A Novel Virtual View Distortion Estimation for Depth Maps Coding

LI Chun-hua AN Ping ZUO Yi-fan
YANG Chao ZHANG Zhao-yang
College of Communication and Information Engineering
Shanghai University
Shanghai, P.R.China
anping@shu.edu.cn

LI Chun-hua
College of Information Science and Engineering
Hebei University of Science and Technology
Shijiazhuang, P.R.China
lch@hebust.edu.cn

***Abstract*—an efficient virtual view distortion estimation algorithm for depth maps coding is proposed in this paper, in which the virtual view distortion is calculated based on the analysis of occlusion relationship change. It is judged whether the depth coding errors alter the occlusion relationship. If true, wrong occluded pixels or wrong un-occluded pixels are further distinguished. The wrong occluded pixels should be un-occluded in the virtual view, but are occluded for the depth map coding error. While the wrong un-occluded pixels are reverse. All the pixels in depth maps are classified into four types: the base pixels which neither are occluded nor alter the occlusion relation, the occluded pixels, the wrong occluded pixels and the wrong un- occluded pixels. The virtual distortion is calculated according to the characteristics of different depth map pixel types. Experimental results show that the proposed virtual distortion estimation algorithm is effective. When the traditional Rate-Distortion Optimization (RDO) in depth map coding is replaced by the estimating algorithm proposed in this paper, the average coding bitrates reduce about 20%. It demonstrates the better subject quality also.**

Keywords-Three dimensional video (3DV); depth coding; virtual view distortion; Rate-Distortion Optimization (RDO); Multiview plus depth (MVD)

I. INTRODUCTION

With the development of three-dimension video (3DV) application, such as three dimension television (3DTV) [1] and free viewpoint television (FTV) [2], the confliction between the huge data volume and limited transmission bandwidth or storage resource becomes obvious. One of the promising solutions is using multi-view plus depth (MVD) [3] format to represent 3DV, by which only a few views of texture and depth to be stored or transmitted. At the decoder, arbitrate view can be synthesized by depth map based rendering (DIBR) [4].

Depth maps record depth information of its corresponding pixel in texture video, which can be viewed as grey-scale images sequence. The common approach to compress it is applying the popular video coding standard, e.g. H.264/AVC [5]. In order to improve the coding efficient of depth maps, many algorithm take advantage of the statistical features of the depth maps [6-12]. Although all of them moderately advance the depth coding performance, they still apply the traditional Rate-Distortion Optimization (RDO) to make mode decision which balances the depth maps coding error and rate. But the use of depth maps is different from the texture images, which is not displayed directly but help virtual view synthesis. As a result, the depth map coding errors degrade the synthesized virtual view quality by providing wrong geometric information. So, it is necessary to amend the traditional RDO criterion in depth maps coding for the synthesized virtual view quality.

To design a more reasonable RDO criterion, the relationship between the depth map distortion and virtual view distortion must be considered. The experiments revealed the phenomenon that depth map coding error causes virtual view distortion varying with the virtual viewpoint location [13]. Kim et al. built new RDO criterions using lineal model and autoregressive model estimating virtual distortion to improve the depth map coding [14, 15]. Motion sensitivity model is presented to analyze the relationship between virtual view distortion and depth map coding error [16-17]. Oh et al. pixel-wisely calculated the virtual view distortion caused by depth coding error [18-19]. Even though these coding error-virtual view distortion models of depth maps ameliorate the RDO criterion to some degree, they all assumes that the occlusion relationship formed during DIBR is not been altered by depth map coding error. But depth map coding errors alter the occlusion relationship in fact, especially in edge areas, which seriously depraved the virtual view quality when QP increases.

Aim at this problem, an efficient virtual view distortion estimation algorithm for depth maps coding is proposed in this paper, in which the virtual view distortion is pixel-wisely calculated based on the analysis of occlusion relationship change. The experiment shows it is efficient to put up the depth map coding performance.

