GEOMETRIC MAPPING ASSISTED MULTI-VIEW DEPTH VIDEO CODING

Qiong Liu[1,2] *Yongbing Zhang*[3] *Xiangyang Ji*[1] *Qionghai Dai*[1]

[1]Department of Automation, Tsinghua University, Beijing 100084 China
[2]Department of Electronics and information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China
[3]Graduate School at Shenzhen, Tsinghua University, Shenzhen, China,

ABSTRACT

Multi-view plus depth (MVD), as a video representation supporting view synthesis based on depth video, has attracted more and more attention for the free view video (FVV) application. It is a challenge to efficiently compress the multi-view depth data in MVD format. In this paper, we explore the geometric relationships in 3D space and propose a geometric mapping assisted (GMA) multi-view depth video coding algorithm. The proposed GMA utilizes the mapped depth image as a reference candidate during prediction. Furthermore, the inpainting method is employed to fill in the holes in mapped depth images. Experimental results demonstrate the gains of up to 2.45 dB for the depth coding, as well as better quality of synthesized views.

Index Terms— video coding, inter-view prediction, depth, geometric mapping,

1. INTRODUCTION

Free View Video (FVV), which is represented by *Multi-view plus Depth* (MVD), is an attractive future video application that is characterized by enabling the user to interactively select the desired viewpoint [1]. The representation of MVD is a data format consisting of multi-view texture video and associated depth video. The depth video provides 3D space information of each pixel for image synthesis when virtual view is selected by user. The MVD data causes a vital problem for data storage and transmission. Efficient compression techniques both for multi-view texture video and depth video are urgently in need.

The standardization of the technologies in this field is the standard of *Multi-view Video Coding* (MVC) which was developed by the Joint Video Team (JVT) of ITU and MPEG and was finalized in 2008 [2]. It makes a first step towards for the FVV standardization. As a second step, MPEG started to work on 3D video recently [3], focusing on FVV from a normative point of view, including its representation, generation, processing, coding and rendering of MVD data. The higher bit-rate of depth is desired for maintaining the rendering quality especially around object edges.

Similar to texture video coding, there is a trade-off between bit-rate of depth coding and quality of synthesized view. There are several techniques developed in the literature to reduce the bit-rate by considering the characteristics of the depth. For example, the context of the depth video is usually considered much sparser than the texture video. Based on this character, a reduced resolution version of depth can be encoded by conventional compression techniques [4]. This kind of approaches can reduce the bit-rate substantially, but the up-sampling method should be carefully designed to avoid quality loss. Another kind of approach focused on encoding the edges of the depth particularly, such as a multi-layered depth coding [5], edge-based filter for depth coding [6] and platelets-based encoding method [7].

All of the work described above does not consider about the geometric relationship of multi-view depth. Multi-view depth data represents the geometry location of 3D-point in different view. The relationship of different views of depth follows the rule of geometric constrain of 3D space, which means that the depth can be mapped from one view to another. The mapped depth properly has higher correlation of a certain view. Inspired by this idea, we propose a geometric mapping assisted (GMA) multi-view depth video coding algorithm aiming to improve the coding efficiency.

This paper is organized as follows. Section 2 represents the details of the GMA, including the prediction structure of GMA and the refinement of the mapped depth. In the Section 3, the efficiency of GMA is evaluated through the depth coding rate, depth quality and rendering quality. A conclusion is given in section 4.

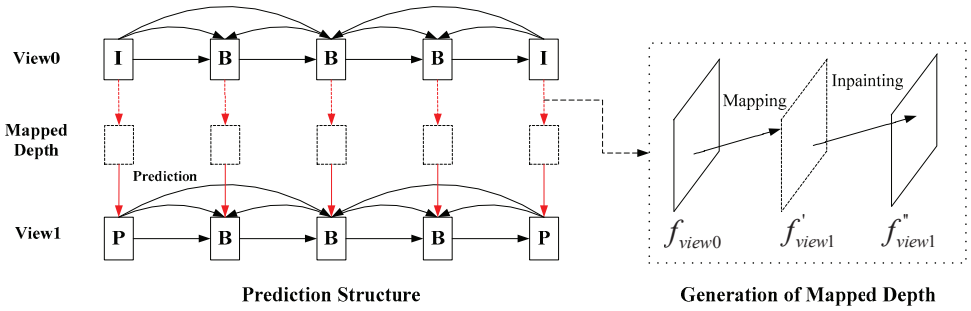


Figure 1 The Prediction Structure of Geometric Mapping Assisted Multi-view Depth Coding

