An Improved Stereo Video Coding Scheme Based on Joint Multiview Video Model

Zhipeng Lin, Yu Xiang, Lin Luo, Li Zhao
School of Information Science and Engineering
Southeast University
Nanjing China
seulzp@163.com

***Abstract*—Stereo video technology is an active research area in current video signal processing filed. More and more researchers are engaged in this field. Based on the Joint Multiview Video Model proposed by JVT (Joint Video Team), three existing stereo video coding schemes are studied in this paper. An improved scheme is proposed after the stereo video MacroBlock types are analyzed. The experimental results show that the encoding complexity of the improved scheme for the right view reduces obviously. Compared to the joint estimation scheme, the encoding performance of improved scheme only reduces a little.**

Keywords-stereo video compression; macroblock type; joint multiview video model

I. INTRODUCTION

Stereo video technology can bring people stereoscopic vision effect in the case of normal vision conditions. It can be applied to many areas such as medical treatment, light field rendering of big molecules, live broadcast of sports and entertainment programs, stereoscopic television and advertisement, games and cartoon, virtual reality, video surveillance and real time reconstruction of stereoscopic battlefield environment. But all the applications mentioned above depend on the effective storage or transferring of the stereo video. A stereo video signal contains two video streams called the left view and the right view respectively, and the quantity of its data is two times larger than that of the normal video. Without the effective stereo video compression technology, it will be impossible to realize the effective storage and transferring of stereo video signal. So it is necessary to study the high performance stereo video coding methods, and more and more researchers have been engaged in this area in recent years.

In 2001, MPEG set up a work group called 3DAV [1][2] to study interactive media application, three dimensional audio, three dimensional video and the standardization of relative technology. At the MEPG meeting held in Klagenfurt, Austria, July of 2006, MPEG decided to allow the JVT (Joint Video Team) to take charge of the work of the standardization of the multiview video coding. Currently, JVT's work mainly concentrates on three aspects based on the H.264 standard: SVC (Scalable Video Coding), MVC (Multiview Video Coding) and SEI (Supplemental Enhancement Information). In the aspect of MVC, JVT has made great progress, and it has proposed a reference model called JMVM (Joint Multiview video Model).

In JMVM, if we only consider the case of two views, the left view adopts the normal H.264[3][4][5] coding method which only has motion estimation[6], that is to say, the reference frames of present frame are those frames which located in the GOP(Group of Pictures)containing the present frame; the right view adopts motion estimation and disparity estimation, which means the reference frames of present frame can be the frames in the GOP containing that frame and the corresponding frame of the left view. So the coding of the right view will be more complicated because of adopting more reference frames. In this paper, for JMVM[7][8], based on the analysis of relationship between MacroBlock types and the quantity of the moving objects contained in the MacroBlock, we propose an improved scheme: it chooses different reference frames for B frame and then realizes the adaptive choice of doing motion estimation or doing joint estimation. It can reduce the coding complexity of right view with only a little performance lost.

II. INTRODUCTION OF JMVM

Brought forward by the group of multiview video coding project of JVT, JMVM is a reference model of multiview video coding based on H.264.

JMVM7.0 is used in this paper, and the number of views is two, since the stereo video coding methods discussed here are based on the binocular view. The GOP size and the IntraPeriod are both 12, and the maximum number of forward or backward reference frames is set to 2 (the reference frame from the left view is thought of as the forward reference frame).

The reference frame model can be descried as follows:

Left view:

I, BBBBBBBBBBB, I, BBBBBBBBBBB, …

Right view:

I, BBBBBBBBBBB, P, BBBBBBBBBBB, …

Here we should know that the real value of QP (Quantization Parameter) can be changed by changing the value of BasisQP (Basic Quantization Parameter) to control the bit rate in JMVM. The relationship between QP and BasisQP can be described in the equation QP = BasisQP + DeltaLayerXQuant, where DeltaLayerXQuant is the quantization offset of the coding frame.

