Multi-view Video Coding Based on fast inter mode selection

Sen Wang[1], Yingyun Yang[2],Huabing Wang[3]

Abstract. Multiview video coding mode selection is divided into inter mode selection and intra mode selection, This paper by calculating the adaptive threshold and the complexity of inter-frame motion, proposed a fast inter mode selection algorithm based on the motion complexity and texture information. This algorithm makes use of inter macroblocks spatio-temporal correlation, Early analysis of the movement complex of the frame around, reducing the alternative models. According to the texture information on pattern further screening, thus greatly improving the pattern search speed, while the selection of adaptive threshold ensures coding versatility. Experimental results show that compared with JMVC8.5 FastSearch algorithm,the new algorithm can save about 60% of encoding time,while maintain high RD performance.

Keywords: JMVC FastSearch, Fast inter-mode selection, Motion complexity, Texture information.

1 Introduction

Multi-view video (MVC) technology is one of the hot spot of current research . It is widely used in Free-viewpoint TV (FTV), 3DTV, 3D films, virtual reality and so on. However, Opposed to the single view video, multi-view video requires processing more data. MVC is an on-going development of the H.264/AVC video coding standard [1]. Joint Video Team (JVT) has delivered the working MVC as well as its reference software, previously known as JMVM (Joint Multi-view Video Model) [2] and recently updated as JMVC (Joint Multi-view Video Coding) [3].

In JMVC, the best coding mode is determined by exhaustively searching over all possible partition modes using rate distortion comparison in motion estimation for each macroblock[7][8], which results in extremely large encoding times. This paper proposed a fast inter mode selection algorithm based on motion complexity in image region and texture information, unnecessary modes are excluded in advance by calculating the adaptive threshold and the motion complexity,then coding efficiency improved. Compared with the fast algorithm in image-based texture analysis proposed by literature [9] and the fast search algorithm TZS1 proposed by literature [10] ,we can save even more time.

2 Fast Mode Decision

In JMVC, Inter prediction and inter-view prediction spent most of the entire coding time.The inter prediction and inter-view prediction using the same prediction model, that a variety of sizes prediction

1 Sen Wang (✉)
College of Information Engineering, Communication University of China, Beijing, China
e-mail: 377901250@qq.com
2 Yingyun Yang
College of Information Engineering, Communication University of China, Beijing, China
3 Huabing Wang
College of Information Engineering, Communication University of China, Beijing, China

mode, including SKIP, Inter16 × 16, Inter16 × 8, Inter8 × 16, Inter8 × 8Frext, Inter8 × 8, Intra16 × 16, Intra8 × 8 and Intra4 × 4 and other models[5][6]. In order to get a good picture quality and access to balance between the code rate, JMVC used for each prediction mode full traversal methods[4]. While traversing can get the best encoding mode, but the full search method is based on computational complexity increases exponentially at the expense of wasting a lot of time to search for some unlikely to be coding mode block mode. If there are effective fast mode selection algorithms in advance exclude some unnecessary coding mode, will effectively reduce the computational complexity, greatly improve the coding efficiency.

2.1 Adaptive threshold determination method of motion complexity

Since the continuity of the video image content, the majority of zone has a strong time correlation and spatial correlation.According to the motion complexity of adjacent images, stereoscopic video images can be roughly divided into the background area, sub-regional background and foreground region. Background area and complexity background area changes little with time and has a strong spatial correlation, its motion complexity is low for large size segmentation; sport zones changes larger with time, its motion complexity is hihg for small size segmentation, as figure 2.1.Analysis of correlation of the current macroblock and predicted macroblock, calculate the macro block motion complexity.We define a symbol Mmc(motion complexity) that represents the current macroblock motion complexity:

$$\Delta f(x,y) = f(x,y) - f(x',y') \tag{2.1}$$

$$Mmc = \frac{1}{M \times N} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} |\Delta f(x,y) - Average| \tag{2.2}$$

Where f (x, y) represents the luminance of the current frame’s pixel (x, y), f(x',y') represents the luminance of the reference frame corresponding to the pixel (x', y');M × N represents width and height of the current macroblock, is normally set to 16 × 16; Average represents the current macroblock average luminance in residual graph, calculated as:

$$Average = \frac{1}{M \times N} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \Delta f(x,y) \tag{2.3}$$

