AN ADAPTIVE FAST MULTIPLE REFERENCE FRAMES SELECTION ALGORITHM FOR H.264/AVC

Peng Wu, Chuang-Bai Xiao

College of Computer Science and Technology,
Beijing University of Technology, Beijing, China

ABSTRACT

To make full use of the temporal correlation of video sequences, H.264/AVC adopts multiple reference frames to enhance the coding quality. While the performance being improved, the complexity of computation has been increased linearly, too. In this paper, we study motion characteristic of video sequences and spatial correlation in video frames first, and then we propose an adaptive fast multiple reference frames selection algorithm. It can decrease the number of reference frames for motion compensation, and reduce the complexity of coding adaptively according to the features of video sequences. The results show that our algorithm achieves 45% coding time saving on average with unnoticeable quality loss.

Index Terms—video coding, H.264/AVC, multiple reference frames, adaptive coding, motion compensation

1. INTRODUCTION

H.264/AVC [1] is the newest video coding standard released by Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG in 2003. By contrast with H.263 and MPEG-2, H.264/AVC introduces some new technical developments, such as intra prediction, variable block-size motion compensation with small block sizes, weighted prediction, multiple references frame motion compensation (MRF-MC) etc. While these methods enhance the coding performance of H.264 in both PSNR and bit-rate, they also make the coding complexity increase significantly at the same time.

Motion Estimation (ME) and Compensation (MC) is one of the most time-consuming parts in H.264/AVC, lots of algorithms have been proposed to improve its performance. UMHexagonS [2] and EPZS [3] are their representatives which are embodied by the reference software JM. They can reduce the coding complexity without sacrificing the quality.

But these improvements are only for single reference frame MC (SRF-MC). In order to get more accurate result of ME, H.264/AVC allows video blocks to use more than one reference frames for ME and MC. This scheme upgrades the coding quality and decreases the bit-rate efficiently at the expense of lengthening the coding time linearly. After many years of research on this area, improved methods are emerging [4-7]. Among them, Shen et al. [7] put forward an adaptive and fast multiframe selection algorithm (AFMFSA) that uses the correlation among the neighboring blocks and information of ME from previous searched reference frames. It is proved to be effective, but they did not take the various block modes into consideration. In fact, blocks with different size have distinct spatial and temporal correlation.

In this paper, we study the multiple reference frame technology further, and propose a novel adaptive fast multi-frames selection algorithm. Our scheme can select reference frames by using the characteristic of video sequences. It can reduce the computation on ME and MC effectively without the loss of PSNR.

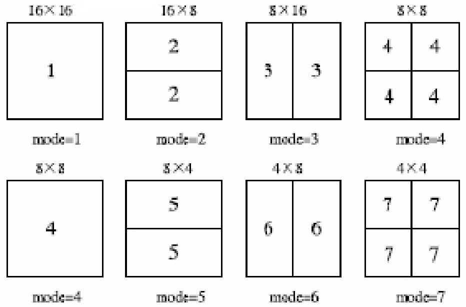

Fig.1 Macroblock and sub-macroblock partitions for motion estimation in H.264/AVC

2. OBSERVATION ON MULTI-FRAME MOTION COMPENSATION FOR H.264/AVC

2.1 MRF-MC of H.264/AVC

To make full use of temporal redundancy of video sequences, H.264/AVC adopts MRF-MC technology. There are seven block mode (Fig.1) for inter prediction in H.264/AVC. For each mode, a block needs to use N frames for MC, and then selects the frame which can minimize the value J in cost function (1) as the best reference frame of current block.

$$J\ (mv \mid ref\ ,\ \lambda_{\mathrm{MOTION}}) = SAD(src, des(mv \mid ref)) + \lambda_{\mathrm{MOTION}} \bullet R((mv - pmv) \mid ref) \quad (1)$$

This work was supported by Beijing Natural Science Foundation（4072004）and Scientific Research Common Program of Beijing Municipal Commission of Education (KM200510005012).