978-0-7695-3557-9/09 $25.00 © 2009 IEEE
DOI 10.1109/ETCS.2009.249

III. TRADITIONAL STEREO VIDEO CODING SCHEMES

As is known from the above, stereo video usually includes two channels, which are the left view and the right view, and the two channels have significant correlation. By the significant correlation considered or not, the stereo video coding scheme can be classified into three types [9], that is both channels use motion estimation independently, or the left view uses motion estimation while the right view uses disparity estimation with the motion estimation neglected, or the left view uses motion estimation while the right view uses motion estimation and disparity estimation.

Scheme I: in this scheme, both channels use motion estimation independently. It does not consider the correlation between the two channels, and it just considers the temporal correlation among the frames in a single channel and the spatial correlation within the present frame;

Scheme II: in this scheme, the left channel uses motion estimation while the right channel uses only disparity estimation. It is the same as scheme I when coding the left channel but it does not consider the temporal correlation when coding the right view.

Scheme III: in this scheme, the left channel uses motion estimation while right channel uses both motion estimation and disparity estimation. It considers the temporal correlation among frames in the present GOP in the right channel, inter-view correlation between the present frame in the right channel and the corresponding frame in the left channel and the spatial correlation within the present frame when coding the right channel. It is the same as scheme I and scheme II when coding the left channel. This scheme is used by JMVM.

Note: In scheme II, there will be no B frames when coding the right channel, since the present frame has only one reference frame which is its corresponding frame in the left channel.

IV. MACROBLOCK TYPE ANALYSIS AND CODING SCHEME IMPROVEMENT

Since B frame has more reference frames and costs more time to finish coding than other types of frames, we will just consider the improvement of coding scheme for B frames while the I and P frames remains to be coded using the schemes defined in JMVM.

A. MacroBlock Type Analysis

MacroBlock type in JMVM is defined as follows:

For I frame, there are four types called Intra16, Intra8, Intra4 and PCM respectively;

For B frame, there are nine types called Direct [10], 16x16, 16x8, 8x16, 8x8, 8x8Frext, Intra16, Intra8, Intra4 respectively;

For P frame, there are nine types called Skip, 16x16, 16x8, 8x16, 8x8, 8x8Frext, Intra16, Intra8, Intra4 respectively。

JMVM chooses scheme III which will get the best coding performance for stereo video coding, but the complexity of the encoder goes much higher. Here we will try to seek an improved scheme to reduce the complexity of the encoder in the case that the coding performance goes down indistinctly.

We will first consider the features of MacroBlock types in B frames in JMVM, since we mainly consider the coding scheme improvement of the B frames of the right view.

Through the experiment, we give percentages of Direct MacroBlocks in B frames when BasisQP=37 in Table I.

TABLE I. PERCENTAGE OF DIRECT MACROBLOCKS IN B FRAMES

Sequence	booksale	soccer	puppy
Proportion（%）	84.5625	79.9796	96.6019

From Table I, we know that most MacroBlocks belong to the type of Direct when BasisQP =37. Besides, puppy sequence has the largest proportion of Direct MacroBlocks, soccer sequence has the smallest proportion of Direct MacroBlocks, and booksale sequence is intermediate.

We know that puppy sequence has the minimum number of moving objects, soccer sequence has the maximum number of moving objects, and booksale sequence is intermediate.

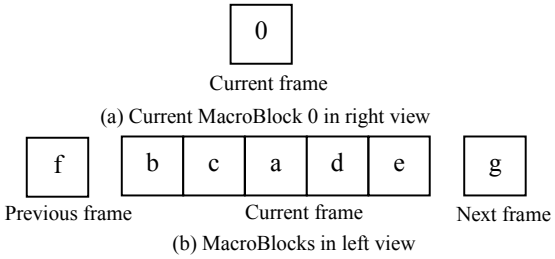
So the proportion of Direct MacroBlocks in B frame can reflect the number of moving objects in the frame, and a B frame containing larger proportion of Direct MacroBlocks contains more moving objects.

B. Our Improved Coding Scheme

On the basis of the analysis in section *A* of this part, we can judge whether the corresponding MacroBlock of present frame in left channel has few or many moving objects according to the information of the types of the corresponding MacroBlock and its surrounding MacroBlocks of present frame and the corresponding MacroBlocks of previous and next frame in left channel.

