

## Fast Inter Coding Mode Selection Algorithm for H.264

DONG Hai-yan, ZHANG Qi-shan

(School of Electronic Information and Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

**Abstract :** In order to reduce the encoding complexity of macroblock coding mode decision in H.264/AVC, a selective smaller block-size searching algorithm and a selective intra coding mode searching algorithm are proposed by using the high correlation among coding modes and in spatial and temporal domains of video sequence. Simulation results demonstrate that the proposed algorithm can provide significant improvement in computational requirement, with negligible small picture quality degradation and slight bit rate increase.

**Key words:** H.264 ;video coding ;motion estimation ;inter frame coding ;intra frame coding ;mode decision ;rate distortion

**H.264 快速帧间编码模式选择算法.** 董海燕, 张其善. 中国航空学报(英文版), 2006,19(3): 219-224

**摘要 :** 为了减少 H.264/AVC 帧间编码模式选择的计算复杂度, 利用编码模式之间的相关性以及视频序列时空域的相关性, 提出了有选择性的小块搜索技术和有选择性的帧内编码模式搜索技术。模拟结果表明, 该算法在保持率失真性能的前提下可以大幅度减少模式选择的计算复杂度, 有利于 H.264 的实时应用。

**关键词 :** H.264 ; 视频编码 ; 运动估计 ; 帧间编码 ; 帧内编码 ; 模式选择 ; 率失真

文章编号 : 1000-9361(2006)03-0219-06

中图分类号 : TP391

文献标识码 : A

H.264/AVC, the new international video coding standard jointly developed and approved by the MPEG group ISO/IEC and the VCEG group of ITU-T, can provides an improved coding efficiency and a significant improvement in flexibility for effective use over a wide range of networks<sup>[1]</sup>. While H.264/AVC still uses the concept of block-based motion compensation, it provides some significant changes, such as enhanced motion compensation capability using variable block sizes<sup>[2]</sup>. In former standards as MPEG-4 or H.263, only blocks of the size 16×16 and 8×8 are supported. Compared to the fixed block size prediction, the variable block-size prediction can reduce the prediction error so that higher compression efficiency will be achieved. However, the encoding complexity and time increase linearly with the number of block-sizes used.

In the H.264 coding algorithm, block matching motion estimation is an essential and effective part of the encoder to reduce the temporal redundancy among frames. In motion compensation of H.264, the macroblock has various block sizes of 16×16, 16×8, 8×16 and 8×8, and the corresponding modes are Mode(1), Mode(2), Mode(3) and Mode(P8×8) separately<sup>[3]</sup>. In Mode(P8×8), each 8×8 block can be further divided into partitions with block sizes of 8×4, 4×8 and 4×4. Each 8×8 block can independently has any combinations of 8×8, 8×4, 4×8 and 4×4. What's more, the inter frame coding modes also support Mode(0), Mode(I16MB) and Mode(I4MB). Thereinto, Mode(0), namely SKIP, is a direct copy from the previous frame<sup>[4]</sup>; Mode(I16MB) and Mode(I4MB) are the intra modes predicted from encoded adjacent pixels. All the available segmentations of a macroblock sup-

ported by H.264 motion compensation are shown in Fig.1.

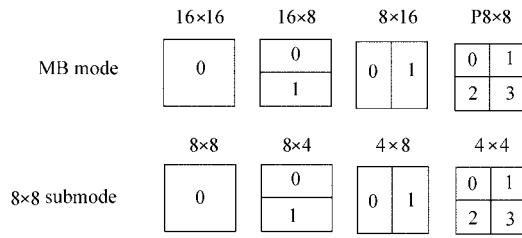


Fig.1 Variable block-size prediction as defined in H.264

In order to choose the best block size for a macroblock, the H.264 standard makes use of the computationally intensive Lagrange rate-distortion (RD) optimization<sup>[5]</sup>. The general equation of Lagrange RD optimization is given as

$$J_{\text{mode}}(s, c, \text{MODE} | \lambda_{\text{mode}}) = \text{SSD}(s, c, \text{MODE}) + \lambda_{\text{mode}} \cdot R(s, c, \text{MODE}) \quad (1)$$