2. GEOMETIC MAPPING ASSISTED MULTI-VIEW DEPTH VIDEO CODING ALGORITHM

The MVC, which is based on H.264 coding technologies, is extended to multi-view from single view by adding interview prediction. The inter-view prediction technique investigates the correlations between different views. Generally, the depth video is compressed as gray-scaled video signals. However, the depth video represents the distance information in 3D space, which is quite different from the natural video signals. It can be mapped from one view to another according to the geometric relationship. Inspired by this, we propose the GMA multi-view prediction algorithm in this section.

2.1 The prediction structure of GMA

Figure 1 illustrates the prediction structure of geometric mapping assisted multi-video depth coding. In order to clarify our idea, we give an example on the case of two views in the following description. For instance, view 0 is considered as the basis view and view 1 can be predicted from view 0. The prediction structure including both temporal and inter-view prediction is provided in Figure 1. For multi-view video, the coding efficiency of inter-view prediction has significant effects on the coding performance. However, the depth images are much smoother than texture images, which easily result in mismatches of motion estimation. Especially, there are lower correlations between views with wide baseline or large view-angle. Unfortunately, the limited number of cameras is quite desired in order to reduce the vast amount of data for the application of FVV. Therefore, we propose to utilize the geometric mapping to eliminate the redundancy of multi-view depths, which is robust in the case of wide baseline or large view-angle.

In our proposed algorithm, the reference frame f_{view0} from view 0 is mapped into f'_{view1} at first. The mapped frame f'_{view1} is very similar as the corresponding frame of view 1, but it suffers from the "hole" caused by occlusions. In order to get better performance, f'_{view1} is further refined into f''_{view1} which is considered as the reference frame of the corresponding encoding frame in view 1.

The geometric mapping is the key point in the proposed method. For a 3D-point X, the corresponding points of view 0 and view 1 are denoted as X_0 and X_1. Suppose the coordinates of X_0 and X_1 are $[X_0, Y_0, Z_0]^T$ and $[X_1, Y_1, Z_1]^T$ respectively. The depth values of X in view 0 and view 1 are $z_0(x_0, y_0)$ and $z_1(x_1, y_1)$. The depth can be mapped from one view to another by the transformation of the coordinates. For two arbitrarily positioned cameras, let $[\mathbf{R}]_0$ and $\mathbf{t}_0$ ($[\mathbf{R}]_1$ and $\mathbf{t}_1$) denote the rotation matrix and the translation vector required to align the camera of view 0 (view 1) coordinates. Then the coordinates of the left and right views, $\mathbf{X}_0 = [X_0, Y_0, Z_0]^T$ and $\mathbf{X}_1 = [X_1, Y_1, Z_1]^T$ are related by

$$\boldsymbol{X}_1 = [\boldsymbol{R}]_{10}\boldsymbol{X}_0 + \boldsymbol{T}_{10} \tag{1}$$

with

$$[\boldsymbol{R}]_{10} = [\boldsymbol{R}]_1[\boldsymbol{R}]_0^T;\ \boldsymbol{T}_{10} = \boldsymbol{T}_1 - [\boldsymbol{R}]_1[\boldsymbol{R}]_0^T\boldsymbol{T}_0 \tag{2}$$

The rotation matrix and translation vector are related to the angle θ and camera parameters, which is described in detail by [8]. The case of parallel camera setting is derived by setting θ as 0. In this way, the depth map can be mapped from one view to another by Equation (1) and (2), especially the case of the parallel cameras is solved by Equation (3).

$$\begin{cases} x_1 = x_0 - \dfrac{fb}{z_0} \\ y_1 = y_0 \end{cases} \tag{3}$$

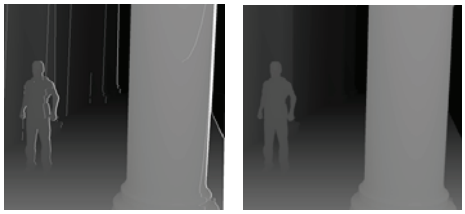
where f is the focal length and b is the baseline length.

