A Fast Selection Algorithm for Multiple Reference Frames in H.264/AVC*

Meng Qing-lei , Yao Chun-lian, and Li Bo

Digital Media Laboratory, School of Computer Science and Technology
Beihang University, Beijing 100083, China
mql198029@gmail.com

Abstract. The newest video coding standard H.264/AVC provides multiple reference frames motion estimation in the spatial region, and the optimal frame is selected by RDO (Rate Distortion Optimization) with high coding complexity. However, the coding efficiency only depends on the attribute of sequences, not on the number of reference frames. In this paper, statistical characteristics of the best reference frame with variable block size are studied, and a fast algorithm that takes into account the correlation is proposed. The reference frame of block mode may be chosen based on the computing result of the above block mode. Experimental results show that with similar Distortion performance, the algorithm can efficiently reduce the computational complexity by 19% averagely.

1 Introduction

The newest video coding standard H.264/AVC [1] is developed by the Joint Video Team (JVT) which was organized by ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG) in 2001, which can typically outperforms all existing standards. H.264/AVC is similar to other standards such as MPEG-4 Video, which consists of a hybrid of temporal and spatial prediction, in conjunction with transform coding. But H.264 includes a number of new techniques such as variable block size, enhanced intra/inter prediction, 4×4 integer transform, adaptive in-loop deblocking filter, refined motion-compensated prediction, and new entropy coding, etc. Compared with the H.263 and MPEG-4(advanced simple profile), H.264/AVC can reduce 40%~50% bits-rate while keeping the equivalent video quality. However, the compression performance comes at a high computational cost [2].

In order to enhance the compression efficiency of P type frame, the motion estimation in H.264/AVC uses variable block size and multiple reference frames, which can greatly reduce prediction errors and obtain better performance. Reference software of H.264 adopts full search mode for each encode size block in every reference frame, and the optimal result is selected based on RDO, which contributes to heaviest computational load. To satisfy the requirement of real time, studying the fast algorithms how to reduce the code complexity becomes a key issue for specific encoder/decoder

* This work was supported by the NSFC (60573150), the National Defense Basic Research Foundation, and the Program for New Century Excellent Talents in University, and the research was made in the State Key Lab of Software Development Environment.

I. King et al. (Eds.): ICONIP 2006, Part II, LNCS 4233, pp. 458 – 465, 2006.

applications. Currently the research and implementation work mainly focus on mode decision process and achieves fairly good results. The main idea is as follows: forecasting the most coding mode based on the nature of sequences (Movement, venation, etc.); then using effective threshold value mechanisms for early withdrawal, which thereby reduces predictive modes and improves coding speed. In [3], candidate modes for current macroblock are first inferred from given coded adjacent macroblocks by adopting motion information and ratios of defined mode, and the final selection is made by a RDO approach. In [4], the threshold value is dynamically updated for each block mode in order to stop prediction quickly and correctly. Similar ideas are also explored in [5~6]. However, it can be seen that the computation is in proportion to the number of search frames, so it is necessary to reduce the multiple reference frames number. In [7~8], a fast motion estimation algorithm is proposed that takes into account the correlation of motion vectors in multiple frames, and a minor search windows is needed. [9] proposed a new idea that several conditions are used to decide whether it is necessary to search more reference frames. But algorithm simply adopt full search when the rest reference frame is beneficial.

A fast selection algorithm for multiple reference frames (FSAMR) in H.264/AVC is proposed in this paper, which reduces the encoding time by 19% averagely and can be combined with other methods such as [3~6] to further improve the speed. The paper is organized as follows: Section 2 briefly introduces the mode decision algorithm of multiple reference frames in H.264/AVC and gives the benefits of multiple reference frame prediction. In section 3, we analyze the statistical characteristics of the best reference frame among variable block size, and describe the details of our proposed fast algorithm for multiple reference frames selection. Finally, experiment results and concluding remarks are given in Section 4 and 5, respectively.