where MODE is a mode of potential inter-frame coding macroblock modes,  $J_{\text{mode}}(s, c, \text{MODE} | \lambda_{\text{mode}})$  is the rate-distortion cost (RD cost) corresponding to MODE,  $s$  indicates the original video signal,  $c$  denotes the coded video signal,  $\lambda_{\text{mode}}$  is the Lagrange multiplier,  $R(s, c, \text{MODE})$  represents the bits associated with choosing MODE, including the bits for the macroblock header, the motion information and all DCT coefficients, and the  $\text{SSD}(s, c, \text{MODE})$  (Sum of Square Difference) is the sum of square differences between the original signal and the reconstructed signal and can be computed as

$$\text{SSD}(s, c, \text{MODE}) = \sum_{i=1, j=1}^{B_1, B_2} |s[i, j] - c[i - m_x, j - m_y]|^2 \quad (2)$$

with  $B_1$  and  $B_2$  being the block sizes which can be 16, 8, or 4 and  $\mathbf{m} = (m_x \ m_y)^T$  denoting the motion vector.

In this paper, a fast inter frame coding mode selection algorithm by using the statistical results of adopted probability for all modes and high correlation between spatial and temporal domains is presented. Instead of searching through all the possible block sizes, the proposed algorithm tries to skip the block sizes that tend to give no performance gain. This is very useful for real-time

applications.

The rest of the paper is organized as follows. In Section 1, the inter frame coding modes are analyzed. The fast inter frame coding mode selection algorithm is proposed in Section 2. The simulation results are presented in Section 3. The conclusions are shown in Section 4.

## 1 Analysis of Inter Frame Coding Modes

As is well known, all potential macroblock modes for H.264 inter-frame coding are Mode(0), Mode(1), Mode(2), Mode(3), Mode(P8×8), Mode(I16MB) and Mode(I4MB)<sup>[6]</sup>. The optimal mode which produces the least RD cost is selected. The H.264 standard employs a brute force algorithm to search through all possible block sizes to find a motion vector for each macroblock. Therefore, the encoding complexity and time increase linearly with the number of modes used. The high coding efficiency achieved by H.264 is at the cost of a very computationally-intensive procedure of finding the “best” coding mode.

The size of block affects the performance of the motion estimation directly. The effective approach is to adapt the block size to the picture characteristics. In general, a large partition size is appropriate for flat and homogenous regions of the frame and a small partition size may be beneficial for high details and complex motion regions<sup>[7]</sup>.

Since the flat and homogenous areas tend to have similar motion, it should not be split into smaller blocks. Choosing a large partition size means that a small number of bits are required to signal the choice of motion vectors and the type of partition but the motion compensated residual may contain a significant amount of energy in frame areas with high detail.

On the contrary, since high-details or high motion areas tend to have complex video content or high motion information and small block sizes afford good approximation to the natural object boundaries, the motion estimation may be more accurate if the areas are encoded by smaller block

sizes. It can be proved that smaller motion compensation block sizes can produce better motion compensation results; however, a smaller block size leads to increased complexity due to more search operations required and an increase in the number of motion vectors that need to be transmitted. Sending each motion vector requires bits to be sent and the extra overhead for vectors may outweigh the benefit of reduced residual energy<sup>[8]</sup>. Choosing a small partition size may give a lower-energy residual after motion compensation but requires a larger number of bits to signal the motion vectors and choice of partition. The choice of partition therefore has a significant impact on compression performance.

Fig.2 shows a typical frame of QCIF sequence “News”<sup>[9]</sup>. The block sizes depicted in Fig.2 are chosen by original full and brute search scheme in H.264. It’s easy to see that flat and homogenous areas such as the background, the black suit of the man and the white suit of the woman are encoded using  $16 \times 16$  block sizes. On



Fig.2 Example of inter coding modes in H.264

the other hand, high-details and high motion areas, such as the boundary regions and the dancers in the upper part of the image, are encoded using smaller block sizes. The simulation results of actual sequence prove above analysis again.

## 2 Proposed Fast Mode Selection Algorithm

In order to further understand the application proportions of different coding modes in actual sequences, several typical video sequences are selected and tested. The adopted percent of each mode for inter frame coding is shown in Table 1.