The rest of this paper is organized as follows. The effect of depth map coding error on virtual view distortion is analyzed in section 2, the proposed virtual distortion estimation algorithm is presented in section 3, experimental result is provided in section 4 and section 5 gives a conclusion of the paper.

II. EFFECT OF DEPTH MAP CODING ERROR ON VIRTUAL VIEW DISTORTION

Multi-view plus depth (MVD) is widely adopted in three dimensional video (3DV) systems, in which depth maps supply geometry information to synthesize the virtual view. The distortion of the compressed depth maps is propagated into the virtual view during 3D wrapping process. In order to accurately estimate the synthesized view distortion caused by the coding error of the depth maps, it is essential to analyze the relationship between them in detail.

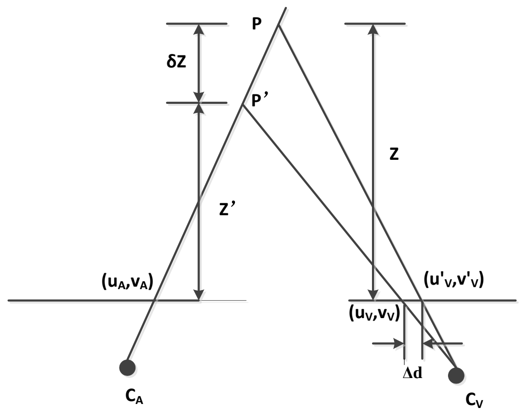

Figure 1. Virtual view distortion caused by depth distortion

As shown in Figure 1, the depth maps distortion propagated into the virtual view through geometric change in the scene. An image coordinate (u_A, v_A) in C_A view should be mapped into $P(x, y, z)$ in the world coordinate with noise-free depth map $D(u_A, v_A)$, while it is mapped into $P'(x, y, z)$ with depth map coding error $\Delta D(u_A, v_A)$. As a result, the rendered pixel position is shifted from (u_V, v_V) to (u'_A, v'_A). For a parallel camera arrangement, the disparity error Δd can be written as [20]:

$$\Delta d = B \cdot f \cdot \Delta D(u_A, v_A)(1/Z_{near} - 1/Z_{far})/255 \qquad (1)$$

In (3), B is baseline, f is camera focal length of horizontal direction, Z_{near} and Z_{far} represent the nearest and farthest depth value, which correspond to 255 and 0 in depth map respectively.

The depth map coding error wrongly shifts the rendered pixel position in the virtual view, which may further alter the occlusion relationship. Then some occluded pixels become un-occluded, while some pixels are in the reverse case. Suppose curve $A(u', v)$ links all the un-occluded pixels whose positions are wrapped by original depth map, and curve $B(u', v)$ links those wrapped by the decoded depth map, the virtual view distortion caused by depth coding error is:

$$SSD_{virtual} = \int \left| A(u', v) - B(u', v) \right|^2 du' \qquad (2)$$

III. VIRTUAL VIEW DISTORTION ESTIMATION

The proposed virtual distortion estimation algorithm calculated the virtual view distortion by pixel-wise analysis of occlusion relationship change. On the basis, depth maps pixels are classified into four types: the base pixels, the occluded pixels, the wrong occluded pixels and the wrong un-occluded pixels.

The virtual view distortion aroused by the coding error of the base pixels is called basic virtual view distortion in this paper. The occluded pixels are neglected because they can't produce distortion in virtual view. The wrong occluded pixels and the wrong un-occluded pixels lead to the virtual view distortion change for they alter the occlusion relationship. The adjustment is called subsidiary virtual view distortion, which can be divided into wrong occluded distortion and wrong un-occluded distortion. Thus, Equation (3) is approximated by the sum of the basic virtual view distortion and subsidiary virtual view distortion:

