Three-Dimensional Image Completion For Depth Image Based Rendering

Chung-Te Li, Pei-Kuei Tsung, Kuan-Yu Chen, Chao-Chung Cheng, and Liang-Gee Chen
DSP/IC Design Lab, Graduate Institute of Electronics Engineering
National Taiwan University
Taipei, Taiwan
{ztdjbdy, iceworm, stevenky, fury, lgchen}@video.ee.ntu.edu.tw

Abstract—**3D video and the corresponding applications are getting more and more attentions in the recent years due to the vivid viewing experiment it can bring. In a 3DTV system, the occlusion handling is one main design challenge. The quality of the previous works is not good enough. Further, the computational complexity from the image based processing grows dramatically while the requirement of the virtual view numbers grows up. In this paper, a three-dimensional image completion scheme is proposed. By three-dimensional voxel transformation and completion, the 3D structure of the whole scene is generated and the occlusion part of virtual views from any view-point can be rendered based on the same voxel set. Furthermore, the proposed algorithm also provides 0.78 to 1.26 dB PSNR gains more than the PSNR from the prior-arts.**

Keywords-3DTV; DIBR; vitural view syethesis; inpainting;

I. INTRODUCTION

Nowadays, 3D video is getting more and more attentions due to the vivid experiment it can bring. In the 3DTV system, depth image based rendering (DIBR) is the key technology to generate images for both eyes [4]. One major challenge in the DIBR is how to handle the occlusion part. In the recent years, some prior-arts are proposed.

Oh, K.-J. et al. provide a foreground-background depth-aware inpainting in [1], which considers depth information when inpainting holes. Chia-Ming Cheng et al. provide a trilateral filter based approach with directional inpainting in [2]. This kind of inpainting provide sharper image quality in hole region, with a trade-off when the directional prediction is wrong. C. Fehn provides a simple linear interpolation in [4]. The computation is very light. M. M. Oliveira et al. provides another edge preserving fast inpainting in [5]. However, the quality of the synthesized frame is still low due to the lack of 3D structural considerations. Furthermore, the conventional inpainting based schemes require a frame-level processing at every output virtual view frames. Therefore, the complexity grows dramatically while the required view number is getting higher and higher. This is not acceptable in free viewpoint 3DTV systems.

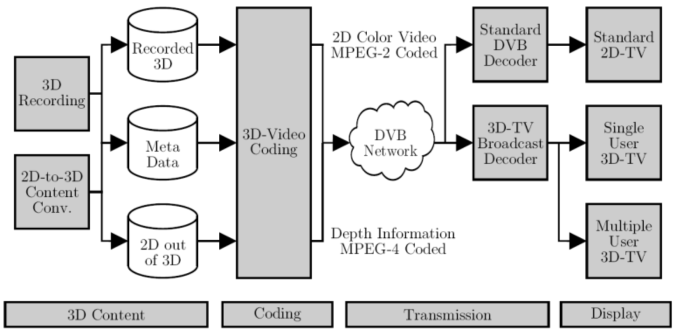

Figure 1. The ATTEST signal processing and data transmission chain. It consists of four different functional blocks: 1) 3D content generation; 2) 3D video coding; 3) Transmission; 4) Virtual view synthesis and 3D display. [7]

Besides, the qualities of 3D experience are also very important. Our target of this work is to provided a suitable algorithm which provides lighter computation when the required view number is high and better perceptual 3D experience to the viewers.

In this paper, a different kind of three-dimensional image completion scheme is proposed to solve the occlusion handling in a 3DTV system. This paper is organized as follows. Section II reviews the 3DTV system. Then, the problem statement and our proposed algorithm are provided in section III. In section IV, the experiment results are shown. Finally, we conclude in section V.

II. REIVEW OF 3DTV SYSTEM

In 2002, The European IST project ATTEST is proposed [6]. It is designed to be a novel, flexible, and backward-compatible broadcast 3DTV system. The details are shown in Figure 1. A complete 3DTV system is composed of 1) 3D content generation, 2) 3D video coding, 3) 3D video transmission, and 4) Virtual view synthesis.

