	1267200 (-0.00%)
	FMDP	33.91(-0.02)	2107.67(+2.5%)	4041472(-7.9%)	1136000(-10.35%)
	PBMMS	33.88(-0.02)	2080.94(+1.2%)	1486988(-66.13%)	731104(-40.40%)
Container	*EMMS*	35.88	219.68	4390848	1267200
	FMMS	35.86(-0.02)	221.22(+0.7%)	2106000(-52.03%)	1267200(-0.00%)
	FMDP	35.79(-0.08)	218.05(-0.07%)	2767344(-36.97%)	737568(-41.80%)
	PBMMS	35.86(-0.02)	220.63(+0.4%)	1414352(-67.79%)	607456(-52.06%)
Tempete	*EMMS*	34.45	1273.90	4390848	1267200
	FMMS	34.43(-0.02)	1274.26(+0.2%)	3184624(-27.47%)	1267200(-0.00%)
	FMDP	34.23(-0.22)	1261.23(-0.1%)	3787968(-13.73%)	1063456(-16.08%)
	PBMMS	34.43(-0.02)	1280.33(+0.5%)	1466736(-66.60%)	793696(-37.37%)
Stefan.cif	*EMMS*	35.35	1293.04	4390848	1267200
	FMMS	35.34(-0.01)	1286.06(-0.5%)	3201520(-27.09%)	1267200(-0.00%)
	FMDP	35.15(-0.20)	1282.17(-0.8%)	3423216(-22.03%)	957984(-24.40%)
	PBMMS	35.36(+0.01)	1311.87(+1.4%)	2619808(-40.33%)	805568(-36.43%)
Stefan.qcif	*EMMS*	34.21	398.46	1097712	316800
	FMMS	34.19(-0.02)	398.21(-0.06%)	811248(-26.10%)	316800(-0.00%)
	FMDP	33.78(-0.43)	393.64(-1.2%)	825088(-24.84%)	232896(-26.48%)
	PBMMS	34.21(-0.00)	402.28(+0.9%)	664280(-39.49%)	203360(-35.80%)

5. CONCLUSION

In this paper, a fast macroblock mode selection algorithm in H.264/AVC is proposed. By predicted to the different probability subset and then combining two types of classifiers, each macroblock can be efficiently categorized to different motion search paths with corresponding early termination scheme. Extensive simulation results ran with a number of different sequences over a big range of bitrates show that the algorithm can dramatically reduce the number of motion search times while maintaining similar rate distortion performance. This algorithm is very efficient in real-time applications of H.264/AVC video coding.

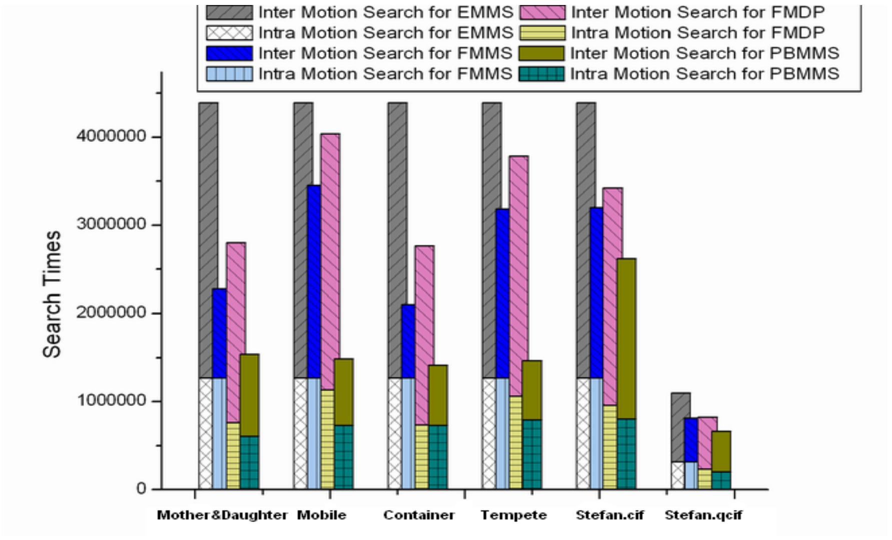

Figure 4. Column graph of the calculated 4x4 block motion search times in EMMS, FMMS, FMDP and PBMMS

