Stereoscopic Video Compression Based on H.264 MVC

Sumei Li Chunping Hou Yicai Ying
School of ElectronicInformation Engineering
Tianjin University
Tianjin, China

Xiaowei Song Lei Yang
School of ElectronicInformation Engineering
Zhongyuan University of technology
Zhengzhou,China

***Abstract*—A new MVC structure is proposed for stereo video compression.Coding performance of the MVC structures is evaluated by DPSNR (Disparity Peak Signal-to-Noise Ratio) of five well-designed stereo video experiments. The experiments' results show that the classification performance, compared with the reference MVC structure, is improved in the new-designed coding structures .It is also proved by the rate-distortion curves that the coding performance of the proposed MVC structure is better than that of the reference MVC structure.**

Keywords- Stereo Video ; DPSNR ; H.264 MVC

I. INTRODUCTION

Video quality has been developed from black-and-white TV to color TV, from analog TV to HDTV. The next generation of video milestone will be multi-view three-dimensional TV [1]. When it comes to talk about stereo video transmission or storage, large volume of the original video data needs to be compressed to save it. Therefore, it's of great value to study high-speed, high-efficiency and high-quality stereo video compression algorithm.

Joint Video Team is researching on the development of the H.264 extension including MVC(Multi-view Video Coding). MVC inherits H.264's excellent features which are high-quality coding efficiency, free coding structure and good network compatibility. Besides, several new coding tools such as temporal scalable structure, viewpoint scalability, and hierarchical B pictures are introduced in the MVC structure [2]. In fact, stereo video is a special type of multi-view video. Stereo video can be previewed in 3D when cameras are at normal eye spacing (2.5 inches). Obviously, MVC is perfectly suitable for stereo video compression.

This paper is organized as follows: The reference MVC structure is explained in section II. A proposed MVC structure is presented in section III. The experiments are described in section IV. Finally, the conclusion is summarized in section V.

II. THE REFERENCE MVC STRUCTURE

The coding scheme of the reference MVC structure [3] is shown in Fig.1. This scheme uses a prediction structure with hierarchical B pictures for each view. Additionally, inter-view prediction is applied to every 2nd view. For an even number of views, the prediction scheme of the last view is a mix of even and odd views. As there is just one neighboring view for inter-view prediction, it starts and ends with P-frames and B pictures have only one inter-view reference. I-frames start each GOP (Group of Pictures) to allow synchronization.

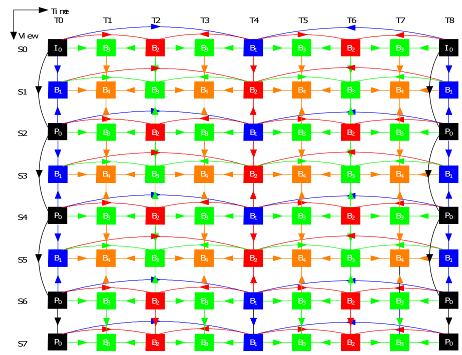


Figure 1. The reference MVC structure

The reference MVC structure is divided into three levels: View S0 is the first level. View S2, S4 and S6 are added in the second level. The remaining views belong to the third level.

III. THE PROPOSED STRUCTURE

H.264 MVC is developed for the compression of multi-view video. Although stereo video is a special kind of multi-view video, some aspects of MVC structure need to be adjusted when used for stereo video compression. If the relationship between adjacent views is a stereo pair, three views (S0, S2 and S1) need to be transmitted at least in order to form a stereo video in the reference MVC structure, which wastes bandwidth of a view.

The proposed MVC structure is shown in Fig.2. Inter-view prediction is applied to most of the views. It is divided into eight levels: S3 is the first level. S4 is added in the second level. S2 is added in the third level. S5 is added in the fourth level. S1 is added in the fifth level. S6 is added in the sixth level. S0 is added in the seventh level. And S7 belongs to the eighth level.

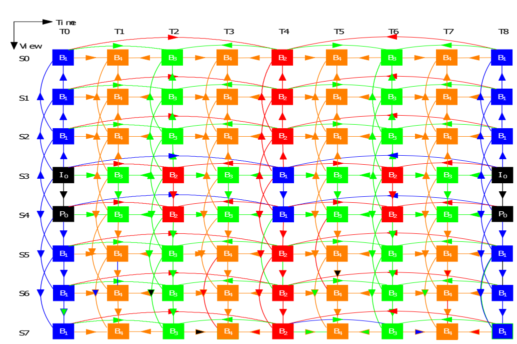


Figure 2. The proposed MVC structure

Figure 3. Eight-views stereo display

In the proposed MVC structure, when view S3 and view S4 are transmitted, stereo video is done. No bandwidth is wasted.