1-4244-1484-9/08/$25.00 ©2008 IEEE

with *ref* being the current reference index number, *mv*=(mv$_x$, mv$_y$)^T being the motion vector in reference *ref*, *pmv*=(pmv$_x$, pmv$_y$)^T being the prediction for motion vector, λ_{MOTION} being the Lagrange multiplier. R(·) denotes the bits used for encoding the reference index number and the difference between *mv* and *pmv*. The *SAD* which means sum of absolute differences is depicted as follow:

$$SAD(src, des(mv \mid ref)) = \sum_{x=1,y=1}^{B1,B2} \mid src(x,y) - des(x - mv_x, y - mv_y \mid ref) \mid \quad (2)$$

with B1, B2=4, 8 or 16, representing the width and length of a block respectively. *src* being the original block value, *des* being the coded block value which has a distance of *mv* from original location in reference *ref*.

Contrary to SRF-MC, the computing time needed by MRF-MC has been lengthened linearly. If the max number of reference frames (*MAXREF*) is equal to five, the time for MC is five times as much as before. Su et al. [8] explained why exploiting MRF can promote the coding performance in detail. But it does not mean that every video block should use as much as *MAXREF* frames to do the MC. Exploiting MRF-MC with fixed number of reference frames directly without any optimization is inefficient and it is difficult to satisfy many applications for the real-time requirements.

2.2 Observation on Distribution of Best Reference Frame

Table I shows us the statistical probability distribution of best reference frame, the efficiency of MRF-MC can be seen through *NACT* defined in (3).

Table I
Distribution of best reference frame of various sequences

Sequence	P_{ref0}(%)	P_{ref1}(%)	P_{ref2}(%)	P_{ref3}(%)	P_{ref4}(%)	*NACT*
Mobile	48.1	14.9	14.7	11.6	10.7	2.2
Tempete	53.3	14.2	13.8	10.1	8.6	2.1
Waterfall	71.0	10.7	7.6	6.1	4.6	1.6
Foreman	75.1	10.1	6.2	4.5	4.2	1.5
Paris	91.4	3.9	2.1	1.4	1.2	1.2
News	93.2	2.7	1.6	1.3	1.2	1.1
Average	72.2	9.2	7.6	5.8	5.2	1.6

$$NACT = \sum_{x=0}^{MAXREF-1} (x+1) * P_{refx} \quad (3)$$

P_{refx} represents the probability of choosing *refx* as the best reference frame. Using the data in table I, we can compute *NACT* which means the expected number of reference frames used by video sequences. In other words, if the sequence wants to get the best coding performance, *NACT* reference frames are need on average, and extra frames may not get more gain. The greater the *NACT* value, the more efficient of using MRF-MC. Conversely, less *NACT* refers to lower efficiency.

We can also find from table I that best reference frames are apt to be the neighbor of current frame. Overall, there is $P_{ref0} + P_{ref1} > 80\%$ for all video sequences. It is decided by the temporal correlation of video sequences.

2.3 Observation on Distribution of Mode Decision

Table II shows the distribution of mode decision of different sequences, including the mode0 (skip mode). Due to the proportion of INTRA mode in P frames is so small, we do not take it into consideration. In the sequences which have large static areas and smooth motion, such as News and Paris, their Macroblocks (MBs) are prone to select mode0 (skip mode) or large partition size mode (16x16,16x8 or 8x16) as their best coding mode. On the contrary, the sequences which have dramatic movement or scene change frequently, such as Tempete and Mobile, smaller block size (8x8, 8x4, 4x8 or 4x4) will be more suitable for them to describe the complicated motion details.

Table II
Distribution of best coding mode of various sequences

Sequence	P_{mode0}(%)	P_{mode1}(%)	$P_{mode2-3}$(%)	$P_{mode4-7}$(%)
Mobile	3.4	23.4	27.1	46.1
Tempete	11.1	26.4	27.7	34.8
Waterfall	31.2	38.9	23.3	6.6
Foreman	32.0	28.9	25.2	13.9
Paris	60.5	14.3	7.6	17.6
News	75.8	11.1	6.5	6.6

We can draw two conclusions from table I and table II as follows: Firstly, the MBs coded in large partition size modes are more inclined to choose neighbor frames as their best reference frame than the ones encoded in P8x8 modes; Secondly, sequences that most of whose MBs are encoded in large partition size have stronger temporal correlation than other sequences.