2 Overview of Multiple Reference Frames Prediction

2.1 Description of Multiple Reference Frames Prediction

H.264/AVC standard has extended the block based motion compensation by introducing tree structured variable-block size to approximate the shape of the moving objects within the MB more accurately. The size of a block can be 16×16 , 16×8 , 8×16 and 8×8 for motion compensation. In case 8×8 size is chosen, it can be further divided into smaller block size 8×4 , 4×8 and 4×4 . Besides the seven different sizes, an inter macroblock can also be coded in the Intra mode (Intra 4×4 and Intra16×16) and so-called SKIP mode. For this mode, neither a quantized prediction error signal, nor a motion vector or reference index parameter, has to be transmitted. H.264/AVC also supports multi-frame motion-compensated prediction. That is, more than one prior-coded frame can be used as a reference for motion-compensated prediction. The reference software of H.264/AVC JM94 performs full search to find the motion vector for each block in different sizes from previous one to five reference frame, shown as Fig.1 .

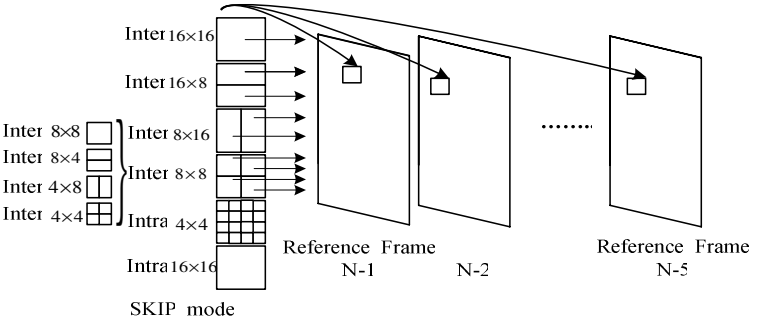

Fig. 1. Variable block size and multiple reference frames motion estimation

H.264/AVC selects the best mode and reference by using the RDO, which means that the final decision is made by minimizing the Lagrange formula (1):

$$J_{MODE} = D_{MODE}(ref) + \lambda \times R_{MODE}(ref) \tag{1}$$

Where ref being the reference frame position; J_{MODE} being R-D cost of the corresponding mode; λ being the Lagrange multiplier; R_{MODE} being the total bits-rate including the motion vectors, block mode, all transform coefficients, etc; D_{MODE} being the distortion between original frame and reconstructed frame.

The above scheme can gain better coding efficiency, but the RD cost of every mode and reference frame should be computed based on the actual rate and distortion, which are obtained only after compression and decompression. So the transform/inverse transform, quantization/inverse quantization, entropy coding have to be used repeatedly, and the complexity of computation is enhanced notably.

2.2 Benefits of Multiple Reference Frame Prediction

The video compression standard such as H.263 and MPEG-4, use single decoded frame as the reference frame, which can achieve better predictive results in most cases, except that there are some non-compensation regions in some special circumstances. However, multiple reference frames can gain better prediction result. The major reasons are described as follows, and the further details can be referred to [7,9]:

1) Due to the repetition of motion, objects or veins may have a better appearance at previous several frames than the latest reference frame.

2) Some parts of the objects or background may be covered by a moving object, which happens in many sequences, the hidden parts can not find a proper match in the latest reference frame, and may be found in the previous pictures when they were uncovered.

3) The shake and telescopic of the camera will lead to a rapid scene switch, and the same object position is in the difference reference frame.

4) Other reasons, why motion estimation of multiple reference frame get better performance than single reference, include the change of lighting and shadow, the sampling of picture, etc.

3 A Fast Multiple Reference Frames Selective Algorithm

3.1 Analyze of Multiple Reference Frames Prediction

In order to statistic and analysis of multiple reference frames motion estimation, we did some experiments on different sequences based on the H.264/AVC reference software JM94. we selected six typical sequences which are QCIF format (176×144) provided by MPEG standard: type A sequence, Mthr_dotr, Container with simple veins or slow movement; type B sequence, Foreman, Coastguard with middle veins or movement; type C sequence, Mobile, Bus with complex veins or intense movement. Table 1 lists the main parameters, and the conditions for all tests will be consistent.

