Fast and Adaptive Inter Mode Decision for Pslices in H. 264

DONG Yu-tao, FANG Xiang-zhong
(*The Institute of Image Communication and Information Processing, Shanghai Jiaotong University, Shanghai* 200240)

Abstract A new fast and adaptive inter mode decision algorithm for Pslices in H. 264 is proposed in this paper. In the algorithm, SKIP mode is decided to be the best mode without Lagrange rate-distortion optimization(RDO), P8 × 8 mode is decided to be removed from the candidate modes before Lagrange RDO. With the above decision of the two special modes, the entire encoding time is greatly reduced. Simulation results show that the proposed algorithm can save 60% ~ 80% entire encoding time with negligible variance in PSNR and bit rate compared to the H. 264 reference software JM 10. 2.

Keywords mode decision, H. 264, adaptive

H. 264 中 Pslices 帧间快速自适应模式判决算法

董彧焘 方向忠
(上海交通大学图像通信与信息处理研究所，上海 200240)

摘 要 为了进一步提高编码效率，提出了一种新的 H. 264 中 Pslices 帧间快速自适应模式判决算法。在该算法中，SKIP 模式无需经过拉格朗日率失真优化就可以判定为最佳模式；P8 × 8 模式也可以在拉格朗日率失真优化之前从候选模式中排除。通过对以上两种特殊模式的判决，可以大大减少整个编码时间。仿真结果表明，和 H. 264 参考软件 JM10. 2 相比，该算法在整个编码时间上节省 60% ~80% 的时间，而在 PSNR 和码率上的变化则可以忽略不计。

关键词 模式判决 H. 264 自适应
中图法分类号：TN919. 81 **文献标识码**：A **文章编号**：1006-8961(2007)10-1749-04

1 Introduction

H. 264 is the latest international video coding standard. Based on previous successful video coding standards, many new features are added in H. 264, such as variable block size motion estimation, integer DCT-like transform and so on. Compared to previous standards, H. 264 can improve the coding efficiency about 50% [1]. However, H. 264 is more complex than previous standards due to enabling many new features. In all the new features of H. 264, the most time-consuming function block is the variable block size motion estimation (ME). H. 264 supports not only 16 × 16 macroblock(MB) mode in ME, but also 16 × 8, 8 × 16, 8 × 8, 8 × 4, 4 × 8 and 4 × 4 for prediction flexibility. In order to decide the best mode for a MB, it is necessary to compute the rate-distortion costs (RDcost) for all possible modes above according to Eq. (1) as follows[2]:

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

where QP is the MB's quantization parameter, λ_{MODE} is the Lagrange multiplier for mode decision and is computed empirically based on QP[2], SSD is the sum of squared difference between the original block s and its reconstruction block c, and $MODE$ is one of the available modes for ME. The final decided mode is the one with the least RDcost. In recent literatures, fast mode decision is proved to be an effective way to lower the

收稿日期：2007-06-10；改回日期：2007-06-30
第一作者简介：董彧焘（1977 ~ ），男，博士。主要从事视频编码研究。E-mail：ytdong@sjtu.edu.cn

computational complexity of ME in H.264. In [2], Wu *et al* proposed a fast mode decision algorithm by using edge information and MB difference in video sequences. In [3], Jing and Chau proposed an algorithm by comparing the mean absolute difference (MAD) of current frame and the MAD of current MB to determine the potential modes. In [4], Feng *et al* proposed a fast inter mode decision algorithm by organizing the candidate inter modes to some mode groups according to the characteristics of residual. Different from the above algorithms, a new fast and adaptive inter mode decision algorithm for Pslices is proposed in this paper. In the new proposed algorithm, SKIP and P8 × 8 modes are decided fast and accurately by analyzing both the spatial and temporal information of each MB.

This paper is organized as follows. Section 2 details the proposed fast and adaptive inter mode decision algorithm. Section 3 shows the simulation results and section 4 concludes this paper.

2 Fast and adaptive inter mode decision algorithm

Based on the H. 264 reference software JM 10.2[5], 10 different QCIF-format video sequences are encoded with full mode decision. The percentage of each mode in every encoded sequence is plotted in Tab. 1. It can be concluded from Tab. 1 as follows:

1) The final selected modes are mainly 16 × 16, 16 × 8, 8 × 16, P8 × 8 and SKIP.

2) SKIP mode has the highest probability to be the best mode.