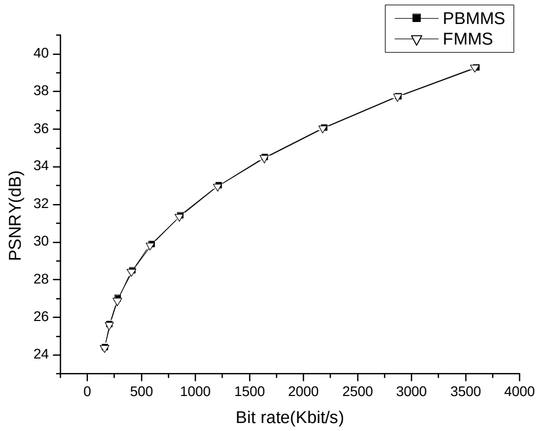

Figure 5. (a) Comparison of rate distortion performance

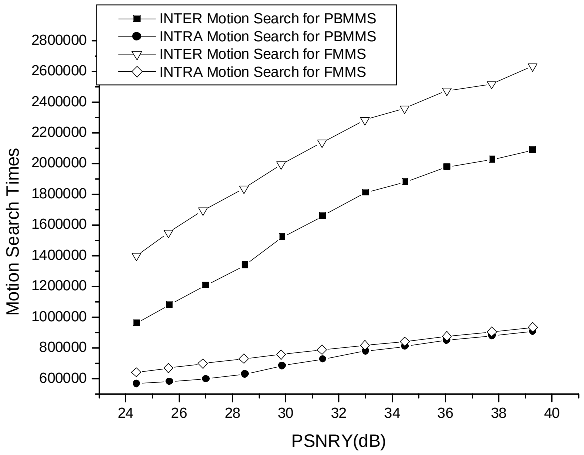

Figure 5. (b) Percentage of motion search

REFERENCES

1. ITU-T Rec.H.264, ISO/IEC 4496-10 AVC, “Advanced Video Coding Final Draft International Standard”, JVT-G050r1, Geneva, Switzerland, May 2003.
2. T. Wiegand, H.Schwarz, and A.Joch, “Rate constrained coder control and comparison of video coding standards,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, pp. 688-703, July 2003.

3. Turaga, D.S.,Tsuhan Chen, "Estimation and Mode Decision for Spatially Correlated Motion Sequences" *Circuits and Systems for Video Technology, IEEE Transactions on* , Vol. 11 , pp. 1098-1107 , Oct. 2001.
4. Turaga, D.S., Tsuhan Chen, "Classification Based Mode Decisions for Video over Networks" *Multimedia, IEEE Transactions on* , Vol. 3 , pp. 41-52, March 2001.
5. J. Lee and B. Jeon, "Fast Mode Decision for H.264 with Variable Motion Block Sizes", *International Symposium on Computer and Information Sciences (ISCIS)*, vol. 2869, pp. 723-730 Nov. 2003.
6. J. Lee and B. Jeon, "Pruned Mode Decision based on Variable Block Sizes Motion Compensation for H.264", *Multimedia Interactive Protocols and Systems (MIPS)*, vol. 2899, pp. 410-418, Nov. 2003.
7. J. Lee and B. Jeon "Fast mode decision for H.264", *Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG(ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6)10th Meeting*JVT-J033, Thailand, 8-12 December, 2003.
8. K. P. Lim, S. Wu,"Fast INTER Mode Selection", *Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG(ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6) 9th Meeting* JVT-I020, San Diego, 2-5 September, 2003.
9. Feng Pan, Xiao Lin,"Fast Mode Decision for Intra Prediction", *Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG(ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6)7th Meeting*JVT-G013, Thailand, 7-14 March, 2003.
10. Ming Yang, Wensheng Wang, "Fast Macroblock Mode Selection Based On Motion Content Classification in H.264/AVC ", *IEEE International Conference on Image Processing (ICIP)*, pp. 741-744, October 24-27, 2004 .
11. Che-Yu Chang, Chia-Ho Pan2, Homer Chen "Fast Mode Decision For P-Frames In H.264", *Picture Coding Symposium (PCS)*, San Francisco, California, USA, December 2004.
12. Ludmila I. Kuncheva, Fuzzy classifier design, chapter 5, Physica-Verlag Heidelberg, New York, 2000.