New Approach to Stereo Video Coding for Auto-stereo Display System

BIN YE[1], GANGYI JIANG [1, 2], FUCUI LI[1], MEI YU [1, 2]

[1] Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China
[2] National Key Laboratory of Machine Perception, Peking University, 100871, China

Abstract: This paper propose a new stereo video coding approach for auto-stereo display system. Firstly, we encode stereo video with H.264 standard compatible. Secondly, we implement disparity estimation to get the disparity information, and then encode them. After decoding these bitstreams, we render virtual views based on disparity information for our auto-stereo display. Exhaustive experiments are conducted on our approach and results are given in detail. Furthermore, we compare the proposed method with other two main stereo video coding methods to demonstrate the advantages. Experimental results show that the proposed method can achieve satisfactory subjective quality of rendered virtual views.

Keywords: Stereo video coding; 3DTV; Auto-stereo display; low computational complexity, View rendering

1 Introduction

3D video have received significant attention in research and development in recent years[1]. It is largely because that 3D video can give viewers a strong sensation of depth just like the real world. 3D video can be widely used in 3DTV[2], Entertainment and Medical Visualization[3] etc..

Auto-stereo displays allowing free 3D viewing with the naked eyes are more comfortable to the viewer, and they are probably candidates for future 3DTV. However, traditional stereo video coding methods[4,5] are no longer suitable to these auto-stereo displays . Because such displayers usually require depth/ disparity information to render virtual images. Such as auto-stereo display, needs at least eight views of a scene.Usually, the receiver is a simple terminal with low processing capacity. Thus, it is difficult for receiver side to implement disparity estimation and render of virtual images in real time.

State-of-the-art system approaches on 3DTV, like the ATTEST proposal, are usually based on the video-plus-depth data corresponding to only a single, central viewing position[6,7]. Video plus depth method allows rendering of virtual views close to the available camera position[8]. For example, a second view corresponding to a stereo pair can be synthesized from video and depth. Although, the problem of rendering is solved by providing with the depth information, quality of virtual views decreases with distance from the available camera, due to effect referred as exposure or disocclusion[9]. Therefore, single video plus depth is not sufficient to deal with the situation when there are lots of occlusion areas. In this paper, we propose a new stereo video coding approach to deal with these problems. Experimental results shows that the proposed method is quite efficient.

2 New stereo video coding approach for auto-stereo display

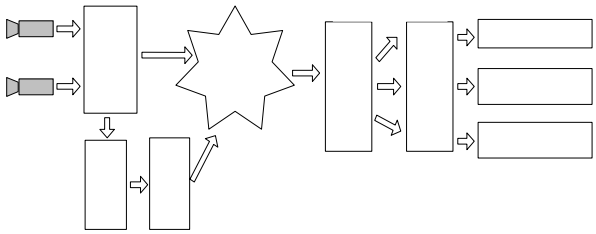

Fig.1.The framework of our 3D system

In our stereo video coding approach, both left and right views as well as disparity information are transmitted to the receiver side. There are two main advantages of the proposed approach. One is that with disparity information, rendering process will be achieved with low complexity. Also, a high adaptability to 3D display properties, viewing conditions and user preferences can be achieved with disparity information. The other is that the problem of disocclusion can be alleviated significantly by providing the receiver side with both left and right views. Additionally, the proposed method has the property of error concealment. The error caused by packet loss of the right view can be concealed by left view and the corresponding disparity information. In this case, our stereo video coding method is more robust. Our 3D display is developed by NewSight Corporation[10] allowing several users to watch the same 3D scene from different perspectives. NewSight use wavelength selective filter array to achieve selective propagation direction for the image information

displayed on the screen. The framework of our 3D system is illustrated in Fig.1, which includes stereo capturing, stereo video coding, transmission, decoding, virtual views rendering and 3D display. Key parts of our 3D system are given in detail in the following paragraphs.