2.4 Observation on Correlation of Neighboring Regions

Besides temporal correlation, the spatial correlation of neighboring regions is also an essential characteristic of video sequences. Natural sequences are usually composed of various independent objects with different range of movement. The regions which belong to the same object often have similar motion vectors (MVs). In other words, these regions are in same direction and homologous motion extent. The smaller the blocks are, the stronger correlation of neighboring regions is. On the contrary, large size blocks are easier to contain more than one motion objects, and it's hard to predict current block's motion status by coding information of adjacent blocks.

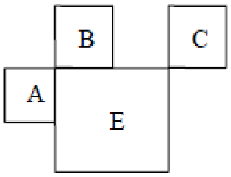

Fig.2 Current block E and its neighbors A、B、C

For utilizing the regional correlation sufficiently, we firstly need to judge whether it exists the correlation in current area or not. So we put forward a criterion of regional correlation. Fig.2 shows current block E and the location of its three adjacent blocks A, B and C. Because A, B and C have been coded already, we can obtain MV_A, MV_B and MV_C which denote the best MV of the three 4x4 blocks respectively. The proposed judging criterion is depicted as follows:

$$\sum_{i=x,y} | MV_E(i) - MV_S(i) | < \mathrm{T1} \quad S \in \{A, B\ or\ C\} \quad (4)$$

MV_E represents the MV of current block E gained after the ME in *ref0*. If this criterion is satisfied, we can deem that block E has correlation with relevant neighboring region. Threshold T1 reflects the strength of regional correlation. Smaller T1 signifies stronger regional correlation, and vice versa. T1 is also a compromise of coding quality and complexity. Although smaller T1 will bring better coding quality, it will cause higher complexity at the same time. We set T1=3 in this paper.

3. PROPOSED FAST SELECTION ALGORITHM

Based on the analysis above, we propose a fast MRF selection algorithm which takes advantage of the sequences' own characteristics.

3.1 Fast MRF Selection of Large-Size Block Modes

For MBs encoded in large partition size (16x16, 16x8, 8x16) modes, temporal factor usually has more significant effect than spatial factor. In this case, we consider the redundancy between current frame and past frames plays an important role in the selection of MRF.

For the case that the *MAXREF* is five, compute P_{mode} of the nearest five coded pictures. We define MRatio as the sum of P_{mode0} and P_{mode1}. As MRatio reflecting the temporal correlation to some extent through the observation of sections 2.2 and 2.3, we make it determine the number of reference frames M used for ME. One can adjust the performance of this algorithm by modifying the range of MRatio of each level. In this paper, we set:

$1 \geq$ MRatio ≥ 0.9, M=1; $0.9 >$ MRatio ≥ 0.7, M=3;
$0.7 >$ MRatio ≥ 0.5, M=4; Else M=*MAXREF*;

It is noticeable that there are several statuses will affect this approach a lot such as coding for complicated texture, periodic motion, and local scene change. All situations could cause the decline of corresponding blocks coding performance obviously. So we add a measure to keep the quality.

$$J_{blockE} \leq \mathrm{T2} \quad \mathrm{T2} = Max(J_{blockA}, J_{blockB}, J_{blockC}) \qquad (5)$$

In (5), $Max(\cdot)$ is the max function and J_{blockE} represents the minimum J of current block E after M reference frames' MC. A, B and C are the neighbors have the same block size with E here. If J_{blockE}>T2 after M frames' MC, we think the result is not satisfying and more reference frames are needed.

3.2 Fast MRF Selection of Smaller-Size Block Modes

P8x8 mode (8x8, 8x4, 4x8 or 4x4) has smaller block size which means it owns stronger regional correlation than large partition size block modes as discussed above. We can eliminate unnecessary reference frames from this character.