Starting from ATTEST, virtual views of the 3D scene are able to be generated at the receiver side by depth-image-based rendering (DIBR) techniques. Show different virtual views to left and right eyes of the viewer provide 3D experiences.

According to the efforts of researchers, the quality of rendered virtual view is quite important for depth perception. In [8], an extension of 2D-metric which involves the measure of the disparity map distortion was proposed and tested with a methodology for subjective assessment (SAMVIQ) of stereo images. Original left view, right view and ground truth depth map are required.

In [9] and [10], stereo sense assessments based on HVS are further proposed. Their works show the evidence that virtual view quality is highly related to perceptual quality of 3D experiences.

This work is mainly focusing on the part of virtual view synthesis and the corresponding quality of virtual views, which are highly dependent to perceptual quality of 3D experiences.

III. Depth-Based Image Rendering and Hole Filling

Problem Statement

Figure 2 shows the conventional view synthesis flow. Due to the promotion and 3D display, the required number virtual views First, all the reference views are warped to the virtual view by depth image based rendering (DIBR) techniques. Because the occlusion regions in virtual views do not exist in original view, the obtained virtual views are not complete. That is there are some holes in virtual views due to occlusion. An example of occlusion between original view and virtual views is shown in Figure 3. The pixels belong to region V is the occlusion region in virtual views.

Each warped frame contains holes due to the depth discontinuity like the blue part in Figure 2. Then, all warped frames are processed with simple or iterative post-processing to fill the holes. Nevertheless which filling algorithm is used, the more the number of hole is, the more the corresponding filling computation is. Thus, in a free viewpoint 3DTV system, the computation will grow incredibly large due to the mass increasing of required virtual views.

In the other hand, the quality of the virtual views are also very important, which is address in section II. Therefore, our targeting problem is to propose an algorithm, which provides better virtual view quality and less-overhead with the increasing of required virtual views in a short summary.

Proposed 3D Voxel Completion

In the above, the target characteristics of the proposed algorithm are mentioned. Because of the requirement of better visual quality, the proposed method is designed for reconstructing more reliable data. In the other hand, due to the computational concern, our proposed method is also designed for reconstructing reusable data for different virtual views. Considering both the two characteristics, our proposed method is considered in 3D voxel space.

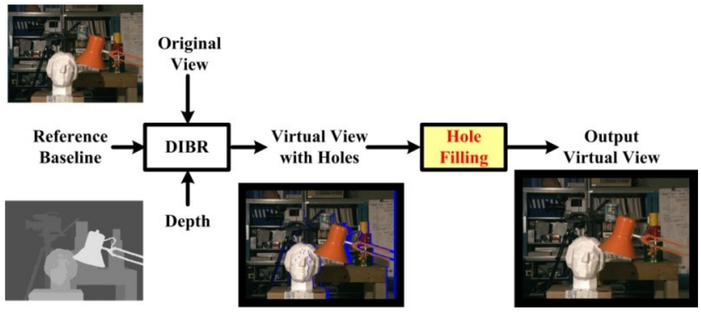

Figure 2. Conventional view synthesis flow

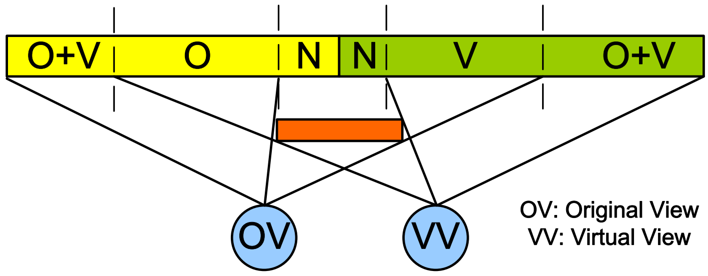

Figure 3. Occlusion discussions of original view and right view. Pixle labeled by O and V are visible in the orignal and virtual views, respecitvely Pixle labeled by N are invisible in both the first and second views.