2.1 H.264 compatible Stereo video coding

As the distance of the two viewpoints of stereo video is small which is approximately the distance between two eyes, there is a high correlation between the left and right view. In order to exploit this correlation, in our stereo video coding which is based on H.264, right frames are predicted from both left and right frames to reduce the bandwidth for transmission while left frames are predicted only by previous left frames. Fig.2 shows the stereo video coding structure in which MCP denotes Motion Compensation Prediction and DCP denotes Disparity Compensation Prediction. Further, we use Supplemental Enhancement Information(SEI) to make the stereo video streams can be decoded by a standard H.264 decoder. SEI is designed for some special applications. Decoder can get additional information through SEI messages, such as buffer period, scene information etc.. The SEI payload type for stereo video is 23. According to H.264 standard Annex D, we set syntax left_view_self_contained_flag as 1 and right_view_self_contained_flag as 0 to indicate that left frames only use MCP and right frames use both MCP and DCP. By using the SEI, decoder can decode both left and right video bitstreams or only left video bitstream depending on user's choice. Thus, view scalability is enabled in our stereo video coding.

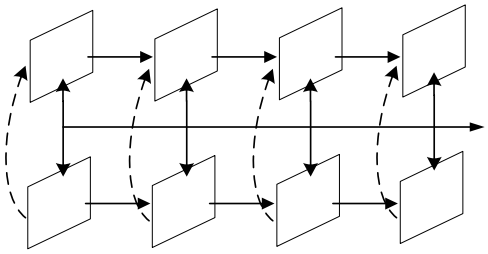

Fig.2. Structure of stereo video coding

2.2 Disparity estimation and views rendering

In order to get the disparity map with high quality, disparity estimation is conducted in pixel accuracy. We adopt block based matching algorithm to each pixel, and the cost function $f(d)$ is defined as follows:

$$f(d) = \arg\min \left| MAD(d) + \frac{\lambda}{4}\sum_{i=1}^{4} |d_i - d| \right| \quad (1)$$

$$MAD(d) = \frac{1}{N_x N_y}\sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-1} |I_L(x+i, y+j) - I_R(x+i+d, y+j)| \quad (2)$$

In the equations above，MAD denotes the mean absolute difference of gray value of matching blocks in left and right images, d denotes the disparity value of current pixel. λ is smoothness parameter. $I_L(x,y)$、$I_R(x,y)$ are gray values of coordinate（x,y）in left and right images respectively, N_x、N_y stand for block size. As shown in Fig.3, d_1、d_2、d_3、d_4 are neighboring disparity values of current pixel's disparity.

d_1	d_2	d_3
d_4	d	

Fig.3. Locality of neighboring disparity values

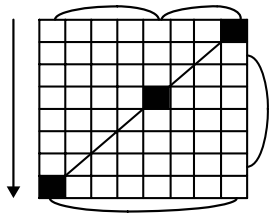

Fig.4. Rendering method from left and right views

Eq. (1) adopt the method of average smooth for disparity values. The added second term penalizes when a pixel has a different disparity value with the neighboring pixels. The effect of smooth is controlled by the parameter λ. Increasing λ results in smoother disparity maps. The principle behind smoothing is that in disparity estimation process, it can be assumed that neighboring pixels should have similar disparity values because disparity values are similar in the same object. We will compare coding performance of disparity maps in terms of smoothness in the next section.

Fig.4 illustrates the principle to render virtual views for a given position α between a stereo image pair. In the ideal case, the relation of the gray values among left, right and interpolated view is given as

$$I_\alpha(x,y) = (1-\alpha)\cdot I_L(x+\alpha\cdot d, y) + \alpha\cdot I_R(x+(\alpha-1)\cdot d, y) \quad (3)$$

where $I(x, y)$ stands for the gray value of a pixel at position (x, y) in the virtual view.