In view of the preceding analysis, we find the fact that the *ref0* is more likely chosen as the best reference frame than others. So the *ref0* is perceived as an obligatory reference frame used for MC. Acquire the MV of block E (MV_E) after do ME in *ref0*, and then we could judge whether there is regional correlation between block E and its adjacent blocks respectively by (4). If the formula is satisfied, it is considered that the two blocks may have strong regional correlation and the best reference frame of them may be identical. We set the flag bit $\mathrm{FLAG_S}$=1 (S=A, B or C); get best reference frame number of corresponding block S ($\mathrm{REF_S}$); append it into the candidate pool of reference frame (CPRF).

In the situation that current block E and all of its adjacent blocks do not meet the relevant criterion or all the neighbors are unavailable, we think E may be at the edge of different motion objects, and all reference frames should be used for MC to get more accurate ME results. Besides this situation, using the frames in the CPRF to do the MC is enough. Extra frames may not bring any more gain.

3.3 Summary of Proposed Algorithm

The process of proposed algorithm is summarized as follows:

Step 1) Compute neighbor reference frames' statistical distribution of mode decision.

Step 2) For each mode in large-size block modes, set number of reference frames M by MRatio.

Step 2.1) Do M reference frames' MC.

Step 2.2) If J_{blockE} > T2 and M < *MAXREF*, M=M+1, do MC in the (M-1) th reference frame, go to Step2.2, else go to Step3).

Step 3) For each mode in P8x8 modes:

Step 3.1) Check the regional correlation between current block and its neighbors with formula (4). If none of them satisfy the correlation criterion or are available, go to Step 3.3).

Step 3.2) Construct the candidate pool of reference frame according to the judgement of regional correlation. Go to Step 3.4).

Step 3.3) Using all reference frames to do the ME and MC. Go to Step 4).

Step 3.4) Using frames in CPRF to do the ME and MC.

Step 4) End

4. EXPERIMENT RESULTS

The proposed algorithm was tested on 6 CIF (352x288, 30Hz) sequences which represent various glass of motion, including 150 frames of Mobile, Tempete, Waterfall, Foreman, Paris and News. The scheme is implemented based on reference software JM11.0. The detailed simulation settings are as follows: Microsoft Windows XP, P4 2.4GHz CPU, 512M RAM. RDO is set to low complexity; search range is 32; number of reference frames is 5; QP=28; using P-frames coding only with first I-frame; rate control is off; entropy coding method is set to CABAC; ME exploits the fast search algorithm UMHexagonS.

Fig.3 depicts the rate-distortion (RD) performance of sequence Mobile and Foreman ($QP \in \{20,24,28,32,36\}$), with cross representing conventional five reference frames algorithm; triangle representing proposed five reference frames algorithm; pentacle representing single reference frame algorithm (SRFA). Table III makes comparison between the proposed algorithm and the conventional algorithm (CA) with full reference frames. The experimental results show that our algorithm achieves 45% coding time saving on average contrast with CA. Meanwhile, the performance of our algorithm is almost the same as the original, with negligible loss in PSNR and bit-rate.

Table III Performance comparison between CA and proposed algorithm

Sequence	△PSNR(db)	△Bitrate(%)	Tsaving(%)
Mobile	-0.05	0.30	36.7
Tempete	-0.03	0.36	39.4
Waterfall	-0.10	-0.69	53.1
Foreman	-0.07	-0.38	43.1
Paris	-0.02	0.56	50.7
News	-0.03	0.39	51.3
Average	-0.05	-0.04	45.7

$$\text{Tsaving}=\frac{\text{time}_{\text{CA}}-\text{time}_{\text{Proposed}}}{\text{time}_{\text{CA}}}*100\%$$

Table IV Number of reference frames in use actually

Sequence	CA	AFMFSA	Proposed
Mobile	5	3.30	2.55
Tempete	5	2.96	2.56
Waterfall	5	2.65	1.80
Foreman	5	2.10	2.00
Paris	5	1.74	1.62
News	5	1.50	1.48
Average	5	2.38	2.00

Table IV shows the average number of used reference frames for CA, AFMFSA [7] and our proposed algorithm, respectively. It can be seen our scheme can save about 60% computation load of ME compared with CA on average; and our method is also more effective than the AFMFSA, especially for sequences with complicated motion, such as Mobile and Tempete.