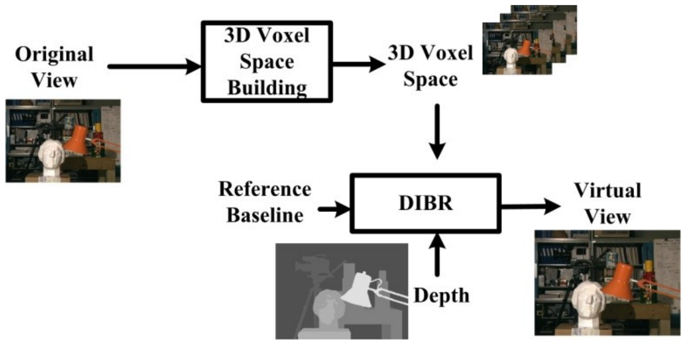

Figure 4. Proposed view synthesis flow

The flow of our algorithm is shown in Figure 4. First of all, a three-dimensional coordinate is generated by the frame width, height and the number of depth layers. In this new coordinate, each pixel is seen as a "voxel" based on its own depth value. These pixels from the input image frame are then set as the "surface" of the 3D scene structure. Second, all the voxels in the coordinate can be classified into three types: surface voxels, transparent voxels, and occlusion voxels. The transform function is:

$$\mathcal{V}(x, y, z) \triangleq \begin{cases} Surface, & if\ z = D(x,y) \\ Transparent, & if\ z > D(x,y) \\ Occlusion, & if\ z < D(x,y) \end{cases} \quad (1).$$

After the voxels transformation, the problem on the occlusion handling in the synthesized views can be translated to the occlusion voxels completion on the original views.

For each depth layer d = 0 to max
(from farthest to nearest)

1) Building a sub-image I_d, such that

$$I_d(x,y) = \begin{cases} V((x,y),D(x,y)) & \text{if} \quad D(x,y) < d \\ Holes & \text{if} \quad \text{others} \end{cases}$$

2) Inpainting I_d (with algorithm from [3] as example)
3) Filling the occlusion regions of V(x,y,z) with $V(x,y,d) = I_d(x,y)$ if $V(x,y,d)$ is occlusion (note: filling is only limited to occlusion parts defined in (1))

TABLE I. LAYER-BY-LAYER INPAINTING PROCEDURES FOR 3D VOXEL SPACE.

In this paper, voxels is filled in an up-down layer-by-layer order. In each depth layer, occlusion voxels are only filled by voxels on or below the current depth layer. Here the inpainting algorithm from [3] is used as the base-line algorithm. The detail layer-by-layer inpainting procedures are shown in Table I.

After the layer-by-layer hole filling in 3D voxel space, the occlusion regions in 3D voxel space are all filled. In Figure 4, an example of voxel space is shown layer by layer in original viewing direction. This example is divided into 16 depth layers, named L1 to L16. For transparent voxels, the nearest surface voxels with the same x-y coordination are shown. Therefore, the image shown in the most top layer, L16, is the same as the original frame.

In Figure 5, although there are lots of voxel are too vague, especially for voxels in the deepest layers, this does not ruin the final visual quality. That's because these voxels will not be chosen to shown in the virtual view when the baseline of the virtual view is selected suitably. That means the baselines of the virtual views are not too large. However, it's not very efficient when considering computation resources. All of the occlusion regions are inpainted. Given the settings of all the requiring views, the requirement of inpainting an occlusion voxel can be pre-derived. We try to adopt this concept to reduce the computation complexity.

Besides, with the increasing of the number of the virtual views, the 3D voxel space does not need to rebuild. Thus, the proposed method can be a useful solution for free viewpoint 3DTV, which requires large amounts of virtual views.

After the building of 3D voxel space, depth image based rendering is applied. Because there are no occlusion voxels more, just as shown in Figure 4, there is no requirement of the hole filling on rendered virtual views because all the 3D voxel space is complete.