3 Experimental results

We perform experiments on the test sequences of real data called "booksale". Fig.5 show the left and right original images of the test sequences. "booksale" is a standard stereo video test sequence and the image resolution is 320×240. Table 1 gives the H.264 coding parameters for both two color

video and disparity information.

Table 1. The parameters of H.264 coding

Frame Number	30
Search Range	± 32
Entropy Coding	CABAC
Basis QP	20, 24, 28, 32
FME	On

Fig.5. "booksale" left and right original image

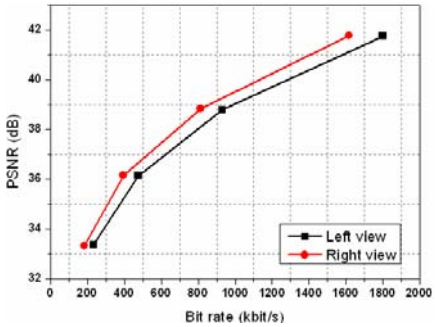

Fig.6. "booksale" coding efficiency

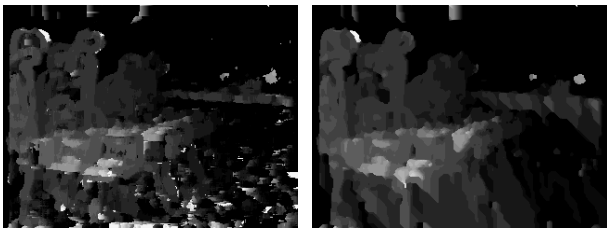

Fig.7 (a) Fig.7 (b)

Fig.7 "booksale" disparity map (a) λ=0; (b) λ=1

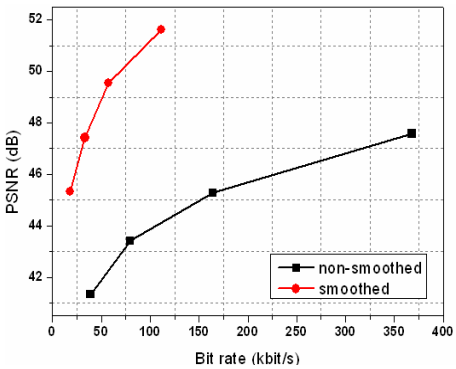

Fig.8. "booksale" disparity coding efficiency

Fig.6 shows the stereo coding result in terms of Rate-Distortion performance. As we can see from Fig.6, the coding efficiency of right view slightly outperforms the left view. This is because the correlation between the left and right views of the test sequence "booksale" is exploited by the DCP. Disparity maps derived by the algorithm presented in previous section are shown in Fig.7. 7(a) shows the disparity map with the smoothness parameter λ=0 and 7(b) shows the disparity map with λ=1. It is obvious that 7(b) is much more smooth. Thus, higher coding efficiency can be expected from the smoothed disparity map. The RD curve demonstrated in Fig.8shows that smoothed disparity maps have a remarkably higher coding performance than the non-smoothed. Compare Fig.6 with Fig.8, we can see that the bit rate of disparity information is very small comparing with the two color video. For instance, while bit rate of left view and right view is 930.1 kbit/s and 812.78 kbit/s respectively, the bit rate of non-smoothed disparity maps is 164.5 kbit/s and the smoothed is 57.26 kbit/s. Therefore, the overhead of our system is extremely small. Fig.9 shows the eight images of "bookslae". The top left is a left view image, the last is a right view image and the others are six rendered virtual view images using the smoothed disparity map.

Fig.9. booksale left、right and six rendered virtual images

If we just use left view and the corresponding disparity information, the virtual image of right view rendered is shown in the Fig.10. As can be seen form the picture, the disoccluded areas in the right view can not be obtained. The quality of these areas decrease significantly to the extent that is unacceptable to the viewer. Thus, the single video-plus-depth method is not sufficient when there are lots of disocclusion phenomenon between the left and right views. In contrast, if both left and right views are available, the problem of disocclusion can be partially solved.