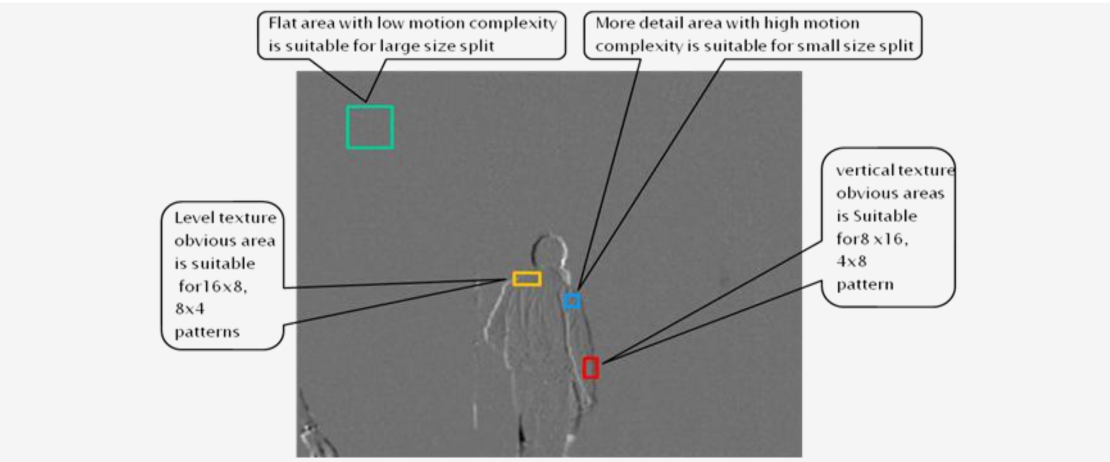

Fig. 2.1 The proposed determination method according to the motion complexity (exit)

In general, most background or sub-background of adjacent images are stationary or almost no movement, so the lower complexity of its movement, Mmc is small;while to the foreground, the higher complexity of its movement, low correlation, Mmc is large; Therefore, by setting a threshold value to determine the macroblock in which region, suppose set two thresholds TH1 and TH2 (TH1 <TH2), Defined as follows: flat background Mmc <T1; complex background T2 <Mmc <T1; foreground Mmc> T2. We define the background is mainly selected pattern set A = {16 × 16}, sub-background is mainly selected pattern set B = {16 × 16,16 × 8,8 × 16}, foreground mainly selected pattern set C = {16 × 16,16 × 8,8 × 16,8 × 8,8 × 4,4 × 8,4 × 4}.

The size of TH1 and TH2 are generally set by the actual experience, but for very different sequence of textures, you need to manually set a different value and then run the encoding process, which gives the actual coding trouble. This paper presents an adaptive threshold method to automatically set TH1 and TH2 according to the image texture , to improve coding practicality. First calculate the average complexity of all macroblock in motion residual figure Avg_mmc:

$$Avg_{mmc} = \frac{1}{H \times W} \sum_{x=0}^{H-1} \sum_{y=0}^{W-1} Mmc \tag{2.4}$$

Where W denotes the number of 16×16 macroblocks in horizontal direction of the motion residual figure, H denotes the number of 16×16 macroblocks in vertical direction of the motion residual figure. We put the brightness difference less than 3 pixels as background region in residual figure, then calculate the background in the entire residual image proportion WEIGHT, then calculate adaptive threshold TH1, then:

$$TH1=Avg\_mmc \times WEIGHT \tag{2.5}$$

Because in an image, the proportion of the background region is the proportion of flat background region X times,then TH2=TH1×(X+1), Experiments in this article we assume that the image sub-background region and background region accounts for a percentage of the ratio of 2, then:

$$TH2=TH1 \times (2+1) \tag{2.6}$$

2.2 According to the texture information narrowed pattern set

When a macroblock in the horizontal direction belong to the same object, the horizontal direction is relatively smooth texture, the texture is relatively complex vertical direction; when one macro block in the vertical direction are the same object, the vertical direction is relatively smooth texture, horizontal texture is relatively complex. Therefore through the texture feature of the macroblock to select the appropriate division pattern: For the size of 16 × 16 macroblock, if the macroblock to be coded belonging to the same in horizontal direction, then the choice of 16 × 8 greater chance of block mode, otherwise select 8 × 16. We define two macro blocks can be expressed the horizontal and vertical direction texture variables:

$$D_h = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} | \Delta f(x,y) - \Delta f(x,y-1) | \tag{2.7}$$

$$D_v = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} | \Delta f(x,y) - \Delta f(x-1,y) | \tag{2.8}$$

Where Dh denotes texture features in horizontal direction; Dv denotes texture features in vertical residual figure the next line corresponding to the brightness value of the pixel, △ f(x-1, y)represents the brightness values before a pixel in the residual figure. When Dh >Dv, the block texture tends to be horizontal, 8×16 mode and 4×8 mode can be excluded in advance; When Dh< Dv the block texture tends to be vertical, 16×8 mode and 8×4 mode can be excluded in advance.