Through the analysis of H.264 performances on different video sequences, it can be known that the mode adopted as the optimal inter frame coding mode most often is Mode(0) and the proportion of Mode(0) in all possible inter modes even can be up to 87%. What’s more, it can be seen that the larger block sizes are selected as the best mode more often than smaller block sizes. On average, the application frequency of Mode(1) is higher than that of Mode(2) or Mode(3), and similarly Mode(2) or Mode(3) are chosen more often than Mode(P8×8). Furthermore, the two intra modes have got little adopted probability and the probabilities of both Mode(I4MB) and Mode(I16MB) are less than 3%.

The simulation results demonstrate that the larger block sizes ( $16 \times 16$ ,  $16 \times 8$  and  $8 \times 16$ ) are selected as the final coding mode more often than the smaller block size ( $8 \times 8$ ,  $8 \times 4$ ,  $4 \times 8$  and  $4 \times 4$ ) because most regions in video sequences belong to stationary or homogenous regions.

**Table 1 The adopted percents of all inter coding modes**

Sequence	Mode(0)	Mode(1)	Mode(2)	Mode(3)	Mode(P8×8)	Mode(I16MB)	Mode(I4MB)
Suzi_qcif.yuv	54.69	27.81	5.97	7.21	3.11	0.56	0.66
Silent_qcif.yuv	71.32	11.69	4.57	5.49	6.08	0.62	0.23
Football_qcif.yuv	14.09	19.54	9.49	10.23	31.7	11.52	3.43
Akiyo_cif.yuv	87.39	7.57	1.97	2.17	0.87	0.000 36	0.000 07
Mobile_cif.yuv	11.24	33.96	16.81	16.34	21.47	0.000 48	0.001 4
Coastg_cif.yuv	24.62	36.34	12.67	11.47	13.41	1.07	0.41
Stefan_cif.yuv	31.12	27.01	11.74	10.45	18.02	0.73	0.94
Average	42.07	23.42	9.03	9.05	13.52	2.07	0.81

From the results above, it can also be seen that the probability of a macroblock to have an intra mode in P slice is less than 0.03. What's more, the RD cost computation for intra macroblocks is about 5 times higher than that for inter modes<sup>[10,11]</sup>. Although the exact figure depends on specific input video characteristics, the small probability of intra mode (Mode(I4MB) and Mode(I16MB)) in real video sequences suggests that the motion search and RD cost computations for the two intra modes can be omitted whenever it is possible. Unless these above modes can meet the defined requirements, the two intra coding modes, Mode(I4MB) and Mode(I16MB), are checked finally to find the best mode producing minimum RD cost.

In other words, the smaller block sizes (8×8, 8×4, 4×8 and 4×4) and two intra modes have the minimum probability to be selected as the final coding mode of inter macroblocks. If the coding modes that make small contributions can be skipped, then larger numbers of unnecessary operations can be avoided. Therefore, the coding speed will be improved much.

In order to reduce the encoding complexity of macroblock coding mode decision in H.264/AVC, a fast inter coding mode selection algorithm is proposed. The fast inter coding mode selection algorithm includes a selective smaller block-size searching algorithm and a selective intra coding mode searching algorithm.

### 2.1 Proposed selective smaller block-size searching algorithm

The selective smaller block-size searching algorithm makes full use of the correlations among the inter frame coding modes. The least RD cost is selected among the block-sizes 16×16, 16×8, 8×16 and 8×8. On one hand, if the RD cost of 8×8 is not the least, then the RD cost of smaller block-size (8×4, 4×8 or 4×4) being the least is impossible, and the searching process for the smaller block-size (8×4, 4×8 or 4×4) is unnecessary. On the other hand, if the RD cost of 8×8 is the least, further searching for the smaller block-size (8×4, 4×8 or 4×4) is needed.

### 2.2 Proposed selective intra coding modes searching algorithm

As it is well known, the inter coding modes mainly use the temporal correlation between the current coding frame and reference frame, and the intra coding modes mainly use the spatial correlation between the current coding macroblock and adjacent pixel. To certain macroblock in P slice, if the temporal correlation is less than the spatial correlation, thus, the probability of intra coding mode to be selected as the final coding mode for the macroblock is lower. In this case, the searching process for these intra coding modes should be skipped to decrease the coding complexity without obvious influence on coding performance.

