2004 International Conference on Image Processing (ICIP)

FAST MACROBLOCK MODE SELECTION BASED ON MOTION CONTENT CLASSIFICATION IN H.264/AVC

Ming Yang, Wensheng Wang

State Key Lab on Microwave & Digital Communications
Dept. of Electronic Engineering, Tsinghua University
Beijing, 100084, P.R.China.
Email:{yangm01,wangwensheng97}@mails.tsinghua.edu.cn

ABSTRACT

In H.264/AVC, new coding techniques introduce many optional macroblock encoding modes. Exhaustive search over all possible modes can achieve optimal coding efficiency but is highly computationally expensive. In this paper, a fast macroblock mode selection (FMMS) algorithm based on the motion content classification is proposed. Each macroblock is categorized into complex motion or simple motion contents by a fuzzy classifier, then different mode search orders with distinct early termination schemes are employed according to the classification. The proposed method can be readily incorporated with fast motion estimation algorithms, and simulation results show that it can further save 40%-70% of the block distortion calculations on the basis of conventional fast motion estimation algorithms, while maintaining similar rate distortion performance.

1. INTRODUCTION

The incoming video coding standard H.264/AVC [1] introduces many advanced coding techniques such as multiple reference picture prediction, tree-structured motion estimation [1], and generalized B-picture [2], which enhance the coding efficiency significantly while dramatically increase the computational complexity. The new features result in much more macroblock(MB) encoding modes than in the standards MPEG-II, H.263[3]. The available MB modes in H.264/AVC include two INTRA coding modes such as the INTRA-4$\times$4 and INTRA-16$\times$16, and various motion compensated coding modes such as the SKIP/DIRECT, INTER-16 $\times$ 16, INTER-16 $\times$ 8, INTER-8 $\times$ 16, and INTER-8 $\times$ 8, while only 4 MB modes are available in H.263 [4]. In the INTER-8$\times$8 mode, each 8$\times$8 sub-macroblock can be further independently coded with sub-macroblock types such as the INTER-8$\times$8, INTER-8$\times$4, INTER-4$\times$8, and INTER-4$\times$4. Exhaustively examine all modes and select the one with the minimum rate distortion cost (rd_cost) [4] as the best mode will achieve optimal coding performance, however, the expensive computational cost obstructs it from practical use.

Different MB modes are suitable for different kinds of motion contents. For a macroblock with rectilinear motion of uniform velocity, the rd_cost of the INTER-16$\times$16 mode with one motion vector is much likely lower than the rd_cost of the INTER-4 $\times$ 4 mode with 16 motion vectors. Thus, if the motion content of the current macroblock can be estimated, the mode set to be searched can be restricted with early termination techniques.

Motivated by the above observation, we propose an effective fast macroblock mode selection algorithm. Each macroblock is categorized into two classes, complex motion macroblock (CMB) and simple motion macroblock (SMB) with a fuzzy classifier [5] that exploits the spatial correlation among MBs. Afterwards, different mode search orders are performed with distinct early termination schemes according to the classification. Thus substantial computations can be reduced due to the limited mode subset to be searched.

The problem of multi-frame selection in H.264/AVC has been addressed in [6], while the focus of the FMMS is the MB mode selection, so in the experiments one reference frame is employed for P-pictures and two for B-pictures.

The paper is organized as follows. In Section 2, the fuzzy classifier to categorize macroblocks in terms of the motion contents is described. In Section 3, the early termination schemes and the entire macroblock mode selection process are presented. Simulation results are presented in Section 4 and concluding remarks are given in Section 5.

2. MACROBLOCK CLASSIFICATION

The best macroblock mode with the minimum rd_cost reflects the characteristic of its motion content to some extent. If the INTER-16 $\times$ 16 mode provides the minimum rd_cost among all modes, this MB generally have a constant

This work was supported by the National Natural Science Foundation of China (60272020.) in part by the National Fundamental Research Program (G1998030406.)