IV. EXPERIMENTS AND RESULTS

A. Experiment Material

A kind of eight-view stereo LCD(Liquid Crystal Display) is shown in Fig.3 . All experimental stereo videos described below are designed by computer for the eight-view stereo LCD. The three videos(Fig.4), named Blueheaven(Fig.4(a)), cassinoavi4d(Fig.4(b)) and Hacker(Fig.4(c)), contain 100 frames of pictures which are sampled in the form of YUV4:2:0. The frame rates of videos are 25 frames per second. There are eight views in a video, and the resolution of each view is 512×288. Any adjacent views make up of a stereopair.

Blueheaven (Fig.4(a)): Three magic cups are shown in the Blueheaven video. The color of the magic cup is changing from pink to blue. Full screen of transparent bubbles are moving at random in the background. cassinoavi4d (Fig.4(b)): A trademark called Cassino is displayed in the center of the picture, rotating around its vertical center on the horizontal surface. There are many Cassino signs dancing like falling leaves in the background. Hacker(Fig.4(a)): Two stereoscopic figures “68” are rotating around their vertical center, where two transparent goblets are rotating inversely. The background looks like the scene of the famous movie “The Matrix”, with words dropping very quickly.

(a)

(b)

(c)

Figure 4. Experiment materials

B. Evaluation Methods

The method of stereo video quality evaluation can be classified into subjective video quality evaluation and objective video quality evaluation. For stereo video, subjective evaluation takes into account six aspects, as follows: Image Quality, Depth, Presence, Naturalness, Visual Comfort, Viewing Experience [4].Considering that about 5 to 10 percent of people belong to stereo blindness [5].before subjective quality evaluation, it is necessary to make sure that the appraisers have normal stereoscopic perception, which will increase the difficulty of subjective evaluation. Of course, but they depend on the user's individual perception and thus, vary from person to person.

Objective video quality evaluation is much easier and cheaper. The most commonly used photo-quality evaluation method is the PSNR (Peak Signal-to-Noise Ratio) [6, 7].

$$PSNR_{dB}=10\log_{10}\frac{(2^n-1)^2}{MSE}$$

MSE is the Mean Squared Error between image before coding and image after decoding. (2n－1)2 is the square of the most likely signal value in the picture. n stands for the number of bits for every pixel.

Y(s, i, j, O/R): the value of Y component's PSNR of pixel whose coordinate is (i,j) in the s-th view. O is short for original view before coding. R is short for reconstructed view after decoding.U(s, i, j, O/R): the value of U component's PSNR of pixel whose coordinate is (i,j) in the s-th view.V(s, i, j, O/R): the value of V component's PSNR of pixel whose coordinate is (i,j) in the s-th view. The resolution of every view is H×V, then,

$$MSE_Y=\frac{1}{v\times h}\sum_{j=1}^{v}\sum_{i=1}^{h}\left(Y(s,i,j,R)-Y(s,i,j,O)\right)^2$$

$$MSE_U=\frac{1}{v\times h}\sum_{j=1}^{v}\sum_{i=1}^{h}\left(U(s,i,j,R)-U(s,i,j,O)\right)^2$$

$$MSE_V=\frac{1}{v\times h}\sum_{j=1}^{v}\sum_{i=1}^{h}\left(V(s,i,j,R)-V(s,i,j,O)\right)^2$$

PSNRY, PSNRU and PSNRV can be used to describe the similarity between video before coding and video after decoding in the same view. The larger value of PSNR, the less difference between video before coding and video after decoding, the more ingredient of the original video is retained. It is proved by lots of experiments that the absolute difference image of stereo pair is similar to the image contour (Looks like Fig.5).