We know the corresponding MacroBlock in the right channel is similar to that in the left channel, and we can also judge whether the corresponding MacroBlock of present frame in right channel has few or many moving objects. The present MacroBlock will be coded using scheme I if it has few moving objects, or it will be coded using scheme III. The reason why we do not consider scheme II will be given in section B of part Ⅴ.

We modify the coding scheme of a MacroBlock (labeled 0) of a B frame in the right view as follows:



(a) Current MacroBlock 0 in right view

(b) MacroBlocks in left view

Figure 1. MacroBlocks considered

As is shown in Fig. 1 (b), the current MacroBlock in the current frame (left channel) is a, its left two MacroBlocks are b and c, and its right two MacroBlocks are d and e (assuming

that b, c, d, e exist). The corresponding MacroBlock in the previous frame (left channel) is f, and the corresponding MacroBlock in the next frame (left channel) is g.

If the type of MacroBlock a, b, c, d, e, f, g (left channel) are all Direct, current frame (right channel) will be coded by selecting the different-time frame in the same GOP (right channel) as reference frame (only considering motion estimation), otherwise it will be coded by selecting the different-time frame of right channel and corresponding frame of left channel as reference frames (joint estimation, considering both motion and disparity estimation).

If any MacroBlock among a, b, c, d, e does not exist (located in the frame edge), current frame (right channel) will be coded by adopting motion estimation.

Because we just consider the B frame, f and g are always attainable.

Note that if the POC (Picture Order Count) of the present frame is n, then the POC of the previous frame is n-1, and the next frame's is n+1.

The reason why we choose the MacroBlocks along horizontal direction is that the vector of disparity is mainly in the horizontal direction.

V. EXPERIMENTS AND RESULTS ANALYSIS

A. About the Experiments

Stereo video sequences chosen in this paper are shown in Table II. Puppy sequence has the minimum number of moving objects, soccer sequence has the maximum number of moving objects, and booksale is intermediate. The software is Visual Studio2005，the hardware platform are Intel(R) Core(TM)2 Duo CPU E6750 @2.66GHz 2.66G, memory 1.96G。

TABLE II. TESTING SEQUENCES

Sequences	Number of Frames	Sampling Format	Size
puppy	90	4：2：0	720x480
booksale	90	4：2：0	320x240
soccer	90	4：2：0	720x480

B. Experiment Results Analysis

In the experiment, we realize the three stereo video coding schemes and the improved scheme described in section *B* in part Ⅳ.

Results of each scheme when BasisQP=37 are shown in Table III including PSNR (Peak Signal to Noise Ratio) of each component, bit rate and time consumed for coding.

By choosing different values of BasisQP and computing the average of PSNR, we get the Rate-PSNR curve, shown in Fig.2. Here values of BasisQP chosen are 40, 37, 30, and 25. Since the sampling format of the sequences used is 4:2:0, we compute the average of PSNR in the equation

$$PSNR = (yPSNR*4+uPSNR+vPSNR)/6. \quad (1)$$

In equation (1), yPSNR, uPSNR and vPSNR are peak signal to noise ratio of Y component, U component and V component respectively.

Time consumed for coding using each scheme varies with the value of BasisQP, so we can get BasisQP-TIME curve by changing the value of BasisQP. See Fig.3.

First, we consider the existing three schemes. They're quite different in coding performance. On the whole, scheme III provides the best coding performance, scheme II provides the worst coding performance, and scheme I is intermediate.

From Fig.2, we can see that the coding performance of scheme II is quite bad, especially to the puppy sequence because puppy sequence has few moving objects and the inter-view correlation is quite small compared to correlation within a channel, so motion estimation at this time is more effective than disparity estimation. We can also analyze scheme II from Table III in which bit rate of scheme II is more than two times higher than that of scheme I and more than four times higher than that of scheme III when using puppy sequence, and yPSNR of scheme II is 2.308db lower than that of scheme I and 1.9081db lower than that of scheme III.