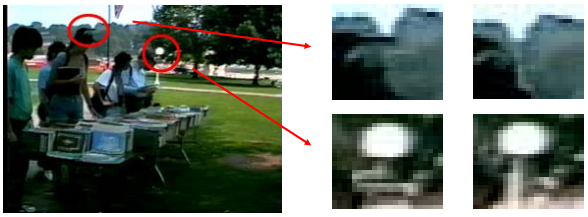

Fig.10. Right virtual image and corresponding original image areas

Table 2. Time consumed by disparity estimation and H.264 decoding process

Test Sequence	Disparity Estimation	Decoding
"booksale"	3725ms	194ms

Stereo video coding method which is only based on the left and right views, should get disparity information on the receiver side to support the auto-stereo display. In our method, the disparity information is also coded and transmitted to the receiver side, so we just need to decode the bitstream to get the disparity information. Table 2 gives the average time per frame consumed by disparity estimation and H.264 decoding process. Experiments are conducted on a 3.00GHz Pentium Ⅳ PC. It is obvious that disparity estimation process is very time-consuming due to its iterant matching process. "booksale" which image resolution is 320×240 needs 3725ms for a stereo image pair in average and the average time for decoding is just 194ms. So it is more reasonable to decode the disparity information rather than to estimate the disparity information in terms of decoding complexity.

4 Conclusion

In this paper, a new stereo video coding approach for auto-stereo display system is proposed. Instead of adopting single video-plus-depth data format, left and right views are both transmitted to receiver side to deal with the problem of exposure. Experimental results show that the final virtual view images interpolated based on the disparity information have satisfactory subjective quality and the bit rate consumed by disparity information is extremely small. Thus, our stereo video coding approach is a very promising method for 3-DTV applications.

Acknowledgement

This work was supported by the Natural Science Foundation of China (grant 60472100, 60672073), NSFC/KOSEF Joint Research Project, the Program for New Century Excellent Talents in University (NCET-06-0537), Natural Science Foundation of Ningbo (grant 2007A610037), the Key Project of Chinese Ministry of Education (grant 206059) , and K.C. Wong Education Foundation in Hong Kong.

References:

[1]. ISO/IEC JTC1/SC29/WG11. Multi-view video plus depth (MVD) format for advanced 3D video systems, W100, San Jose, USA, 21–27 April, 2007

[2]. K. Yun, B. Bae, et al., A DTV-compatible 3DTV broadcasting system, Int. Conf. on Consumer Electronics, Jan. 2006, pp. 149 -150.

[3]. D. Maupu, M.H.Van Horn, S. Weeks, E Bullitt, 3D Stereo Interactive Medical Visualization, IEEE Trans. on Comp.Graph. and Appl., 2005, pp.67-71,

[4]. S. Li, M. Yu, G. Jiang, et al, Approaches to H.264-Based stereo video coding, Int. Conf. on Image and Graphics, 2004, pp. 365-368.

[5]. S. Pehlivan, A. Aksay, C. Bilen, et al., End-to-End Stereo Video Streaming System, Int. Conf. on Multi. and Expo,2006, 2169-2172

[6]. C. Fehn, Depth-image-based rendering (DIBR), compressionand transmission for a new approach on 3D-TV, Proc. of the SPIE, 2004, pp. 93-104.

[7]. P. Kauff, N. Atzpadin, C. Fehn, Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability, SP:IC(22), No. 2, Feb. 2007, pp. 217-234.

[8]. L.Zhang, W. J. Tam, D. Wang, Stereo image generation based on depth images, Int. Conf. on Image Processing, 2004, pp. 2993-2996.

[9]. C. Vázquez, W. J. Tam, et al., Stereo imaging: Filling disoccluded areas in image-based rendering, Proc. of SPIE, Vol. 6392, 2006, pp. 0D1–0D12.

[10]. www.newsight.com