5. CONCLUSION

In this paper, we propose an adaptive fast reference frame selection algorithm based on motion correlation. The impact of temporal and spatial motion correlation on reference frame selection is distinct for different block modes. We take use of spatial correlation to reduce the number of reference frames for large size modes, while utilizing spatial correlation to get rid of the unnecessary ones for small size modes. Experimental results show that our approach can reduce the coding complexity notably while achieves similar gain as conventional full reference frames algorithm.

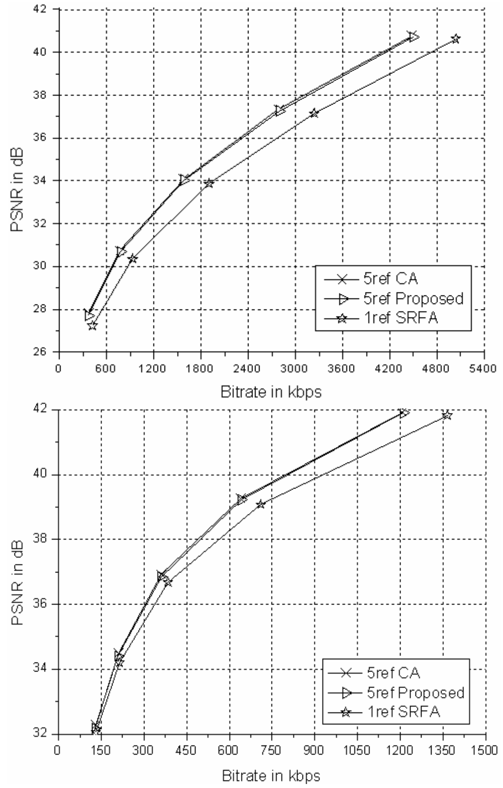

Fig.3 RD curves of Mobile (up) and Foreman (down)

6. REFERENCES

[1] 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 (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC)," March 2003.

[2] Z.B.Chen, P. Zhou, Y. He, "Fast integer pel and fractional pel motion estimation for JVT," JVT-F017.doc, 6th Meeting, Awaji, Japan, Dec.5-13, 2002.

[3] A.M.Tourapis, O.C.Au, M.L.Liou, "Highly efficient predictive zonal algorithms for fast block matching motion estimation," IEEE Trans.Circuits Syst. Video Technol., vol.12, pp.934-947, Oct. 2002.

[4] A.Chang, O.C.Au, Y.M.Yeung, "A Novel Approach To Fast Multi-Frame Selection For H.264 Video Coding," Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). vol.3, pp.413-416, April.2003.

[5] H.J.Li, C.T.Hsu, M.J.Chen, "Fast multiple reference frame selection method for motion estimation in JVT/H.264," Circuits and Systems, 2004. Proceedings. The 2004 IEEE Asia-Pacific Conference, vol.1, pp.605-608, Dec.2004.

[6] S.Saponara, M.Casula, F.Rovati, D.Alfonso, "Dynamic Control of Motion Estimation Search Parameters for Low Complex H.264 Video Coding," IEEE Trans. Consumer Electronics, vol.52, pp.232-239, Feb.2006.

[7] L.Q.Shen, Z.Liu and Z.Y.Zhang et al, "An adaptive and fast multiframe selection algorithm for H.264 video coding," IEEE signal processing letters, vol. 14, pp.836-839, Nov 2007.

[8] Y.P.Su and M.T.Sun, "Fast multiple reference frame motion estimation for H.264/AVC," IEEE Trans. Circuits Syst.Video Technol., vol.16, pp.447-452, Mar 2006.