$$SSD_{virtual} \approx SSD_{basic} + SSD_{\mathrm{subsidiary}} \qquad (3)$$

A. *Occlusion handling*

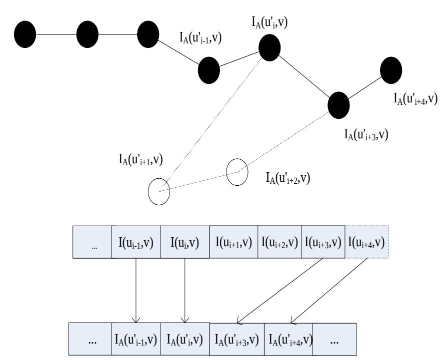

Figure 2. Occlusion phenomenon in DIBR

As shown in fig. 2, when right referrence view renders the intermidiate view during DIBR, some pixels may be occluded by other pixel. It is similiar with left referrence view rendering. We render them pixel-wisely from left to right. If current pixel's rendering location lies on the left of the rightmost rendering location of the pixels before, it is occluded.

The base pixels are those neither are occluded nor alter the occlusion relation. The occluded pixels are normal occluded ones, which are occluding both under the control of noise-free and noise depth map. As normal occluded pixels don't appear in virtual view, they are ignored while calculating virtual view distortion. If the pixel is not occluded when rendering by noise-free depth map while occluded by noise depth map, they are called wrong occluded pixels. The wrong un-occluded pixels refer those pixels which alter the occlusion relation in reverse case.

B. *Basic virtual view distortion calculation*

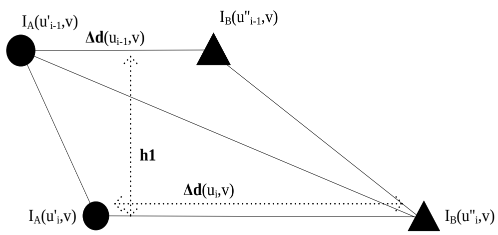

Figure 3. Basic virtual view distortion calculation

As shown in Fig.3, $I(u_{i-1},v), I(u_i,v)$ are the pixels of reference view corresponding to the adjacent basic depth pixels. They are wrapped into $I_A(u'_{i-1},v), I_A(u'_i,v)$ and $I_B(u''_{i-1},v), I_B(u''_i,v)$ in virtual view by noise-free depth map pixels $D(u_{i-1},v), D(u_i,v)$ and noise depth map pixels $D'(u_{i-1},v), D'(u_i,v)$. $I_A(u'_{i-1},v), I_A(u'_i,v)$ and $I_B(u''_{i-1},v)$, $I_B(u''_i,v)$ form a trapezoid. Its upper line and the lower line are the disparity error $\Delta d(u_{i-1},v)$ and $\Delta d(u_i,v)$, which can be calculated by (1). And its height $h1$ is the intensity difference between $I(u_{i-1},v), I(u_i,v)$. So, the basic virtual view distortion is the area of the parallelogram, which can be written as:

$$SSD_{basic}(u_i,v) = [\Delta d(u_{i-1},v) + \Delta d(u_i,v)]\cdot h1/2 = [\Delta d(u_{i-1},v) + \Delta d(u_i,v)]\cdot[|I(u_{i-1},v) - I(u_i,v)|]/2 \quad (4)$$

C. *subsidiary virtual view distortion*

As mentioned above, some depth map distortion results pixel is error occluded or error un-occluded in virtual view. Fig.4 and fig.5 show the virtual view distortion calculation of the two cases separately.