Table 1. Test conditions

UseHadamard	On
SearchRange	16
SymbolMode	UVLC
ReferenceFrames	5
LoopFilter	On
AllMode	On
RDOptimization	On

We encoded 15 frames for each sequence, the structure of GOP (Group of picture) is IPPP (I type frame and P type frame), and the fixed QP is set to 28. Table2 shows the experimental result of coding efficiency with different reference frames. Compared with the single reference frame, ΔPSNR represents the benefits in luminance PSNR and ΔBits (%) denotes the percent of bit-rate variety with further reference frames. From Table2, we can conclude that the coding efficiency only depends on the nature of sequences, not on the number of reference frames. Generally, most benefits depend on the previous two reference frames obviously. We can also find that with the increased number of reference frames, the coding efficiency gain little, but the computational complexity increased sharply.

The mode decision result after motion estimation and intra prediction is also a very important cue [9].In Table3, in the expression A|B, A represents the possibility of a

Table 2. Comparison of the encode with different reference frame

	2 Reference	3 Reference	4 Reference	5 Reference
		ΔPSNR(dB)|ΔBits(%)		
Mthr_dotr	+0.048| +0.7	+0.059 | +0.9	+0.059|+0.9	+0.099|+1.7
Container	-0.028 | -6.1	+0.015 | -12.6	+0.025|-21.5	+0.049|-19.8
Foreman	+0.081| -3.0	+0.134 | -2.4	+0.152|-1.8	+0.174|-1.7
Coastguard	+0.082| +0.6	+0.092 | +0.5	+0.104|+1.8	+0.099|+1.4
Mobile	+0.069| -6.9	+0.129 |-13.5	+0.180|-16.5	+0.189|-19.0
Bus	+0.122| -2.6	+0.140 | -4.3	+0.158|-4.5	+0.177|-4.1

mode that can be chosen after the latest reference frame estimation, B is the possibility of A mode that can keep unchanged after 5 frame searched. Compared with [9], we added the analysis of SKIP mode. From Table3, we can see that: 53% of macroblocks need the latest reference frame; furthermore, when macroblock is split into smaller block size of 8×4, 4×8 and 4×4, there will be a better match on other reference frames; if macroblock adopt 16×16 mode or SKIP mode, there is simply circumstance and no further search is needed; Intra mode is seldom used. Summarily, the multiple reference gains better prediction for some special non-compensation region, and the efficiency depends on the nature of sequences such as veins and movement.

Table 3. Comparison of the encode with different reference frame

	SKIP	16×16	16×8	8×16	8×8	Intra
Mthr_dotr	51 | 90	19 | 54	11 | 73	08 | 49	09 | 62	2 |100
Container	80 | 95	10 | 53	03 | 31	05 | 31	02 | 44	0 | 0
Foreman	31 | 65	27 | 52	10 | 36	17 | 47	15 | 53	0 | 0
Coastguard	19 | 51	40 | 68	14 | 37	12 | 40	15 | 54	0 | 0
Mobile	06 | 25	27 | 42	11 | 28	10 | 34	46 | 58	0 | 0
Bus	08 | 51	36 | 54	17 | 41	10 | 39	28 | 78	1 | 77
Average	33 | 53	27 | 54	11 | 54	11 | 40	18 | 58	0 | 30
	33×53+27×54+11×54+11×40+18×58=53%					

Now we try to find out the correlation of variable block sizes. After motion estimation with 5 reference frames, we can get one best reference for 16×16 block mode, two best references for 16×8 block mode, and two best references for 8×16 block mode. From Table4 result, we can get some very useful information that about 84.5% blocks of 16×8 and 8×16 mode have the same reference frame which is consistent with 16×16 mode, and the percentage in 8×8 mode and further smaller block sizes is 89.8%.

Table 4. Correlation of the best reference frame among variable mode

	16×16 | 16×8 and 8×16	8×8 | 8×4 and 4×8
Mthr_dotr	88.8%	91.7%
Container	95.6%	97.5%
Foreman	81.2%	89.1%
Coastguard	86.7%	90.8%
Mobile	76.2%	83.5%
Bus	78.4%	85.9%
Average	84.5%	89.8%

3.2 Description of Fast Reference Frame Decision

Based on the above statistic and analysis, we propose a fast multiple reference frames selection algorithm for H.264/AVC, which is composed of the following steps:

Step 1: Perform the 16×16 block mode motion estimation referring to previous one to five reference frames, and obtain the best reference frame, noted as F_{16}.

Step 2: Do motion estimation on size of 16×8 and 8×16, and only the latest reference frame and F_{16} frame need to calculate the R-D cost.