0-7803-8554-3/04/$20.00 ©2004 IEEE.

movement in one direction. While if the INTER-8×8 MB mode with the INTER-4×4 sub-macroblock type is the best mode, it may have a very complex motion that its parts have different moving directions or different velocities, e.g rotating motion. The SKIP mode in P-pictures and the DIRECT mode in B-pictures imply a simple and regular motion that can be derived from the previous encoded blocks. In addition, the block motion fields of a real world image sequence have a strong spatial correlation, so the best modes of adjacent MBs may provide a coarse estimation of current MB.

Based on these observations, the macroblocks can be categorized according to the degree of their motion content complexities. However, the motion content complexity is rather a fuzzy concept than a deterministic one, therefore a fuzzy classifier exploiting the spatial correlation is employed to estimate the motion complexity of current macroblock from the best modes of previous encoded macroblocks. The reason why temporal correlation is not exploited is that the successive frames may have different statistic character. For example, B-pictures and P-pictures have quite different macroblock mode distributions.

All macroblocks are divided into two categories 1) SMBs: the best mode of which is one of SKIP/DIRECT, INTER-16×16, INTER-16×8 and INTER-8×16 modes, which implies quiescence or simple and regular motion in one or two directions; 2) CMBs: the best mode of which is the INTER8×8, INTRA-4×4 or INTRA-16×16, which indicates intricate motion or no adequate match block ever found in the motion estimation. A membership is assigned to every MB mode to indicate the probability that it belongs to the CMBs, the assignment are listed in Table 1. Let μ_L, μ_T and μ_R represent the membership of the left, top, and top-right MB, respectively; w_L, w_T and w_R represent the corresponding weight coefficients. The membership of current MB μ_C is the weighted sum of μ_L, μ_T, and μ_R.

$$\mu_C = [w_L, w_T, w_R][\mu_L, \mu_T, \mu_R]^T$$

A μ_{motion}-cut set with $\mu_{motion} \in [0, 1]$ is used to determine the category to which current macroblock belongs. Two bias of real world movement are taken into consideration to choose the weighted coefficients and μ_{motion}. First, if a macroblock belongs to the CMBs, the conditional probability that its neighboring blocks also belong to the CMBs is larger than the conditional probability when a MB is quiescent or simple motion and its neighborhood are also simple motion, so μ_{motion} is set to 0.4. Second, because the movements in horizontal direction are more frequent than that in vertical direction, w_L should be slightly larger than other two coefficients, $w_L = 3/7$, $w_T = 2/7$, and $w_R = 2/7$.

Table 2 reflects the classification accuracy of 6 CIF sequences with 100 frames. Correct Ratio is the ratio of which the MB classification accords with the result obtained from the exhaustive mode selection, SMB Err Ratio is the ratio of which SMBs are mistakenly categorized into the CMBs, the opposite ratio is called CMB Err Ratio.

Table 1. Membership assignment of macroblock modes

MB Mode	Membership	MB Mode	Membership
SKIP	0.0	DIRECT	0.0
INTER-16×16	0.1	INTER-16×8	0.2
INTER-8×16	0.2	INTER-8×8	0.8
INTRA-4×4	1.0	INTRA-16×16	1.0

Table 2. Macroblock classification accuracy

Sequence	SMB Err Ratio	CMB Err Ratio	Correct Ratio
Container	3.39%	2.24%	94.37%
Salesman	4.50%	2.61%	92.89%
Children	8.10%	3.46%	88.44%
Paris	10.81%	5.71%	83.48%
Tempete	14.23%	7.42%	78.35%
Mobile	18.53%	9.04%	72.44%

3. MACROBLOCK MODE SELECTION PROCESS

The macroblock classification in Section 2 provides a coarse estimation, therefore the possible MB mode set should gradually shrink in stead of just being restricted in a definite subset. The macroblock modes can be separated into three groups, the simple motion group including DIRECT/SKIP, INTER-16×16, INTER-16×8, and INTER-8×16; the complex motion group including the INTER-8×8 mode with all sub-macroblock types and the INTRA modes group. Four schemes of mode early termination are performed to skip some mode groups as follows.