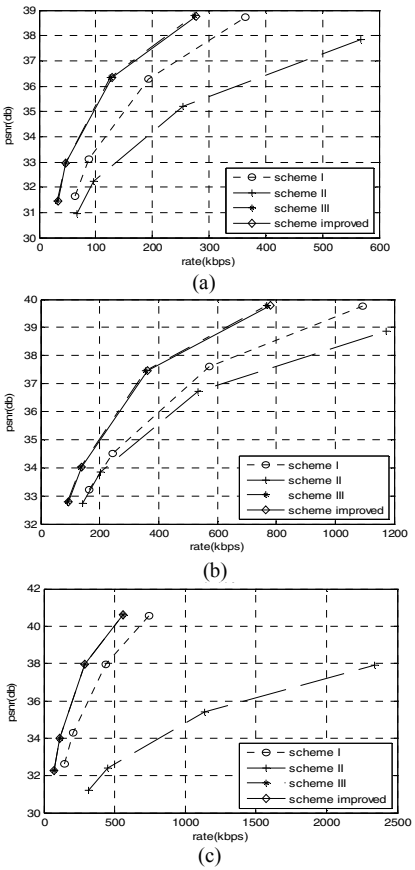


Figure 2. Rate-PSNR curves of different schemes when coding different sequences:
(a) booksale sequence (b) soccer sequence (c) puppy sequence

The coding performance of scheme I is best when coding puppy sequence because the correlation within a channel is larger than inter-view correlation in this case analyzed above. But compared to scheme III, the coding performance of scheme I is still bad. For example, when we use puppy sequence, we can see from Table III that although yPSNR of

scheme I is 0.3999db higher than that of scheme III, the bit rate of scheme I is nearly two times higher than that of scheme III.

Time consumed for coding is also quite different when using different schemes. From Fig.3, we can see that scheme III cost the longest time, scheme II cost the shortest time, and scheme III is intermediate. From Table III, we find that scheme II is more than eight times faster than scheme III, but we don't consider scheme II because its coding performance is too bad as analyzed above. So we just try to find an improved scheme between scheme I and scheme III for encoding every MacroBlock in B frames of the right channel.

Now we discuss the result of improved scheme. From Fig.2 we can clearly see that the coding performance of the improved scheme is quite close to that of scheme III, especially in the case of puppy sequence. From Table III, we can see that yPSNR of the improved scheme is only 0.0015db to 0.0163db less than that of scheme III, and the bit rate of the improved scheme even goes down 0.02% compared to that of scheme III in the case of puppy sequence and rises only 1.79% and 3.15% in the case of booksale sequence and soccer sequence respectively.

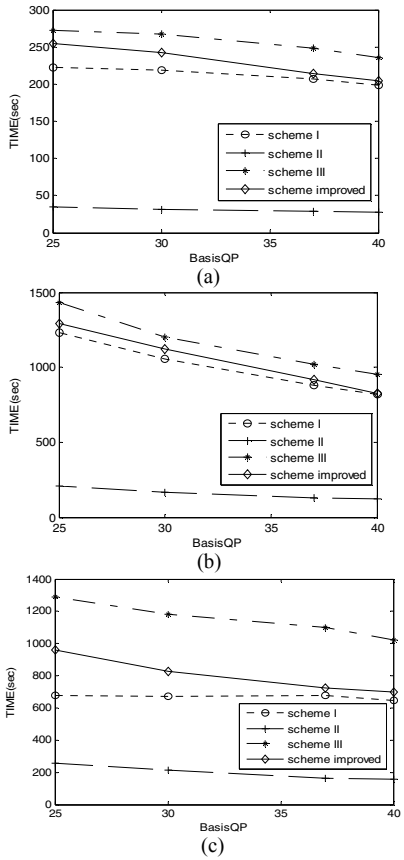


Figure 3. BasisQP-TIME curve of different schemes when coding different sequences:
(a) booksale sequence (b) soccer sequence (c) puppy sequence

In Fig.3, we can see that the time consumed for coding the right view when using the improved scheme is reduced obviously compared to that when using scheme III. This improvement is most obvious in the case of puppy sequence. We can get a quantitative description from Table III that time of the improved scheme goes down 34%, 13.9% and 10.27% in the case of puppy sequence, booksale sequence and soccer sequence respectively.