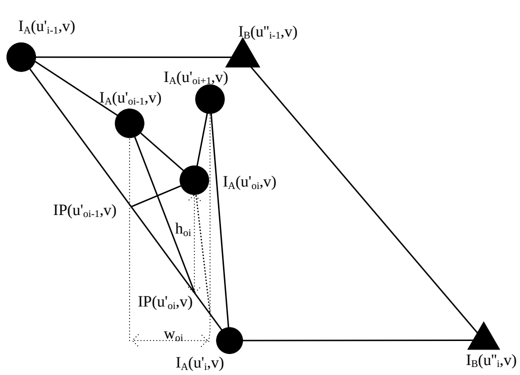

Figure 4. Error occlusion virtual view distortion calculation

In fig.4, under the control of the adjacent error occluded depth map pixel $D'(u_{oi-1},v), D'(u_{oi},v), D'(u_{oi+1},v)$, $I(u_{oi-1},v)$, $I(u_{oi},v)$ and $I(u_{oi+1},v)$ are wrapped to $I_A(u'_{oi-1},v)$, $I_A(u'_{oi},v)$ and $I_A(u'_{oi+1},v)$ respectively. When calculating the basic virtual view distortion, $IP_{oi}(u'_{oi},v)$ is interpolated by pixel $I_A(u'_i,v)$ and $I_A(u'_L,v)$. But, it is $I_A(u'_{oi},v)$ when wrapped by noise-free depth in fact. So, the basic virtual distortion needs to be amended. The adjustment is the area of the triangle composed of $IP_{oi}(u'_{oi-1},v)$, $IP_{oi}(u'_{oi+1},v)$ and $I_A(u'_{oi},v)$. Let w_{oi} and h_{oi} are the base and height of the triangle respectively, it can be written as:

$$SSD_{error\_occlusion}(u_{oi},v) = \frac{1}{2}w_{oi}h_{oi} = \frac{1}{2}|u'_{oi-1} - u'_{oi+1}|[IP_{oi} - I_A(u'_{oi},v)] = \frac{1}{2}|u'_{oi-1} - u'_{oi+1}|[IP_{oi} - I(u_{oi},v)] \quad (5)$$

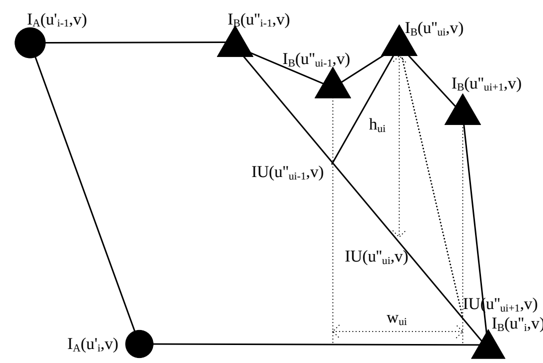

Figure 5. Error un-occlusion virtual view distortion calculation

Similarity, coding errors of error un-occlusion depth map pixels make it necessary to modify the basic virtual distortion calculation also. In Figure 5, $I(u_{ui-1},v)$, $I(u_{ui},v)$ and $I(u_{ui+1},v)$ are wrapped to $I_B(u''_{ui-1},v)$, $I_B(u''_{ui},v)$ and $I_B(u''_{ui+1},v)$ by the adjacent error occluded depth map pixel $D'(u_{ui-1},v), D'(u_{ui},v), D'(u_{ui+1},v)$. When calculating the basic virtual view distortion, $IU(u''_{oi},v)$ which is interpolated by pixel $I_B(u''_i,v)$ and $I_B(u''_L,v)$, replaces the actual $I_B(u''_{ui},v)$ and adds the estimated distortion error. Thus, the basic virtual distortion can be adjusted by calculating the area of the triangle formed by $IP_{ui}(u''_{ui-1},v)$, $IP_{ui}(u''_{ui+1},v)$ and $I_B(u''_{ui},v)$ as (6).

$$SSD_{error\_un-occlusion}(u_{ui},v) = \frac{1}{2}w_{ui}h_{ui} = \frac{1}{2}|u''_{ui-1} - u''_{ui+1}|[I_B(u''_{ui},v) - IU(u''_{ui},v)] = \frac{1}{2}|u''_{ui-1} - u''_{ui+1}|[I(u_{ui},v) - IU(u''_{ui},v)] \quad (6)$$

In summary, the subsidiary virtual view distortion is:

$$SSD_{\text{subsidiary}} = \sum_{v}\Big(\sum_{oi} SSD_{erroe\_occlusion}(u_{oi}, v) + \sum_{ui} SSD_{erroe\_un\text{-}occlusion}(u_{ui}, v)\Big) \quad (7)$$

IV. Experimental Results

In order to integrate the proposed virtual estimation into H.264 reference platform, we extended it by adding original reference color view to aid virtual view distortion calculation. And the virtual view is synthesized by VSRS 3.5. The test sequences "Newspaper", "Kendo", "Book_arrival" and "Balloon" from MPEG group are used to validate the algorithm effective. The synthesized virtual view video by original reference color video and depth maps is used as ground-truth virtual view because not all the videos can be captured at arbitrary virtual view.