Tab.1 The final mode percentage of the 10 test sequences based on the JM 10.2 with full mode decision (IPPP…, QP = 28, 50 frames in all)

Sequence	SKIP(%)	16×16(%)	16×8(%)	8×16(%)	P8×8(%)	Intra(%)
Carphone	34.6	28.3	9.4	8.8	16.4	1.5
Container	77.6	9.3	3.8	2.6	6.7	0.1
Foreman	27.6	28.8	10.3	15.9	17.3	0.1
Hall_monitor	82.3	9.5	0.8	1.1	5.9	0.5
Mother_and_daughterr	71.3	13.1	5.1	4.9	5.6	0.0
News	78.0	5.9	1.8	4.5	9.8	0.0
Salesman	74.0	5.7	3.0	3.3	14.0	0.0
Silent	66.7	11.2	3.9	4.7	12.2	1.3
Susie	43.1	27.6	9.3	11.6	7.2	1.3
Trevor	31.8	20.7	8.8	9.1	29.0	0.5

More over, it is well known that SKIP mode needs the least computation and P8 × 8 mode needs the most computation because 8 × 8 blocks will be further divided into 8 × 4, 4 × 8 and 4 × 4 sub-blocks. If SKIP mode can be decided to be the best mode for a MB correctly in advance, all other modes can be omitted from later Lagrange RDO and a lot of computation can be saved. If P8 × 8 mode can be decided to be removed from the candidate modes, a lot of computation can also be saved. As we all know, MB's variance (MBV) contains lots of spatial information. Generally, the MB with large variance contains complex texture information; and the MB with small variance belongs to flat region. Therefore, MBV's difference (MBVD) reflects the similarity of the current MB and the MB at the same location in previous frame. They can be calculated according to Eq. (2) and Eq. (3) as follows:

$$MBV = \sum_{i=1}^{16}\sum_{j=1}^{16}(P_{i,j} - \bar{P}_{MB}) \quad (2)$$

$$MBVD = MBV(n) - MBV(n-1) \quad (3)$$

where $P_{i,j}$ is the luma value of the pixel at location (i, j) in current MB, $\bar{P}_{MB}$ is the average value of all luma value in current MB, index n and $n-1$ denote the current and the previous frame. Through extensive experiments, it is found that those MBs encoded with SKIP mode all have very small *MBVD* value and those MBs encoded with P8 × 8 mode all have big *MBVD* value, but those MBs with small *MBVD* value are not all encoded with SKIP mode because the time-varing motion is not considered. MB's sum of absolute difference (SAD) is a good measurement for temporal motion information. In order to indicate the motion more accurately, block's SAD (BSAD) is used instead of MB's SAD in the proposed algorithm and it can be calculated by Eq. (4):

$$BSAD = \sum_{i=1}^{8}\sum_{j=1}^{8}(P_{i,j}(n) - P_{i,j}(n-1)) \quad (4)$$

Among those MBs encoded with SKIP mode, it is also found that bigger *MBV* has bigger *MBVD* and *BSAD*. Therefore, it is appropriate to divide the whole range of *MBV* into some subranges and set different threshold for *MBVD* and *BSAD* in different subranges. In the proposed algorithm, the whole range of *MBV* is divided into seven subranges and corresponding thresholds of *MBVD* and *BSAD* are set for each subrange according to Tab. 2. Note that there is no threshold set for *BSAD* in the first subrange because *BSADs* of the MBs in that subrange are all very small.

Tab. 2 Subranges and corresponding *MBVD* and *BSAD* thresholds

Subrange (MBV)	<100	100~500	500~1000	1000~1500	1500~2000	2000~2500	>2500
Threshold T_1 (MBVD)	1.0	1.0	1.5	2.0	2.5	3.0	3.5
Threshold T_2 (BSAD)	–	25	30	35	40	45	50

The proposed algorithm can be described as follows:

Step 1: Calculate the current MB's *MBV* based on Eq. (2);

Step 2: Calculate the current MB's *MBVD* and four *BSADs* according to Eq. (3) and Eq. (4), then select the biggest *BSAD* as the current MB's *BSAD*;

Step 3: Select subrange according to MBV, and then set the corresponding *MBVD* and *BSAD* thresholds;

Step 4: Mode decision:

(a) If the *MBVD* is less than T_1, go to (b) for further decision. Otherwise, SKIP mode is removed from the candidate modes and go to step 5;

(b) If the current MB's *BSAD* is less than T_2, SKIP mode is selected as the best mode. Then go to step 6. Otherwise, SKIP mode is removed from the candidate modes. Further more, P8 × 8 is also removed from the candidate modes because *MBVD* is small, and then go to step 5;

Step 5: Compute the RDcosts of the remaining candidate modes and select the mode which leads the minimal RDcost as the best mode;

Step 6: Encode the current MB with the best mode and then return to step 1 to encode the next MB.