Step 3: If 8×8 size is chosen, it can be further divided into smaller block size 8×4, 4×8 and 4×4, and a macroblock will loop four times for sub-macroblock mode. Perform the 8×8 block mode motion estimation for each reference frame, and obtain the best reference frame F_8 and R-D cost. $J_{8\times8}$, respectively.

Step 4: Calculate the R-D cost $J_{8\times4}$ and $J_{4\times8}$ of the latest reference frame and F_8 frame.If $J_{8\times8} > J_{8\times4} / J_{4\times8}$, the 4×4 mode block search not only the latest reference frame, but also reference frames between F_8 and available furthest reference frame. Otherwise select the best reference between the latest reference frame and F_8.

Step 5: If all block mode have been processed, then process the next macroblock, otherwise jumps to step 3.

In the steps above, only the reference frame selection is modified during motion esti-mation, and it can be integrated with other fast algorithms to reduce complexity further.

4 Experimental Result and Discussions

The proposed FSAMR has been implemented based on JM94. The sequences and encoder condition are the same as shown in section 3.1. We encode 60 frames for each sequence, the structure of GOP is IPPP, the frame rate is 15f/s and the fixed QP value is set to 28. Peak signal noise ratio (PSNR), total motion estimation time, and total bits-rate of P-type frame are used as measurement. The results achieved by FSAMR and full search algorithm are presented in Table 5 and Figure 2. ΔPSNR represents the difference in luminance PSNR, ΔBits and ΔTime is defined as the formula (2):

$$Ratio = \frac{T_{pro} - T_{full}}{T_{full}} \times 100\% \tag{2}$$

Where T_{full} and T_{pro} denote the result of full search and FSAMR, respectively.

Table 5. Comparison results

Sequence	ΔPSNR(dB)	ΔBits (%)	ΔTime (%)
Mthr_dotr	-0.022	+0.26	-20.8
Container	-0.022	-0.21	-20.0
Foreman	-0.037	+0.20	-19.2
Coastguard	+0.001	+0.17	-19.4
Mobile	-0.012	+1.70	-15.0
Bus	-0.024	+0.73	-19.6

As shown in Table5, our proposed algorithm reduces the computational complexity by 19%, meanwhile PSNR only decreases 0.02 slightly, and bits-rate increases only 0.47%, averagely. Besides, it can be seen that algorithm has a high content correlation between image sequences and FSAMR. Because the FSMAR algorithm uses statistical characteristics of the best reference frame among variable block size. Generally, simple veins or slow movement sequence has the stronger relativity, and this algorithm gains the better benefits; conversely, the effect of the fast reference frames selection algorithm may decrease.