2.2 Refinement of mapped depth images

It should be noticed that it is impossible to map all pixels of one view to another because of inter-view occlusions, especially in the case of wide baseline of cameras. As illustrated in Figure 2(a), the white area is the "hole" caused by occlusions, where depth cannot be obtained from mapping. The prediction accuracy will decrease due to these artifacts. To solve this problem, the inpainting algorithm [9] is adopted in our algorithm. An example of depth map

processed by inpainting is provided by Figure 2 (b). From the result, we can see that the all the holes are filled well.

Furthermore, the residual image between the mapped and original depth is compared. It is can be observed from Figure 3 that the depth mapping reduces the energy of residual greatly.



(a) (b)

Figure 2 Example of mapped depth of sequence "*Dancer*" (a) The mapped depth image without inpainting (b) The mapped depth image with inpainting.

(a) (b)

Figure 3 Comparison of the residual of sequence "*Dancer*" (a) with mapped depth (b) with unmapped depth

3. EXPERIMENTAL RESULTS

We have tested the proposed algorithm on two test sequences, "*Dancer*" and "*Ballons*", which are generated by computer graphics and depth estimation respectively [3]. For the sequence of "*Dancer*", views 2 and 5 are selected as encoded views and view 3 is set as the virtual view to be synthesized. For the sequence of "*Ballons*", views 3 and 5 are selected as encoded views and view 4 is set as the virtual view to be synthesized. The proposed algorithm was implemented on JMVC 6.0 [10]. The original version of JMVC 6.0 is considered as anchor to evaluate the performance of our algorithm. The encoding QP is set as 32,36,38,42. The delta QP, differential QP between the basis layer and sub-layer in hierarchical-B picture structure, is set as zero to all layers. We use the rendering software of MPEG 3DV [11] and the texture videos are not encoded.

We implement the proposed GMA on two schemes: 1) encoding the depth with full resolution; 2) encoding the depth in reduced resolution version and including down/up sampling process. The down/up-sampling method is selected from the Scalable Video Coding (SVC) standard [12]. For the first scheme, the experimental results of original and proposed algorithms are given in Table 1 and Table 2. As the fact that the encoding method of the basis view (view 2 for *Dancer* and view 3 for *Ballons*) is the same, we only compare the result of enhanced view (view 5 for *Dancer* and *Ballons*). The bitrate reduction ratio ($\Delta Bitrate$) was calculated based on BD-PSNR [13].

Table 1 Results of sequence "*Dancer*" with Full resolution

QP	Bit-rate of depth		Reconstructed quality	
	original	proposed	original	proposed
32	63.92	62.63	40.89	42.96
36	92.86	91.98	42.89	45.42
38	113.95	112.32	44.15	46.73
42	181.05	160.06	46.80	48.68
Avg. gain	--	-34.34%	--	2.545

Table 2 Results of sequence "*Ballons*" with Full resolution

QP	Bit-rate of depth (kbps)		Reconstructed quality (dB)	
	original	proposed	original	proposed
32	94.68	90.89	36.15	35.98
36	165.91	159.65	38.81	38.74
38	221.77	221.31	40.50	40.41
42	384.40	382.60	42.91	42.89
Avg. gain	--	-0.52%	--	0.02

From Table 1, we can see that the proposed algorithm achieves bit-rate reduction of 34.34% and reconstructed quality gains of 2.545 dB. Table 2 shows the comparison results of sequence "*Ballons*", whose depth information is generated by depth estimation method which doesn't consider the inter-view restriction at all. In other words, the view-dependent geometric relationship is not revealed so much by this sequence. Even for this case, our proposed algorithm can also achieve the gain of 0.02dB. Table 3 and Table 4 show the experimental results for encoding the depth in reduced resolution version with down/up sampling processing. The results are compared with the original method in depth bit-rate and rendering quality.

