# Real-time Stereo View Generation using Kinect Depth Camera

Sang-Beom Lee and Yo-Sung Ho
Gwangju Institute of Science and Technology (GIST)
*261 Cheomdan-gwagiro, Buk-gu, Gwangju, 500-712, Republic of Korea*
E-mail: {sblee, hoyo}@gist.ac.kr Tel: +82-62-715-2258

***Abstract*— In this paper, we propose a stereo view generation algorithm using the Kinect depth camera that utilizes the infrared structured light. After we capture a color image and its corresponding depth map, we preprocess the depth map. The preprocessed depth map is warped to the virtual viewpoint and filtered by median filtering to reduce the truncation error. Then, the color image is back-projected to the virtual viewpoint using the warped depth map. In order to fill out the remaining holes caused by disocclusion, we perform a background-based image in-painting operation. Finally, we obtain the synthesized image without any disocclusion. From experimental results, we have verified that the proposed algorithm generated the natural stereo images in nearly real-time.**

## I. INTRODUCTION

Television realized a human dream of watching a distant world in real-time. Moreover, it has been a great portion of visual system since it was invented. However, the flat scene is still different from the real world. In the conventional broadcasting system, we can only watch two-dimensional scene and we cannot feel the reality. Many TV researchers tried to develop high-definition TV (HDTV) during the last decade but they could not be an alternative to the three-dimensional scene.

We believe that the three-dimensional television (3DTV) is the next-generation broadcasting system in the history of TV. By aiding of advances in display devices, such as stereoscopic or multi-stereoscopic displays, 3DTV provides users with a feeling of 'being there', or presence, from the simulation of reality [1]. In this decade, we expect that the technology will be progressed enough to realize the 3DTV including content generation, coding, transmission, and display.

In 2002, the advanced three-dimensional television system technologies (ATTEST) project began the research for 3DTV [2]. ATTEST introduced a novel 3D broadcasting system including four main stages: 3D contents generation, coding, transmission, and rendering/display. While the previous approach dealt with two stereoscopic video streams - one for the left view and one for the right view - on the broadcasting system, ATTEST adopted novel two streams, the monoscopic video stream and the corresponding depth map stream that is composed of per-pixel depth information.

The virtual image can be synthesized by a depth image-based rendering (DIBR) technique using the color video and the corresponding depth video [3]. We can deal with the depth map as 3D information of the real scene. The virtual image can be generated by following procedures. First, the entire pixel of the color image of the original viewpoint is back-projected to the world coordinate using the camera geometry and the depth map. Then, the points in the world coordinate are re-projected on the image plane of the virtual viewpoint. This procedure is called "3D warping" in the computer graphics literature [4].

Although the DIBR technique is suitable for 3DTV, it has some problems. One of the most significant problems of the DIBR technique is that when we synthesize the virtual image, there are newly exposed areas, which are occluded in the original view but become visible in the virtual images. These areas are called disocclusion. The disocclusion is an annoying problem since the color image and the depth map cannot provide any information to naturally synthesize the virtual image. Therefore, the disocclusion should be filled out so that the virtual image seems more natural.

In order to remove the disocclusion, several solutions were introduced. Those methods are mainly categorized by two approaches: filling out the disocclusion by using near color information such as interpolation, extrapolation, and mirroring of background color, and preprocessing using a Gaussian smoothing filtering [3]. Recently, an asymmetric smoothing filtering is proposed [5]. This method reduces not only the disocclusion but also the geometric distortion that is caused by a symmetric smoothing filter.

While the disocclusion and the geometric distortion are mostly removed by the asymmetric depth map filtering, the synthesized view is deformed due to the distorted depth map. Recently, many solutions based on depth map filtering have been tried to solve the problem about the low depth quality. One of the solutions is the depth map filtering near the object boundary [6]. Although we can reduce the deformation of the depth map by restricting the filtered areas, the depth quality is still unsatisfied.

In this paper, we propose a stereo view generation algorithm. The main contribution of this paper is that we synthesize the virtual image using the original color image and the preprocessed depth map and also we realize the real-time process by aiming at the Kinect depth camera that utilizes the infrared structured light. The virtual image can be obtained by preprocessed depth map in virtual viewpoint and background-based image in-painting process.

## II. DEPTH IMAGE-BASED RENDERING (DIBR) TECHNIQUES

The color and depth video can be used for synthesizing the virtual images in DIBR technique. The block diagram of DIBR technique is depicted in Fig. 1. Each process is explained in detail in this section.