Figure 5. Absolute difference image of cassinoavi4d

DPSNR is the Disparity Peak Signal-to-Noise Ratio between stereopair before coding and stereopair after decoding. It is an objective stereo video quality evaluation method.

$$DPSNR_{dB} = 10\log_{10}\frac{(2^n-1)^2}{DMSE}$$

$$DMSE_Y = \frac{1}{v\times h}\sum_{j=1}^{v}\sum_{i=1}^{h}\left(Y(s+1,i,j,R)-Y(s,i,j,R)-\left(Y(s+1,i,j,O)-Y(s,i,j,O)\right)\right)^2$$

$$DMSE_U = \frac{1}{v\times h}\sum_{j=1}^{v}\sum_{i=1}^{h}\left(U(s+1,i,j,R)-U(s,i,j,R)-\left(U(s+1,i,j,O)-U(s,i,j,O)\right)\right)^2$$

$$DMSE_V = \frac{1}{v\times h}\sum_{j=1}^{v}\sum_{i=1}^{h}\left(V(s+1,i,j,R)-V(s,i,j,R)-\left(V(s+1,i,j,O)-V(s,i,j,O)\right)\right)^2$$

DPSNRY, DPSNRU and DPSNRV can be used to describe the similarity between the difference image before coding and the difference image after decoding in the same stereo pair.

C. *Experiment One:The Relationship between Stereopair*

The purpose of this experiment is to evaluate the degree of difference between the adjacent views, by calculating the PSNR of adjacent views, which is taking one view as the image before coding, and the other view as the image after decoding.

TableI is the result of experiment 1(Blueheaven as the example, the results of the other materials may be gained, too). The following data is average PSNR (or DPSNR) value of 100 frames of pictures in a particular view. In this paper, the unit of PSNR and DPSNR is dB.

These results, together with stereo video characteristics can be summarized to the following conclusions: Firstly, the difference of YUV component between adjacent views has the correlation with particular stereo video. Secondly, the difference of Y component between adjacent views is higher than that of U and V component. Thirdly, the difference of every stereo pair is more or less consistent in a particular stereo video.

D. *Experiment Two:The Relationship of All Views*

The purpose of this experiment is to evaluate the degree of difference between all views, by calculating the PSNR of all views, which is taking the reconstructed view that contains I pictures in the MVC structure as the image before coding, and original views as the image after decoding. To simplify the data analysis, take the results on condition that BasisQP equals to 27 as an example.

TableII ((Blueheaven as the example, the results of the other materials may be gained, too)) is the results of PSNR between reconstructed view that contains I-picture and other reconstructed views.

It is proved by the above experiment results that: Firstly, the average PSNR of proposed MVC structure is higher than that of reference MVC structure. Secondly, the difference of Y component between adjacent views is higher than that of U and V component. Thirdly, the average difference between middle view and other views is smaller than the average difference between the first view and other views. Fourthly, the farther distance of two views, the lower the value of PSNR.

TABLE I. THE DIFFERENCE OF ADJACENT VIEWS IN BLUEHEAVEN

View Pairs		$PSNR_Y$	$PSNR_U$	$PSNR_V$
View_1	View_0	*26.7003*	*37.022*	*40.3963*
View_2	View_1	*26.7928*	*36.887*	*40.3007*
View_3	View_2	*26.7535*	*36.531*	*39.9566*
View_4	View_3	*26.7311*	*36.7495*	*40.2097*
View_5	View_4	*26.6417*	*36.9782*	*40.2576*
View_6	View_5	*26.4663*	*36.3665*	*39.9333*
View_7	View_6	*26.5225*	*36.4364*	*39.7386*
Average PSNR		*26.6583*	*36.7101*	*40.1133*

TABLE II. THE DIFFERENCE OF ALL VIEWS IN BLUEHEAVEN

MVC	View ID	$PSNR_Y$	$PSNR_U$	$PSNR_V$
Reference	View_0	*38.9893*	*41.0695*	*44.2984*
	View_1	*26.7929*	*36.2188*	*39.4741*
	View_2	*22.7452*	*31.9512*	*35.8366*
	View_3	*20.7853*	*29.2627*	*33.5376*
	View_4	*19.5811*	*27.4948*	*31.9822*
	View_5	*18.7498*	*26.3283*	*30.7838*
	View_6	*18.1689*	*25.4171*	*29.8368*
	View_7	*17.7538*	*24.7494*	*29.0149*
	Average	*22.9458*	*30.3115*	*34.3456*
Proposed	View_0	*20.79*	*29.2818*	*33.5416*
	View_1	*22.7648*	*31.8668*	*35.7414*
	View_2	*26.8685*	*35.9122*	*39.2079*
	View_3	*39.0046*	*41.1016*	*44.1959*
	View_4	*26.8267*	*36.0376*	*39.335*
	View_5	*22.6661*	*31.9933*	*35.7911*
	View_6	*20.7405*	*29.2665*	*33.522*
	View_7	*19.582*	*27.4956*	*31.9015*
	Average	*24.9054*	*32.8694*	*36.6546*