The paper utilizes the prediction data obtained from the optimal coding mode among the inter coding modes {0,1,2,3,P8×8} to calculate the temporal correlation of the current coding macroblock. Define TC to indicate the temporal correlation as following

$$TC = \frac{1}{384} \left( \sum_{i=0}^{15} \sum_{j=0}^{15} |Y_{\text{orig}}(x+i, y+j) - Y_{\text{pred}}(x+i, y+j)| + \sum_{i=0}^7 \sum_{j=0}^7 |U_{\text{orig}}(x+i, y+j) - U_{\text{pred}}(x+i, y+j)| + \sum_{i=0}^7 \sum_{j=0}^7 |V_{\text{orig}}(x+i, y+j) - V_{\text{pred}}(x+i, y+j)| \right) \quad (3)$$

where  $(x, y)$  indicates the top left position pixel of current macroblock,  $Y_{\text{orig}}, U_{\text{orig}}$  and  $V_{\text{orig}}$  separately indicate the original pixel values for luminance and chrominance parts of the current coding macroblock,  $Y_{\text{pred}}, U_{\text{pred}}$  and  $V_{\text{pred}}$  separately indicate the predicted pixel values for luminance and chrominance parts when the optimal mode among {0,1,2,3,P8×8} is used to code the current macroblock.

The paper utilizes the 32 adjacent pixels above and left to the current macroblock to calculate the spatial correlation of the current macroblock. Define SC to indicate the spatial correlation as following

$$SC = \frac{1}{64} \left( \sum_{i=0}^{15} |Y_{\text{orig}}(x+i, y) - Y_{\text{rec}}(x+i, y-1)| + \right)$$

$$\begin{aligned}
& \sum_{j=0}^{15} |Y_{\text{orig}}(x, y + j) - Y_{\text{rec}}(x - 1, y + j)| + \\
& \sum_{i=0}^7 |U_{\text{orig}}(x + i, y) - U_{\text{rec}}(x + i, y - 1)| + \\
& \sum_{j=0}^7 |U_{\text{orig}}(x, y + j) - U_{\text{rec}}(x - 1, y + j)| + \quad (4) \\
& \sum_{i=0}^7 |V_{\text{orig}}(x + i, y) - V_{\text{rec}}(x + i, y - 1)| + \\
& \left. \sum_{j=0}^7 |V_{\text{orig}}(x, y + j) - V_{\text{rec}}(x - 1, y + j)| \right)
\end{aligned}$$

where  $(x, y)$  indicates the top left position pixel of current macroblock;  $Y_{\text{orig}}, U_{\text{orig}}$  and  $V_{\text{orig}}$  separately indicate the original pixel values for luminance and chrominance parts of the current coding macroblock;  $Y_{\text{rec}}, U_{\text{rec}}$  and  $V_{\text{rec}}$  separately indicate the reconstructed adjacent pixel values for luminance and chrominance part above and left to the current macroblock.

If  $SC < TC$ , it suggests that the spatial correlation is larger than the temporal correlation for the current macroblock, and the two intra coding modes (I4MB and I16MB) should be searched further. If not, the searching processes and RD cost computations for two intra coding modes are unneeded, so that the unnecessary resource consuming is saved and the coding speed can be improved.

### 3 Simulation Results

The proposed scheme was implemented in the H.264 reference software JM 8.6. The proposed algorithm was compared with the original scheme in the H.264 reference software. The experimental parameters are set as Table 2.

**Table 2 Experimental parameters**

Software	JM 8.6	MV	$\pm 16$
Hadamard	ON	Entropy	CAVLC
QP	32	GOP	IPPP
Reference frames	1	RDO	ON
Sequences			
News	CIF(100 frames)	Mobile	QCIF(120 frames)
Foreman	CIF(80 frames)	Template	QCIF(150 frames)
Bus	CIF(100 frames)	Suzi	QCIF(100 frames)

Table 3 gives the results of PSNR, bit rate and complexity reduction of proposed scheme in the test sequences. In Table 3,  $\Delta_{\text{PSNR}}$  (average difference in PSNR between two methods in dB),  $\Delta_{\text{bits}}$  (average difference in bit-rate between two methods in percent) and  $\Delta_{\text{time}}$  (average difference in encoding time between two methods in percent) are used for the evaluation of coding efficiency.  $\Delta_{\text{time}}$  and  $\Delta_{\text{bits}}$  are separately calculated as following

$$\Delta_{\text{time}} = \frac{T_{\text{orig}} - T_{\text{prop}}}{T_{\text{orig}}} \quad (5)$$

$$\Delta_{\text{bits}} = \frac{B_{\text{orig}} - B_{\text{prop}}}{B_{\text{orig}}} \quad (6)$$

where  $T_{\text{orig}}$  and  $B_{\text{orig}}$  separately indicate the coding time and the total coding bits using the original algorithm,  $T_{\text{prop}}$  and  $B_{\text{prop}}$  separately indicate the coding time and the total coding bits using the proposed algorithm.