Table 3 Results of "*Dancer*" with reduced resolution

QP	Bit-rate of depth (kbps)		Rendering quality (dB)	
	Original-full	Proposed-reduced	Original-full	Proposed-reduced
32	163.20	52.72	36.26	34.78
36	222.11	80.99	37.02	35.39
38	267.31	98.76	37.40	35.94
42	391.54	145.04	37.68	36.38
Avg. gain	--	-22.60%	--	0.165

Table 4 Results of "*Ballons*" with reduced resolution

QP	Bit-rate of depth (kbps)		Rendering quality (dB)	
	Original-full	Proposed-reduced	Original-full	Proposed-reduced
32	179.65	57.61	35.60	35.17
36	304.03	97.12	35.73	35.48
38	404.45	130.71	35.80	35.55
42	687.68	224.02	35.87	35.69
Avg. gain	--	-40.55%	--	0.512

From Table 3 and Table 4, we can see that the rendering qualtiy can be improved by 0.165dB and 0.512dB respectivity. Figure 4 gives an example of RD curve for the coding performance of our proposed algorithm.

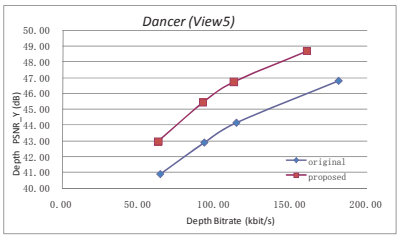


Figure 4 RD curves for enhanced inter-view prediction

From the experimental results above, we can conclude that the proposed multi-view depth video coding algorithm can have better depth coding performance and better rendering quality than the conversional method.

4. CONCULUSION

The data format of MVD, consisting of multi-view texture video and associated depth video, causes a vast amount of data to be stored or transmitted. There are geometric relationships of multi-view depth video. Inspired by the idea, this paper proposes a geometric mapping assisted inter-view prediction algorithm. To demonstrate the efficiency of our GMA prediction algorithm, we carry out the simulations in both the full resolution and reduced resolution with down/up sampling for each test sequence. Experimental results reveal the coding gains of up to 2.545 dB for reconstructed depth views and 0.512 dB for the synthesized views.

5. REFERENCES

[1] H. Ozaktas, L. Onural, *Three-Dimensional Television: Capture, Transmission, and Display*, Springer, Heidelberg, December 2007.

[2] Vetro, A.; Wiegand, T.; Sullivan, G.J.; , "Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard," *Proceedings of the IEEE* , vol.99, no.4, pp.626-642, April 2011

[3] ISO/IEC JTC1/SC29/WG11 N12036, "Call for Proposals on 3D Video Coding Technology", March 2011

[4] Ekmekcioglu, E.; Worrall, S.T.; Kondoz, A.M.; , "Bit-Rate Adaptive Downsampling for the Coding of Multi-View Video with Depth Information," *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, 2008 , vol., no., pp.137-140, 28-30 May 2008

[5] Yea, S., Vetro, A.: Multi-layered coding of depth for virtual view synthesis. *In:Picture Coding Symposium*, Chicago, IL (2009)

[6] Kwan-Jung Oh; Sehoon Yea; Vetro, A.; Yo-Sung Ho; , "Depth Reconstruction Filter and Down/Up Sampling for Depth Coding in 3-D Video," *Signal Processing Letters, IEEE*, vol.16, no.9, pp.747-750, Sept. 2009

[7] Morvan, Y., Farin, D., de With, P.H.N.: Depth-image compression based on an R-D optimized quadtree decomposition for the transmission of multiview images. *In: IEEE International Conference on Image Processing*, San Antonio,TX (2007)

[8] Faugeras, O. *Three-Dimensional Computer Vision_A Geometric Viewpoint,* Cambridge, MA:MIT Press. 1993

[9] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proc. SIGGRAPH 2000*, July 2000.

[10] Chen, Y., Pandit, P., Yea, S., Lim, C.S.: Draft Reference Software for MVC, *Joint VideoTeam (JVT) of ISO/IEC MPEG & ITU-T VCEG, ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6,* Doc. JVT-AE207, London (2009)

[11] Gun Bang , Min Soo Ko, Jisang Yoo, Gi-Mun Um, Won-Sik Cheong, Namho Hur, "Boundary noise removal and hole filling for VSRS3.5 alpha,"*ISO/IEC JTC1/SC29/WG11, M19992,* Jan 2011

[12] H. Schwarz, D. Marpe, T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103-1120, Sep 2007.

[13] G. Bjøntegaard, "Calculation of average PSNR differences between RD-Curves," *ITU-T SG16 Q.6 Document, VCEG-M33,* Austin, April 2001.