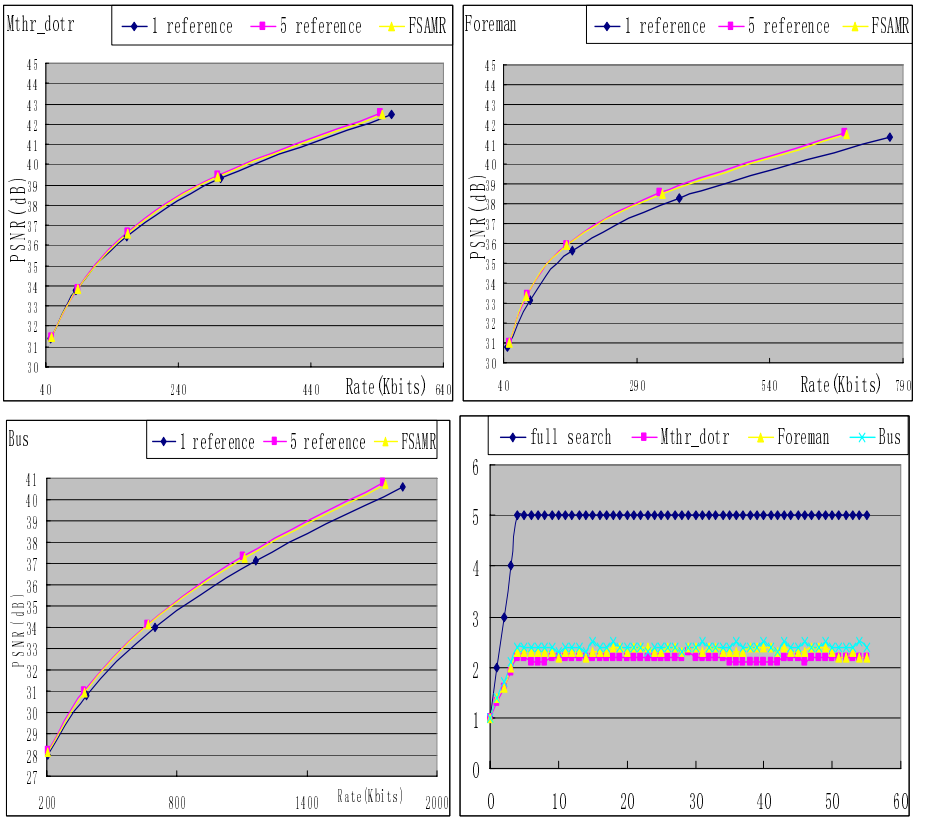

Fig. 2. Rate distortion curves and average searched frame

Fig.2 is the rate-distortion curves and average searched frames between 5 reference frames with full search, 1 reference frames with full search and 5 reference frames with FSAMR for different sequences. It is shown that compared with 5 reference frames with full search, FSAMR can efficiently reduce the number of searched reference frames with similar R-D, and the number of searched frames is no more than 3.

5 Conclusions

In this paper, we propose a new a fast selection algorithm, called FSAMR, for multiple reference frames in H.264/AVC. It is based on an analysis of statistical characteristics of the best reference frame among variable block size. The reference frame of block mode may be chosen based on the computing result of the above block mode. Experimental results show that compared with 5 reference frames search method, the algorithm can efficiently reduce the computational complexity, and meanwhile the degradation of the reconstructive video quality and the increase of the bits-rate are controlled under a reasonable level. Besides, this algorithm can be combined with other methods such as [3~6] to further improve the speed. How to perform a fast mode selection will also be our further work.

References

1. ISO/IEC FDIS 14496-10. Information technology-Coding of audio-visual objects Part 10: Advanced video coding[S]. Final Draft International Standard, 2003.
2. RavasiM, MattavelliM, Clerc C A. Computational Complexity Comparison of MPEG4 and JVT Codecs[S]. JVT-D153r1-L, Joint Video Team of ISO/IEC MPEG&ITU-T VCEG, Klagenfurt, Austria, 2002.
3. Zhu Hong, Wu Chengke, Fang Yong, A Novel Scheme for Fast Mode Decision within H. 264[J], ACTA Ectronica Sinica, 2005, 33(9):99-103(in Chinese).
4. Shen Gao, Tiejun Lun, An Improved Fast Mode Decision Algorithm in H.264 for Video Communications[A], ISSCAA2006 [C], Harbin,China,2006,57-60
5. KWON G, LEE J, YUN J, et al. Fast Inter-prediction method for mobile video communications using H.264/AVC[A]. International Conference on Consumer Electronics 2005[C]. Los Angeles, USA, 2005. 227-228.
6. Xiang Dong, Zhou Jingli, Yu Shengsheng, et al. Macroblock Coding Mode Predictive Method Based on Spatio-Temporal Correlation[J], Mini- Micro Systems ,2006, 27(1):101-103(in Chinese).
7. Ye-Ping Su, Ming-Ting Sun, Fast Multiple Reference Frame Motion Estimation for H.264/AVC[J], IEEE Transactions on CSVT, Accepted for future publication Volume PP, Issue 99, 2006 Page(s):1 - 1.
8. Mei-Juan Chen, Yi-Yen Chiang, Huang-Ju Li, Ming-Chieh Chi, Efficient Multi-Frame Motion Estimation Algorithms for MPEG-4 AVC/JVT/H.264[A], IEEE ISCAS 2004[C], Vancouver, Canada, May 2004,737-740
9. Yu-Wen Huang, Bing-Yu Hsieh, Tu-Chih Wang, Shao-Yi Chien, Analysis and Reduction of Reference Frames for Motion Estimation in MPEG-4 AVC/JVT/H.264[A], IEEE ISASSP 2003[C], Hong Kong, April 2003, 145-148