**Table 3 Experimental results**

Sequences	$\Delta_{\text{PSNR}}/\text{dB}$	$\Delta_{\text{bits}}$	$\Delta_{\text{time}}$
News	0.030	0.12	58.11
Mobile	0.037	1.27	61.91
Foreman	0.026	0.92	57.42
Template	0.009	0.23	60.50
Bus	0.009	0.26	60.07
Suzi	-0.029	0.04	55.65
Average	0.014	0.47	58.94

From the experimental results in Table 3, it can be observed that, compared with the original mode selection scheme, the proposed inter mode selection algorithm reduces the encoding time by 58.94% on average with negligibly small PSNR degradation (0.014 dB) and slight increase in bit rate (0.47%). This means that the encoding time with the proposed algorithm used is about 2.5 times faster than that with the reference software JM7.0. However, the losses in terms of PSNR and bit rate are very small.

### 4 Conclusions

A novel fast inter mode selection algorithm that makes use of the inherent properties of different modes and the spatial correlations and temporal correlations among adjacent frames in video sequences is proposed. Extensive simulations show that average 58.94% computation reduction, average 0.47% bit rate increase and average 0.014 dB PSNR drop can be achieved. Simulation results demon-

strate that the proposed method provides significant improvement in computational requirement, with negligible small picture quality degradation and slight bit rate increase. The computation complexity savings helps in real-time implementation of H.264 video encoder.

### References

- [1] JVT-G050, Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG : Draft ITU-T recommendation and final draft international standard of joint video specification[S]. 2003.
- [2] Ostermann J, Bormans J, List P, et al. Video coding with H.264/AVC: tools, performance, and complexity[J]. IEEE Circuits and Systems Magazine, 2004, 4(1):7-28.
- [3] Kim Y H, Yoo J W, Lee S W, et al. Adaptive mode decision for H.264 encoder[J]. Electronics Letters, 2004, 40(19): 1172-1173.
- [4] Yu A. Efficient block-size selection algorithm for inter-frame coding H.264/MPEG-4 AVC[C]//IEEE International Conference on Acoustics, Speech, and Signal Processing. 2004:169-172.
- [5] Choi W I , Lee J, Yang S, et al. Fast motion estimation and mode decision with variable motion block sizes[C]//Ebrahimi T, Sikora T. Conference on visual communications and Image processing 2003. Bullingham, WA 98227-0010 USA: SPIE, 2003: 1561-1572.
- [6] Peng Y, Tourapis H Y , Tourapis A M , et al. Fast mode decision and motion estimation for JVT/H.264[C]. Proceedings 2003 International Conference on Image Processing. 2003:853-856.
- [7] Ahmad L, Khan N, Masud S, et al. Efficient block size selection in H.264 video coding standard[J]. Electronics Letters, 2004,40(1):19-21.
- [8] Richardson I. H.264 and MPEG-4 video compression, video coding for next-generation multimedia[M]. UK: Wiley, 2003.
- [9] Wu D, Wu S, Lim K P, et al. Block inter mode decision for fast encoding of H.264[C]//IEEE International Conference on Acoustics, Speech, and Signal Processing. 2004:181-184.
- [10] Lee J Y, Jeon B W. Fast mode decision for H.264 with variable motion block sizes[C]//Yazici A, Sener C. 18<sup>th</sup> International Symposium on Computer and Information Sciences(ISCIS2003). Berlin, Germany: Springer-Verlag, 2003:723-730.
- [11] Inchoon C , Jeyun L , Byeungwoo J. Efficient coding mode decision in MPEG-4 PART-10 AVC/H.264 main profile[C]// IEEE International Conference on Image Processing. 2004: 1141-1144.

### Biography:

**DONG hai-yan** Born in 1977, she received B.S. and doctoral degree from Beijing Institute of technology in 1999 and 2004 respectively, and then she became a postdoctoral at Beijing University of Aeronautics and Astronautics. E-mail: d\_haiyan@163.com