Fig. 1 is the flowchart of the inter mode decision process of a MB.

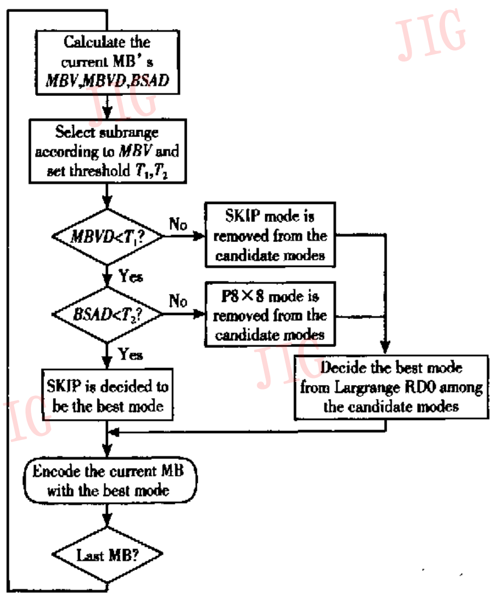

Fig. 1 The flowchart of a MB's mode decision with the proposed algorithm

3 Simulation results

The simulations run on both the standard JM 10.2 and the proposed algorithm with the same encoding parameters and environments. The input video sequences are standard QCIF-format sequences. The encoded picture type is IPPP, the quantization parameter is 28, and the maximal search range is 16. In the simulations, 10 different video sequences are encoded all with 50 frames. Moreover, all the 10 video sequences contain a wide range from low motion to high motion and from simple texture to complex texture, which can prove that the simulations are complete and reliable. The simulation results are illustrated in Tab. 3. It can be easily found that there is almost no change in PSNR and bit rate while the entire encoding time is greatly reduced compared with the JM 10.2. In Tab. 4, compared to the standard reference software, the saved encoding time of the other two fast algorithms is about 50%, while the proposed algorithm reaches 60% ~ 80%, which can prove that the proposed algorithm perform better in reducing the encoding time.

Tab. 3 Comparison between the proposed algorithm and the JM 10.2 in PSNR, bit rate, and saved encoding time

Sequence	ΔPSNR(dB)	Δbit rate(%)	Saved encoding time(%)
Carphone	-0.11	-0.4	75.7
Container	-0.14	0.6	78.9
Foreman	0.0	1.1	60.3
Hall_monitor	-0.03	-1.0	66.6
Mother_and_daughter	-0.08	-1.0	68.5
News	-0.18	0.7	80.6
Salesman	-0.03	-0.1	67.0
Silent	0.0	1.2	77.5
Susie	-0.18	0.2	67.6
Trevor	-0.07	-0.1	66.1

Tab. 4 Comparison between the proposed algorithm and other fast algorithms (Compared to the standard reference software)

Methods	Saved encoding time range(average)
Jing's algorithm[3]	29% ~48% (38%)
Feng's algorithm[4]	46% ~55% (52%)
The proposed algorithm	60% ~80% (71%)

4 Conclusions

In this paper, a new and more accurate fast inter mode decision algorithm is proposed. First, we encode a variety of video sequences with the JM 10.2 and find that SKIP mode has the highest probability to be the best mode. It is well known that P8 × 8 mode is the most time-consuming mode in Lagrange RDO among all the possible modes. Then the algorithm is proposed which decides these two special modes accurately by analyzing both the spatial and temporal information of MB. In the proposed algorithm, *MBV*, *MBVD* and *BSAD* are used together to decide the above two modes. Simulation results show that the proposed algorithm can reduce 60% ~80% entire encoding time of the JM 10.2 with negligible variance in PSNR and bit rate. It can be concluded that the proposed algorithm is more effective and can be applied to real time scenarios.

References

1 Wiegand T, Sullivan G, Bjontegaard G, *et al*. Overview of the H.264/AVC video coding standard [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2003, 13(7): 1 ~ 19.

2 Wu D, Pan F, Lim K P, *et al*. Fast intermode decision in H.264/AVC video coding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2005, 15(6): 953 ~958.

3 Jing X, Chou L P. Fast approach for H.264 inter mode decision[J]. IEE Electronic Letters, 2004, 40(17): 1050 ~ 1052.

4 Feng B, Zhu G, Liu W. Fast adaptive inter mode decision method for Pslices in H. 264 [A]. In: Proceedings of IEEE Consumer Communications and Networking Conference[C], Las Vegas, NV, USA, 2006: 745 ~ 748.

5 Joint Video Team (JVT). reference software JM 10.2 [S/OL], http://bs.hhi.de/~suehring/tml/download.