VI. CONCLUSION

Based on JMVM, three existing stereo video coding schemes are introduced in this paper. By analyzing the characteristic of the MacroBlock types and its relationship with the number of moving objects in MacroBlocks, an improved stereo video coding scheme is proposed. The results of the experiment show that the PSNR of the improved scheme is just about 0.015db less than that of scheme III, and the bit rate of the improved scheme rise less than 3.2% compared with that of scheme III. When we use puppy sequence which has smallest number of moving objects, the bit rate of improved scheme even goes down 0.02%. The coding time of the improved scheme reduces about 10% to 34% compared to that of scheme III.

ACKNOWLEDGMENT

Zhipeng Lin would like to thank the JVT for providing the software of Joint Multiview Video Model for this paper. The stereoscopic sequences for testing are provided by Jianghong Yan from Nanjing University of Posts and Telecommunications in China.

REFERENCES

[1] Aljoscha Smolic, and David McCutchen, "3DAV Exploration of Video-Based Rendering Technology in MPEG", IEEE Transactions on circuits and systems for video technology, March 2004, 14 (3), pp.348-356.

[2] Jose M. Martinez-Ibañez, "MPEG 3DAV AhG Activities Report", 65th MPEG Meeting, Trondheim Norway, July 2003.

[3] ITU-T and ISO/IEC JTC1, "Advanced Video Coding for Generic Audiovisual Services", ITU-T Recommendation H.264 –ISO/IEC 14496-10, 2005.

[4] Thomas Wiegand, Gary J. Sullivan, Gisle Bjontegaard, and Ajay Luthra, "Overview of the H.264/AVC Video Coding Standard", IEEE Transactions on circuits and systems for video technology, 2003, 13(7), pp.560-576.

[5] ISO/IEC JTC1/SC29/WG11 (MPEG), "Coding of audio-visual objects–Part 10: Advanced Video Coding", International Standard 14496-10, ISO/IEC, 2004.

[6] Krit Panusopone, Xue Fang, and Limin Wang, "An Efficient Implementation of Motion Estimation with Prediction for ITU-T H.264 | MPEG-4 AVC", IEEE Transactions on Consumer Electronics, 2007, 53(3), pp.974-978.

[7] Anthony Vetro, Purvin Pandit, Hideaki Kimata, and Aljoscha Smolic, "Joint Multiview Video Model (JMVM) 7.0", Joint Video Team (JVT)26th Meeting: Antalya, TR, 12-18 January, 2008.

[8] ITU-T and ISO/IEC JTC1, "Joint Draft 1.0 on Multiview Video Coding", JVT-U209, Nov 2006.

[9] Yang Hong, "Research on H.264 Based Stereoscopic Video Coding", master's dissertation, college of communication engineering of Jilin University, Changchun, China, 2007.

[10] Alexis Michael Tourapis, Feng Wu, and Shipeng Li, "Direct Mode Coding for Bipredictive Slices in the H.264 Standard", IEEE Transactions on circuits and systems for video technology, Janury 2005, 15(1), pp.119-126.

TABLE III. RESULTS OF EACH SCHEME WHEN BASISQP=37

Sequences	Scheme	YPSNR(db)	UPSNR(db)	VPSNR(db)	RATE(kbps)	TIME(sec)
booksale	Scheme I	30.3609	38.3788	38.9457	88.7178	206.468
	Scheme II	29.2251	38.0179	38.4556	96.3178	29.125
	Scheme III	30.0862	38.3892	38.9294	46.5267	248.464
	Improved scheme	30.0709	38.3908	38.9381	47.3600	213.953
soccer	Scheme I	32.4155	38.6212	38.7840	243.5222	880.906
	Scheme II	31.9295	37.3536	38.0289	203.3622	132.937
	Scheme III	32.0422	37.7313	38.2336	135.7044	1019.5
	Improved scheme	32.0259	37.7479	38.2511	139.9800	914.796
puppy	Scheme I	33.2715	36.8810	35.7525	208.0711	675.953
	Scheme II	30.9635	36.1227	34.3437	448.8133	164.562
	Scheme III	32.8716	36.9694	35.4764	113.6467	1096.72
	Improved scheme	32.8701	36.9678	35.4791	113.6267	723.453