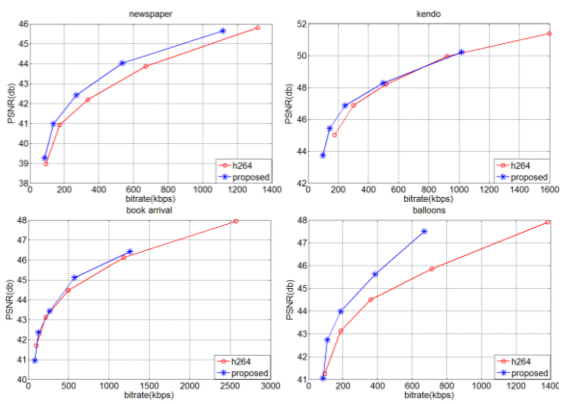

Figure 6. Rate-distortion performance comparison

Fig. 6 provides the rate-distortion performance comparison of the proposed RDO criterion and traditional RDO criterion. It is obviously, the proposed RDO criterion is better than traditional RDO criterion. Table 1 gives the BD-rate and BD-PSNR results of the proposed depth coding vs. H.264, bitrates is reduced about 20% and PSNR is improved about 0.6 on average.

TABLE I. PROPOSED DEPTH CODING V.S H 264

sequence	BD-rate(%)	BD-psnr(dB)
newspaper	-23.7%	0.638
kendo	-19.6%	0.128
book_arrival	-29.86%	0.66
balloon	-30.7%	1.23
average	-25.1%	0.66

The synthesized virtual view distortion aroused by depth map coding error is related to the local feature of responding color image. The more complex the texture of the color image, the more heaven the distortion of the virtual view. Once the wrong occlusion relation occurs, i.e. the foreground pixel wrongly substitutes the background pixel or inverse, exactly estimate the virtual view distortion become more difficult to deal with because the difference value of the foreground pixel and the background pixel is usually large and a little number such depth map pixels would introduce heaven virtual view distortion. And they often locate at the edge areas resulting subject quality of 3DV is affected seriously. The texture of "Balloon" is more complex than "kendo", and its depth map is of more edge pixels than "kendo". Therefore, "Balloon" improves more coding performance than "kendo".

Fig. 7 shows the virtual view images. From them, it is clear that the proposed RDO criterion improves the subjective quality of virtual images.

a) ground truth virtual view image

b) the virtual view image synthesized by H.264

c) the virtual view image synthesized by the proposed algotithm

Figure 7. Subject quality comparison

V. Conclusion

In this paper, a novel virtual view distortion estimation algorithm is proposed based on the analysis of the relationship between the coding error of depth maps and the synthesized virtual view distortion. It calculates the virtual view distortion distinguishing the occlusion relation change. Experiments validate its effectiveness from objective RD performance and subject quality of the syntheses synthesized virtual view image.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China, under Grants U1301257, 61172096 and 60902085, the key Project of Shanghai Science and Technology Commission, under Grant 12DZ2293500.

REFERENCES

[1] L. Onural, A. Gotchev, H. M. Ozaktas, and E. Stoykova, "A Survey of signal processing problems and tools in holographic 3D television," IEEE Trans. Circuits Syst. Video Technol., vol. 17, no. 11, pp. 1631-1646, Nov. 2007.

[2] M. Tanimoto and T. Fujii, "Free viewpoint television", in Proc. 61th Meet.ISO/IEC JTCI/SC29/WG11, Jul.2002, no. M8595.

[3] K. Muller, P. Merkle, and T. Wiegand, "3-D video representation using depth maps," Proc. IEEE, vol. 99, no. 4, pp.643-656, April 2011.

[4] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Multiview video plus depth representation and coding," Proc. IEEE Int. Conf. Image Process., 2007, pp. 201–204.