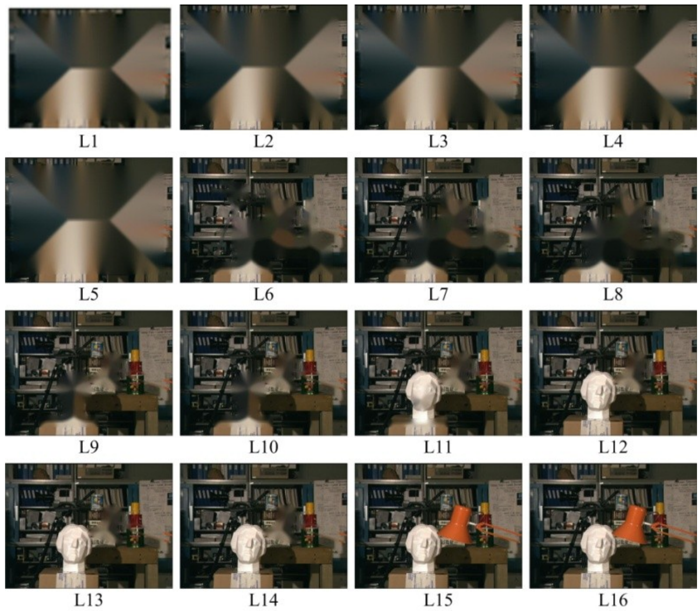

Figure 5. Filled voxels (shown layer-by-layer)

Method	**[1]**	**[2]**	**[4]**	**[5]**	**Proposed**
PSNR (dB)	29.5141	29.7571	29.2869	29.0043	30.5434

TABLE II. PSNR COMPARISON

In a short summary, we proposed a non-post-hole-filling virtual view synthesis method with building 3D voxel space and complete it. For visual quality, this method preserves more data, and never inpaints foreground voxels to backgrounds. For computational concern, the 3D voxel space is built only once even if large amounts of virtual views are required. Therefore, it's computational light when many virtual views are needed.

If the number of needing virtual views is n, the amortized computation for each virtual view on building 3D voxel space will become $\frac{1}{n}$.

IV. EXPERIMENT REUSLT

In this section, both subjective view and objective measurements are shown. For objective measurements, we select PSNR as a performance assessment tool.

The quality comparison on PSNR and subjective views are shown in Figure 6 and Table II. The hole filling methods in [1], [2], [4], and [5] are selected as the reference methods. The proposed method provides 0.78 to 1.26 dB PSNR improvement. In the subjective view comparison, the synthesized frame from the proposed algorithm also shows more details in the occlusion region among the prior-arts.