1) For SMBs, if the best mode of the simple motion mode group is the SKIP/DIRECT or INTER-16×16 mode, skip all other modes.

2) For CMBs, if the best sub-macroblock types of more than two 8×8 blocks are not the INTER-8×8 types, skip simple motion modes.

3) If the best mode is not the INTER-8×8 mode before checking the INTRA modes, skip the two INTRA modes.

4) For B-pictures, if the rd_cost of the DIRECT mode of a SMB is lower than the best rd_cost of one of its left, top, top-right MBs which also belongs to the SMBs, skip all other modes.

In the scheme, the mistakenly categorized MBs still have the chance to experience all MB modes, which guarantees the coding performance. The whole macroblock mode selection process is summarized in flowchart Fig. 1.

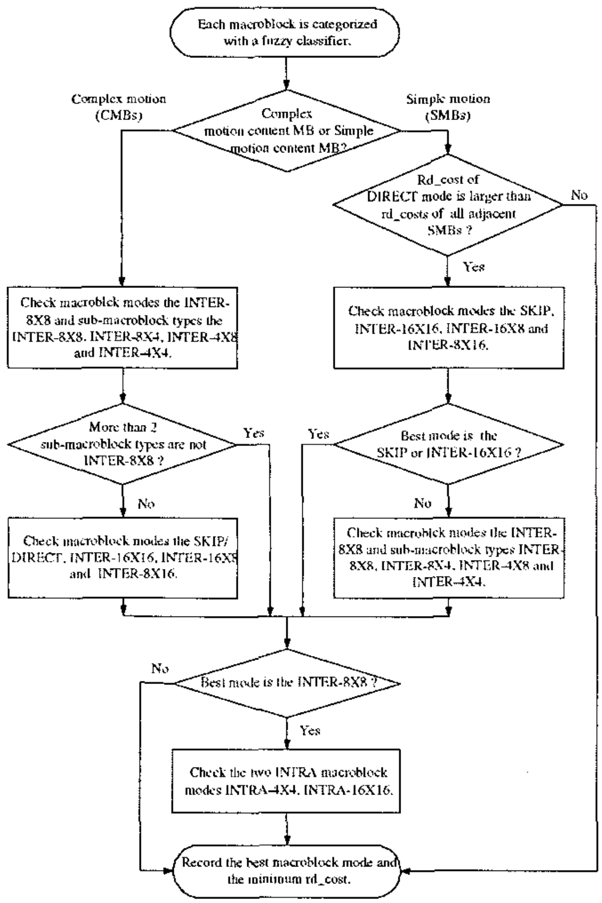


Fig. 1. Macroblock mode selection process.

If the best mode accords with the result of the exhaustive mode selection, it is referred as a Corrent-MB. If the MB has skipped some mode groups except only omitted the two INTRA modes, it is referred as an ET-MB. The number and ratio of Corrent-MBs and ET-MBs of 100 frames in 6 CIF sequences are listed in Table 3, which reflects the performance of the mode selection.

4. SIMULATION RESULTS

The proposed FMMS algorithm is embedded in JM6.1e reference encoder and simulated on the CIF sequences. For each sequence, 101 frames are encoded with the H.264/AVC Main Profile except the multiple reference frames. Q_P factor equals 28 for 1 I-picture and 50 P-pictures, and 30 for 50 B-pictures. The FMMS is compared with the exhaustive macroblock mode selection (EMMS), incorporating with 4 fast motion estimation algorithms respectively, the diamond search (DS) [7], the hexagon-based search (HEXBS) [8], the four-step search (4SS) [9], the block-based gradient descent search (BBGDS) [10], the search window is [-16,+16].