[5] Telecommunication Standardization Sector of ITU, Series H: Audio Visual and Multimedia Systems Infrastructure of Audiovisual Services:Coding of Moving Video, Annex H, Recomm. ITU-T H.264, Mar. 2011.

[6] Da-Hyun Yoon and Yo-Sung Ho,"Fast Mode Decision Algorithm for Depth Coding in 3D Video Systems Using H.264/AVC", PSIVT 2011, Part II, LNCS 7088, pp. 25–35, 2011.

[7] J. Y. Lee, H. Wey, and D. S. Park, "A fast and efficient multi-view depth image coding method based on temporal and inter-view correlations of texture images," IEEE Trans. Circuits Syst. Video Technol., vol. 21, no.12, pp. 1859-1868, Dec. 2012.

[8] Buncha Kamolrat, Anil. Fernando, Marta Mrak, Ahmet Kondoz, "3D Motion Estimation for Depth Image Coding in 3D Video Coding," IEEE Trans. on Consumer Electronics, Volume 55, Issue 2, pp. 824-830,May 2009.

[9] Y. Morvan, P.H.N. de With, and D. Farin, "Platelet-based coding of depth maps for the transmission of multiview images," in Proceedings of SPIE, Stereoscopic Displays and Applications, 2006, vol. 6055.

[10] MAITRE M, DO M N. "Depth and depth-color coding using shape-adaptive wavelets". Journal of Visual Communication and Image Representation, 2010,21(5-6): 513~522.

[11] Wildeboer M.O., Yendo T., Tehrani M.P., etal., "Depth up-sampling for depth coding using view information". Proceedings of 3DTV-Conference 2011: The True Vision Capture, Transmission and Display of 3D Video(3DTV2011), Antalya, Turkey, May, 2011.

[12] S. Milani, G. Calvagno, "A depth image coder based on progressive silhouettes," IEEE Signal Processing Letters, vol. 17, no. 8, pp. 711-714, Aug. 2010.

[13] Merkle, P., Morvan, Y., Smolic, A., Farin, D., Mueller, K., de With, P. H. N., and Wiegand, T., "The effects of multiview depth video compression on multiview rendering," Singal Proc.: Image Comm. 24(1-2), 73-88 , 2009.

[14] W. S. Kim, A. Ortega, P. L. Lai, D. Tian, and C. Gomila, "Depth map distortion analysis for view rendering and depth coding", Proc. IEEE Int. Conf. Image Process., 2009, pp. 721–724.

[15] W. S. Kim, A. Ortega, P. L. Lai, D. Tian, and C. Gomila, "Depth map coding with distortion estimation of rendered view", Proc. SPIE Vis. Inf. Process. Commun., 2010.

[16] Yanwei Liu, Qingming Huang, Siwei Ma, Debin Zhao, Wen Gao. "Joint Video/Depth Rate Allocation for 3D Video Coding based on View Synthesis Distortion Model".Signal Processing: Image Communication, Vol.24, No.8, pp. 666-681, 2009.

[17] Qifei Wang; Xiangyang Ji; Qionghai Dai; Naiyao Zhang , "Free Viewpoint Video Coding With Rate-Distortion Analysis", IEEE Transactions on Circuits and Systems for Video Technology, Vol.22,No.6,pp.875-889, 2012.

[18] B. T. Oh, J. Lee, and D.-S. Park, "Depth map coding based on synthesized view distortion function," IEEE Journal of Selected Topics in Signal Processing, vol. 5, no. 7, pp. 1344 –1352, Nov. 2011.

[19] Li Wang, Lu Yu. "Rate-distortion optimization for depth map coding with distortion estimation of synthesized view", 2013 IEEE International Symposium on Circuits and Systems (ISCAS),May,2013.pp17-20.

[20] Lai, P., Ortega, A., Dorea, C., Yin, P., and Gomila, C., "Improving view rendering quality and coding efficiency by suppressing compression artifacts in depth-image coding", in [Visual Communic. and ImageProc., VCIP 2009], Proc. SPIE (Jan. 2009).