2.3 This paper presents algorithms

The specific steps of proposed algorithm are as follows:

1) According to the macroblock (x', y') of the reference image, the calculation of the current image corresponding to movement complexity Mmc;

2) Adaptive threshold calculation: TH1 and TH2;

3) Compare Mmc with TH1, TH2. a) If Mmc <TH1, the macroblock is in background area, just select {16×16} mode. b) If TH1< Mmc <TH2, the macroblock is in complex background area, select {16×16,16×8, 8×16}. As to complex background area, calculate Dh and Dv. If Dh>Dv, select {16×16,16×8}; else if Dh<Dv, choose {16×16,8×16}; otherwise, choose {8×8Frext，16×16}. c) If the Mmc > TH2, illustrate macro block is in the foreground region, traverse all the interframe modes.

4)Calculate their mode centralized rate-distortion cost of each mode，Choose the optimal mode based on RDcost. Algorithm flow chart is as follows:

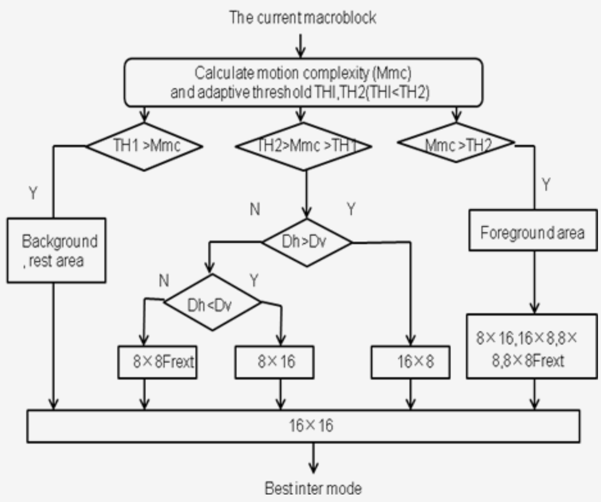

Fig. 2.2 The proposed fast algorithm flowchart

3 Experimental Results

The experiment was based on the configuration of Intel(R) Core(TM) i5-2400 CPU @ 3.10 GHz RAM 2.99GB. Test sequences are ballroom、 Vassar、exit and race1. 100 frames are encoded in each viewpoint. Search mode is fast search, and search range is 16,the GOP length is set as 8. To evaluate the rate distortion performance, respectively in QP for 27,32,37,42 measure PSNR of each viewpoint, the average encoding time and the average bitrates.

In Figure 3.1, JMVC represents JMVC8.5 FastSearch algorithm structure,”improved” represents using improved fast inter mode selection algorithm.

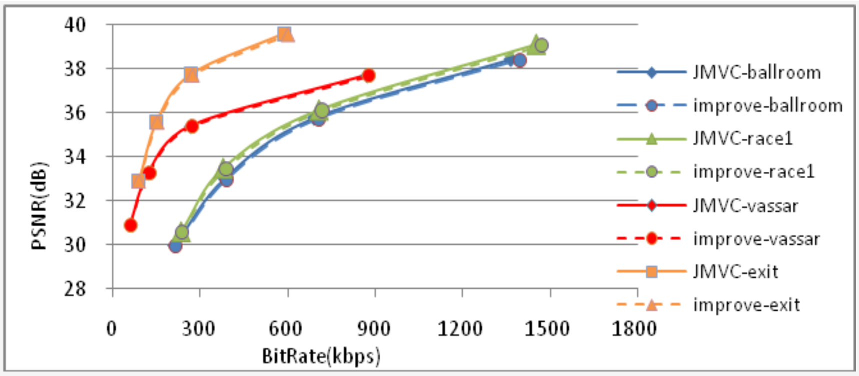

Fig. 3.1 Rate-distortion curve contrast diagram

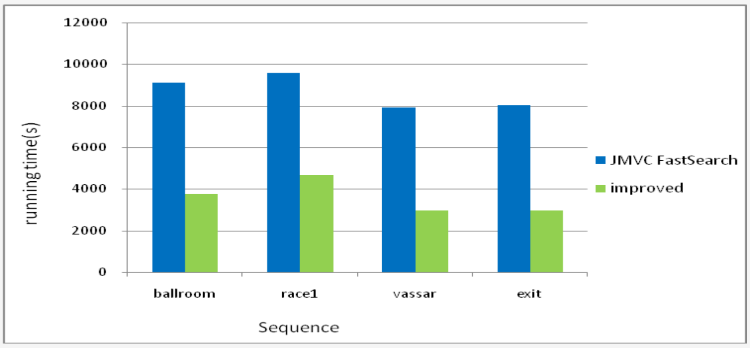

Fig. 3.2 Algorithm coding time comparison chart

Table 3.1 improved algorithm with JMVC original fast algorithm code comparison table