Table 3. Macroblock mode selection accuracy

Sequence	ET-MB	ET-Ratio	Correct-MB	Ratio
Container	33173	83.77%	38369	96.89%
Salesman	29132	73.57%	38785	97.94%
Children	29536	74.59%	37051	93.56%
Paris	25946	65.52%	38286	96.68%
Tempete	18137	45.80%	37172	93.87%
Mobile	15870	40.08%	36644	92.54%

The rate distortion performance is reflected by average peak signal-to-noise ratio (PSNR) of the luminance component and the average bit rate. The complexity is measured with the number of calculated sum of absolute differences (SAD) of 4×4 block in the full-pel motion estimation and the number of calculated sum of absolute transformed differences (SATD) of 4×4 block in the sub-pel motion estimation and the calculations of INTRA mode cost. In the bracket is the saving percent of calculated SAD and SATD numbers.

The simulation results of two representative sequences are shown in Table 4 and Table 5, one with relatively high accuracy in the MB classification, the other with relatively low accuracy. Salesman is a typical video conference sequence with a still camera on human subjects. Tempete contains camera zooming, spatial details and fast random motion. Within 0.1dB PSNR loss and less than 2% bit rate increase, more than 50% of the 4×4 block SAD and nearly 70% of SATD are saved for Salesman, while more than 40% SAD and 50% SATD of 4×4 block are reduced for Tempete.

The performance of FMMS under a wide range of coding rates is illustrated in Fig. 2, which plots the rate distortion curve of Mobile and the corresponding calculation saving percentage against the PSNR employing the DS algorithm. PSNR difference between the EMMS and FMMS is magnified in Fig. 2(a). Fig. 2(b) illustrates that the FMMS can reduce substantial computations in a broad range of bit rates and image qualities.

5. CONCLUSION

In this paper, a fast macroblock mode selection algorithm in H.264/AVC is proposed base on the classification of the motion content complexity. The method can save 40%–70% of the 4×4 block SAD and SATD calculations over the typical fast motion estimation algorithms such as the DS, HEXBS, 4SS, and BBGDS. Accompany with the fast development of video coding techniques, the increasing demand on extensive computation resources indicates that flexible and adap-

Table 4. Salesman CIF 30Hz

Method	PSNRY(dB)	Bit rate(Kbit/s)	SAD	SATD
DS	35.6	159.97	13.62M	74.52M
FMMS	35.56(-0.04)	160.2(+0.16%)	6.06M(-55.50%)	23.38M(-68.60%)
HEXBS	35.6	159.62	11.75M	74.47M
FMMS	35.56(-0.04)	162.6(+1.84%)	5.22M(-55.56%)	23.36M(-68.63%)
4SS	35.61	160.71	17.74M	74.53M
FMMS	35.57(-0.04)	161.3(+0.37%)	7.54M(-57.52%)	23.40M(-68.60%)
BBGDS	35.61	160.39	9.48M	74.40M
FMMS	35.57(-0.04)	162.5(+1.31%)	4.25M(-55.20%)	23.28M(-68.71%)

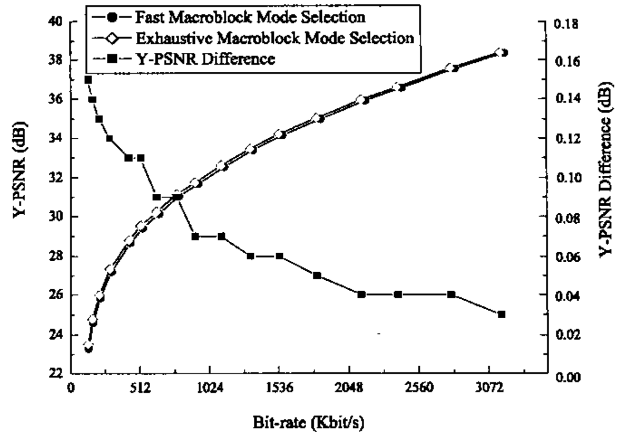
'M' represents one million times.