E. *Experiment Three:Coding Performance*

Rate-distortion performance can be used to describe the balance relationship between video quality and data rate. In the rate-distortion curve's coordinate, PSNR is the abscissa, and data rate is the ordinate. Rate-distortion performance of codec is widely recognized as a codec performance evaluation method. Good codec needs a better rate-distortion performance and relatively lower computational complexity.

Before comparing performance of MVC structure, testing conditions must be consistent and is presented such as Tab.III.. The same group of BasisQP values and DeltaLayerXQuant values are used for the comparison of coding proformance of reference and proposed MVC structure [8].The parameter of BasisQP indicates the basic QP (Quantization Parameter), which is used to control data rate of the stream.

DeltaLayerXQuant (The value of X equals to an integer ranges from 0 to 5) indicates the difference between the actual

TABLE III. TEST CONDITION OF BASISQP

Experiment materials	BasisQP			
Blueheaven	*37*	*32*	*27*	*22*
cassinoavi4d	*37*	*32*	*27*	*22*
Hacker	*37*	*32*	*27*	*22*

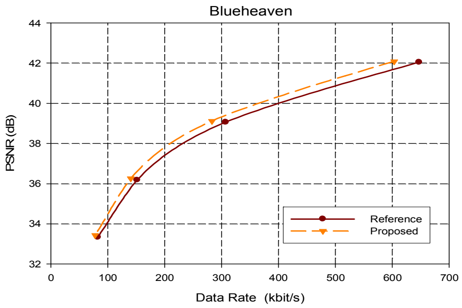


Figure 6. Rate-distortion curves of Blueheaven

quantization parameter and the value of BasisQP. Then, for a particular frame of picture, the actual quantization parameter is determined by QP＝BasisQP +DeltaLayerXQuant.

All of the results are calculated by JMVM_3_0_2 with the same configuration file [9]. It is concluded from Fig.6(Rate-distortion curves of Blueheaven) that the proposed MVC structure has a better coding performance for stereo video.

If a GOP of a particular view contains I picture, the B pictures in the same GOP will have only two inter predictions.

If a GOP of a particular view contains P picture, the B pictures in the same GOP will have an inter-view prediction and two inter predictions.

If a GOP of a particular view contains only B picture, the B pictures will have two inter-view predictions and two inter predictions.

The proposed MVC structure has more inter-view predictions than the reference MVC structure. Therefore, the proposed MVC structure is more complex.

Although the video quality is only slightly improved by increasing the number of inter-view prediction, it is effective to save bit rate by increasing the number of inter-view prediction. Better PSNR and Data Rate performance is gained at the cost of the complexity of MVC structure(such as Fig.7,Fig.8).

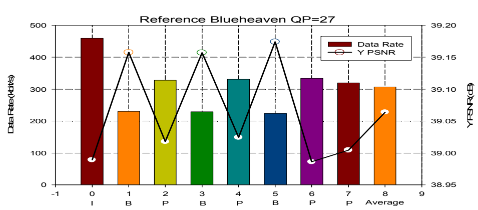


Figure 7. Date Rate and PSNR value of every view in the reference MVC structure

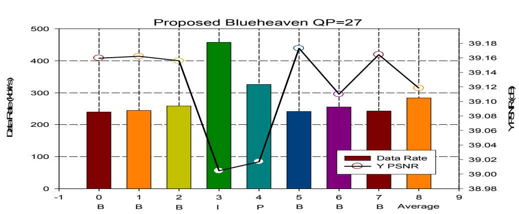


Figure 8. Date Rate and PSNR value of every view in the proposed MVC structure

F. Experiment Four:Classification Performance

This experiment uses the same date as experiment three but analyses from classification performance.

It is proved by Fig.9 that the classification performance of the proposed MVC structure is much better than that of the reference MVC structure.