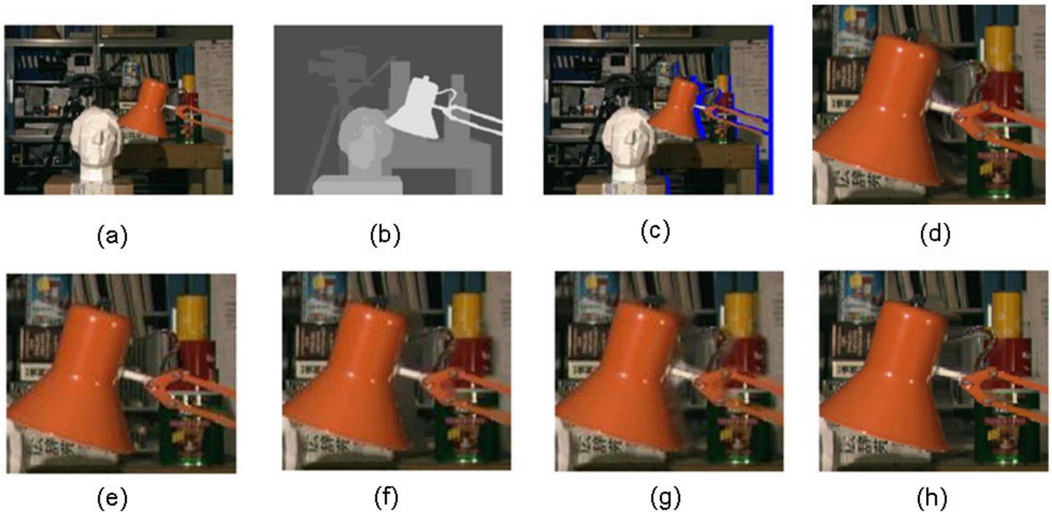

Figure 6. Experiment Results (a) inpit first view (b) depth (c) virtual view w.o. hole filling (d) [1] (e) [2] (f) [4] (g) [5] (h)proposed

V. CONCLUSION

In this paper, we provide a 3D voxel image completion method which is suitable for free viewpoint 3DTV with better visual quality and less computation overhead with the increasing of required virtual views.

In PSNR, the proposed method provides 0.78 to 1.26 dB PSNR improvement. In subjective view, the synthesized image from the proposed algorithm also shows more details among the prior-arts. Furthermore, after the voxel completion, the inpainting complexity can be reduced to only one time no matter how many virtual views are required. The main drawback is that compared to other existing methods, the computation will be too much when only requiring a single virtual view. One of our future works is to optimize the proposed algorithm in such cases with preserving the better visual quality at the same time.

REFERENCES

[1] Oh, K.-J.; Yea, S.; Ho, Y.-S., "Hole-Filling Method Using Depth Based In-Painting for View Synthesis in Free Viewpoint Television (FTV) and 3D Video", *Picture Coding Symposium (PCS)*, May 2009

[2] Chia-Ming Cheng; Shu-Jyuan Lin; Shang-Hong Lai; Jinn-Cherng Yang, "Improved novel view synthesis from depth image with large baseline," *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on* , vol., no., pp.1-4, 8-11 Dec. 2008

[3] A. Telea, "An Image Inpainting Technique Based on the Fast Marching Method," *J. Graphics Tools*, vol. 9, no. 1, pp. 23-34, 2004.

[4] C. Fehn, "Depth-Image-Based Rendering (DIBR), the compression and Transmission for s New Approach on 3DTV", *SPIE Proc. Stereoscopic Displays and Virtual Reality Systems XI*, 2004, pp. 93-104.

[5] M. M. Oliveira, B. Bowen, R, McKenna, and Y.-S. Chang, "Fast Digital Image Inpainting", *Proc. Inter. Conf. Visualization, Imaging and Image Processing*, 2001, pp. 261-266.

[6] Redert, Marc Op de Beeck, Christoph Fehn, Wijnand IJsselsteijn, Marc Pollefeys, Luc Van Gool, Eyal Ofek, Ian Sexton, Philip Surman, "ATTEST: Advanced Three-dimensional Television System Technologies," *3D Data Processing Visualization and Transmission, International Symposium* on, p. 313, First International Symposium on 3D Data Processing Visualization and Transmission (3DPVT'02), 2002

[7] Christoph Fehn, "A 3D-TV System Based On Video Plus Depth Information", Thirty-Seventh Asilomar Conference on Signals, *Systems and Computers*, vol. 2, pp. 1529-1533, Nov. 2003.

[8] A. Benoit, P. Le Callet, P. Campisi and R. Cousseau, "Quality Assessment of Stereoscopic Images", *EURASIP Journal on Image and Video Processing, special issue on 3D Image and Video Processing*, vol. 2008, Article ID 659024, 13 pages, 2008.

[9] J. Yang, Ch. Hou, "Objective Quality Assessment Method of Stereo Images", Proc. *IEEE 3DTV Conference*, Potsdam, Germany, May 2009.

[10] F. Lu, H. Wang, X. Ji, G. Er, "Quality Assessment of 3D Asymmetric View Coding Using Spatial Frequency Dominance Model", Proc. *IEEE 3DTV Conference*, Potsdam, Germany, May 2009.