Table 5. Tempete CIF 30Hz

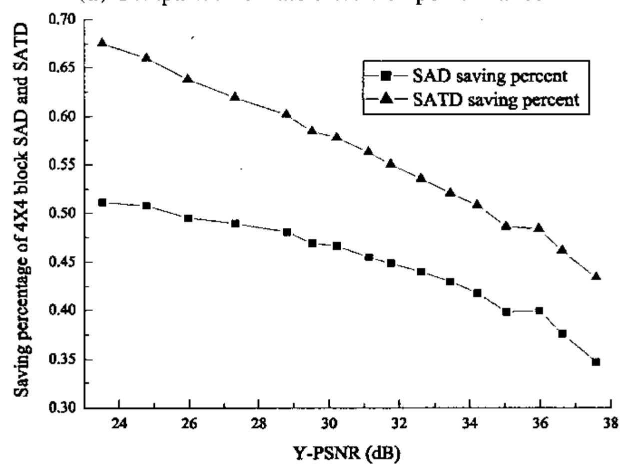
Method	PSNRY(dB)	Bit rate(Kbit/s)	SAD	SATD
DS	34.08	850.48	16.79M	77.85M
FMMS	34.02(-0.06)	865.0(+1.70%)	9.74M(-41.96%)	35.05M(-54.99%)
HEXBS	34.07	851.58	15.09M	77.92M
FMMS	34.01(-0.06)	866.0(+1.70%)	8.68M(-42.46%)	34.99M(-55.10%)
4SS	34.08	853.06	21.78M	77.90M
FMMS	34.01(-0.07)	865.4(+1.44%)	12.3M(-43.57%)	35.05M(-55.00%)
BBGDS	34.07	850.07	11.74M	77.66M
FMMS	34.01(-0.06)	865.9(+1.86%)	6.77M(-42.30%)	34.76M(-55.24%)

'M' represents one million times.

tive encoding mode selection is a promising approach and deserves more consideration.



(a) Comparison of rate distortion performance



(b) Percentage of calculation savings

Fig. 2. Mobile&Calendar CIF 30Hz

6. REFERENCES

[1] "ITU-T Rec.H.264, ISO/IEC 14496-10 AVC, Advanced Video Coding Final Draft International Standard, JVT-G050r1, Geneva, Switzerland," May 2003.

[2] M.Flierl and B.Girod, "Generalized B pictures and the draft H.264/AVC video-compression standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 587–597, July 2003.

[3] "ITU-T Rec.H.263, Video Coding for Low Bit Rate Communication," Feb. 1998.

[4] T. Wiegand, H.Schwarz, and A.Joch, "Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 688–703, July 2003.

[5] Ludmila I. Kuncheva, *Fuzzy classifier design*, chapter 5, Physica-Verlag Heidelberg, New York, 2000.

[6] A. Chang, C. A. Oscar, and Y. M. Yeung, "A novel approach to fast multi-frame selection for H.264 video coding," in *ICASSP'2003*, Hong Kong, Apr. 6-10, 2003, vol. 3, pp. 413–416.

[7] S. Zhu and K. K. Ma, "A new diamond search algorithm for fast block matching motion estimation," *IEEE Trans. Image Processing*, vol. 9, pp. 287–290, Feb. 2000.

[8] C. Zhu, X. Lin, and L. P. Chau, "Hexagon-based search patten for fast block motion estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, pp. 349–355, May 2002.

[9] Po L. M and W. C. Ma, "A novel four-step search algorithm for fast block motion estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 313–317, June 1996.

[10] L. K. Liu and E. Feig, "A block-based gradient descent search algorithm for block motion estimation in video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 419–423, Aug. 1996.