G. Experiment Five:DPSNR Performance

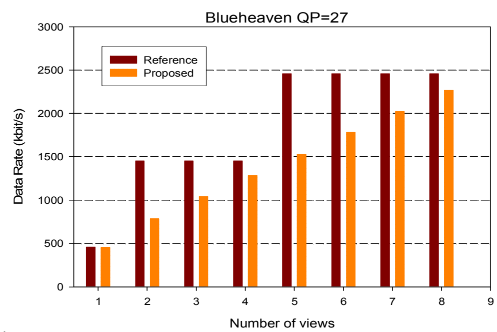
DSPNR is calculated in experiment Five in the reference and proposed MVC structure. Some results list as Tab.IV.

The results of DPSNR suggest that:

Firstly, the proposed MVC structure gains a higher DPSNR value, which means that more ingredient of the original difference image is retained.

Secondly, the higher value of BasisQP, the lower value of DPSNR of all YUV components, the lower stereoscopic impression may be gained from the reconstructed stereo video.

Finally, the DPSNR value of Y component is smaller than that of U and V components.



.

Figure 9. Classification performance of Blueheaven

TABLE IV. AVERAGE DPSNR OF THE REFERENCE AND PROPOSED MVC STRUCTURE IN BLUEHEAVEN

MVC	BasisQP	$DPSNR_Y$	$DPSNR_U$	$DPSNR_V$
Reference	22	*39.6309*	*43.27796*	*46.07646*
	27	*37.06167*	*41.74616*	*44.97637*
	32	*34.80999*	*40.38523*	*44.08911*
	37	*32.72351*	*39.21939*	*43.159*
Proposed	22	*39.66921*	*43.30773*	*46.12473*
	27	*37.1095*	*41.8354*	*45.01553*
	32	*34.88907*	*40.5427*	*44.15799*
	37	*32.82856*	*39.47399*	*43.37411*

V. CONCLUSION

This paper presents an experimental work about Multi-view Video Coding (MVC). It presents an MVC structure for stereo video compression after the analysis of H.264 MVC and the characteristics of stereo video. An objective evaluation of stereoscopic impression is proposed, based on DPSNR (Disparity Peak Signal-to-Noise Ratio), in order to evaluate the coding performance of the proposed MVC structures. The results show that the coding performance of the proposed MVC structure is better than that of the reference MVC structure.

ACKNOWLEDGMENT

The work was supported by the following projects: National Natural Science Foundation of China (NSFC: 60672077）;National Basic Research Program of China (863 Program) （2007AA01Z304、2008AA 01A 320、2009AA01Z306）;Key Technology Support Program of Tianjin （08ZCKFJC27900,09JCZDJC17000）;Tianjin University Project(TJU-YFF-08B63)

REFERENCES

[1] Masayuki Tanimoto, Toshiaki Fujii and Shigeyuki Sakazawa, "Proposal on standardization of Free Viewpoint TV (FTV) system", ISO/IEC JTC1/SC29/ WG11 and ITU-T SG16 Q.6, JVT-T140.doc, July, 2006.

[2] Anthony Vetro, Purvin Pandit, Hideaki Kimata, Joint Multiview Video Model (JMVM) 3.0, ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-V207. doc, January 2007

[3] Richard Hartley, Multiple View Geometry in Computer Vision Second Edition, Cambridge University Press, Second Edition 2003.

[4] Pieter J.H. Seuntiens, Visual Experience of 3D TV, Technische Universiteit Eindhoven, 2006..

[5] W.A. IJsselsteijn, P.J.H. Seuntens, L.M.J. Meesters, State-of-the-art in human factors and quality issues of stereoscopic broadcast television, Deliverable ATTEST/WP5/01, Aug. 2002.pp.2356-2360.

[6] Iain E. G. Richardson, H.264 and MPEG-4 Video Compression Video Coding for Next-generation Multimedia, John Wiley & Sons Ltd.

[7] Stefan Winkler, Digital Video Quality Vision Models and Metrics, Genista Corporation, Montreux, Switzerland, John Wiley & Sons Ltd, 2005

[8] Hideaki Kimata, CE6: View Interpolation Prediction for MVC, ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-V306.doc, January, 2007

[9] JMVM 3.0.2 and JMVM Software Manual, CVS tag: JMVM_3_0_2, Mar. 19, 2007