Sequences	improved algorithm compare with JMVC original fast algorithm				
	△ encoding time(%)	△ bitrate (%)	△ YPSNR(dB)	△ UPSNR(dB)	△ VPSNR(dB)
Ballroom	-58.58	1.84	-0.03	-0.038	-0.03
Vassar	-62.63	1.13	-0.014	-0.004	-0.001
Race1	-51.41	1.25	-0.016	-0.01	-0.007
Exit	-62.89	2.19	-0.024	-0.032	-0.022
Average	-58.88	1.60	-0.021	-0.021	-0.015

Table 3.2 improved algorithm compare with the references [9] and [10] proposed an algorithm

Sequences	△encoding time (%)		△average bitrate (%)		△PSNR(dB)	
	improved	references[9]/[10]	improved	references[9]/[10]	improved	references[9]/[10]
ballroom	-58.6	-48.4/--	1.8	1.6/--	-0.030	-0.030/--
exit	-62.9	-53.1/-58.1	2.2	1.7/2.1	-0.024	-0.007/-0.044
vassar	-62.6	-60.3/-56.3	1.1	0.7/1.7	-0.014	-0.025/-0.022

From Figure 3.1, the rate-distortion curve can be seen: the two algorithms rate-distortion curve almost coincide, indicating that the coding efficiency of the improved algorithm compared with the original algorithm is hardly declined. Figure 3.2 and Table 3.1 can be seen in the improved coding algorithm almost no damage to image quality, the encoding time is greatly reduced, has better real-time. For more background area in Vassar sequence encoding time reduction of about 63%; As for the scene change faster race1 sequence encoding time decreases 52% or so. Description of the algorithm for a sequence with more fixed background and fewer moving object has better optimization results, and therefore more suitable for video conferencing, remote video surveillance and other occasions.

From Table 3.2 can be seen that the proposed algorithm in the almost same image quality with literature [9]and [10], we can save more encoding time.

4 Conclusions

This paper makes use of inter macroblocks spatio-temporal correlation,proposed a fast inter mode selection algorithm based on the motion complexity and texture information. Experimental data show that the algorithm PSNR drop is less than 0.03dB, encoding bitrate rose less than 2.19%, almost without a decline in image quality and coding time saves about 60% on average compare with JMVC8.5 FastSearch algorithm. This algorithm is suitable for all kinds of video sequences, and the sequence with more fixed background is better.

Acknowledgment This work is supported by the National Science & Technology Pillar Program of China under Grant No.2012BAH39F02.

References

1. ISO/IEC International Standard 14496-10, Information technology-Coding of Audio-Visual Objects-Part 10: Advanced Video Coding, third edition, Dec. 2005, corrected version, March 2006.
2. A. Vetro, P. Pandit, H. Kimata, A. Smolic and Y. K. Wang, ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16. Joint Multiview Video Model (JMVM) 8.0, Doc. JVT-AA207, Apr. 2008.
3. Joint Video Team, MVC software manual, Version JMVC 8.5 (CVS tag: JMVC_8_5), March 26, 2011.
4. Description of core experiments in MVC,ISO/IEC JTC1/SC29/WG11MPEG2006/W8019,Montreux,April 2006.
5. Liu Z, ShenLQ, ZhangZY. An efficient intermode decision algorithm based on motion homogeneity for H.264/AVC[J].IEEE Transactions on Circuits and Systems for Video Technology200919(1):128-132.
6. Zeng H Q,Cai C H,MaKK. Fast mode decision for H.264/AVC based on macroblock motion activity [J]. IEEE Transactions on Circuits and Systems for Video Technology200919(4):491-499.
7. HE Yong,LIU Gui-hua,H.264 inter mode selection optimization algorithm,VIDEO ENGINEERING, 2007,31(12)
8. HE Jun-qiu,ZHAO Huan,Based motion complexity fast mode selection algorithm,COMPUTER ENGINEERING AND APPLICATION, 2009,45(7)
9. Huabing Wang ,Yingyun Yang,Sen Wang ,Fast Mode Selection Based on Texture Segmentation in JMVC, 2012 14th International Conference on Communication Technology, 2012 (11);
10. Xiuli Tang, Multiview video coding algorithm for fast search mode，Huaqiao University Thesis，2010.1