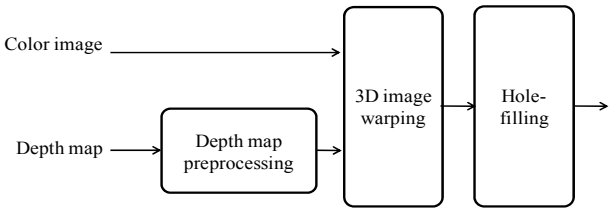
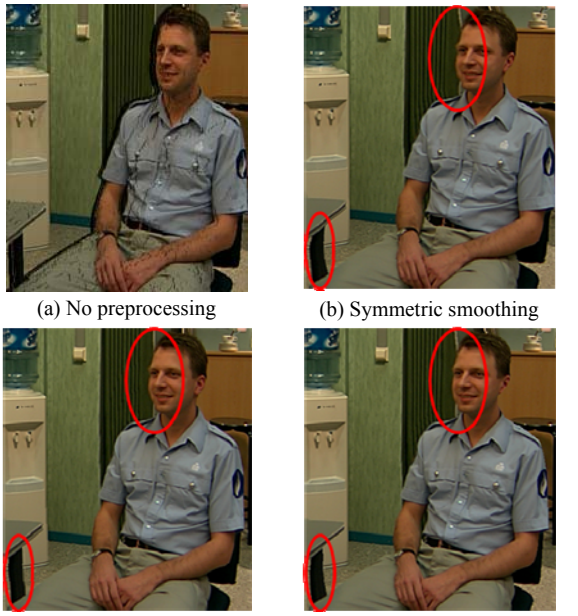



**Fig. 1.** Block diagram of depth image-based rendering technique

### A. Depth Map Preprocessing

When synthesizing the virtual image, we can find the disocclusion. Since there is no information of the disocclusion area, we need to fill out. One of the solutions is preprocessing of depth map using smoothing filter [3]. The main advantage of smoothing is that the sharpness of depth discontinuity is weakened and most disocclusion areas are filled with neighboring pixels.

Figure 2 shows various smoothing results for "Interview". As shown in Figure 2(b), the simple smoothing filter can fills out the disocclusion areas. However, it causes a geometric distortion that the vertical edges of the synthesized image are bent. This problem gives the discomfort to viewers. In order to reduce the geometric distortion, asymmetric smoothing method is proposed [5]. In this approach, the strength of filtering of a depth map in the horizontal direction is less than that in the vertical direction. Figure 2(c) shows the asymmetric smoothing result.



(a) No preprocessing (b) Symmetric smoothing

(c) Asymmetric smoothing (d) discontinuity-adaptive filtering

**Fig. 2.** Smoothing results for "Interview"

The synthesized image after asymmetric smoothing of the depth map has good subjective quality. However, the filtered depth map has many errors. It is desirable that the filter is applied so that the filtered areas are reduced through the prediction of the disocclusion areas. By aiming at this assumption, discontinuity-adaptive depth map filtering is proposed [6]. This approach assumes that the disocclusion area is detected nearby object boundaries and the depth map is filtered only near those regions. Therefore, the filtered region of the depth map is reduced. As shown in Fig. 2(d), the deformation of the object is reduced.

### B. 3D Image Warping

We assume that the camera configuration is parallel for simplicity. There are two approaches of stereoscopic image generation using DIBR technique. One is generating a virtual left image so that the original view is regarded as the right view. Another method is generating both the virtual left and right view by using original view. The first approach has the lowest quality of the left view since this view has the largest disocclusion areas compared to the second method. However, it gives us the highest quality for the right view. We adopt the first method since several conventional works proved that the binocular perception performance is determined by only one view which is higher quality than the other view [7].

Figure 3 shows the relationship of the pixel displacement and the real depth. The new coordinates $(x_l,y)$ of the virtual viewpoint from the original coordinates $(x_r,y)$ according to the depth value $Z$ is determined by

$$x_l = x_r + \frac{fB}{Z} \tag{1}$$

where $f$ represents the focal length of the camera and $B$ represents the distance between cameras.

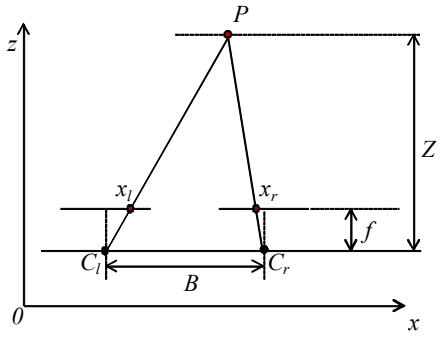



**Fig. 3.** Relationship between disparity and depth

### C. Hole-filling

After depth map preprocessing and 3D warping, most unknown regions of the virtual image are filled out. Due to the truncation error in the 3D warping process, the small-sized holes are remained. Therefore, we need to fill those holes. The common method in this step is linear interpolation using neighbor pixels.

## III. STEREO VIEW GENERATION METHOD

The proposed method exploits the Kinect depth camera, which interprets 3D scene information from a continuously-projected infrared structured light [8]. Figure 4 shows the overall block diagram of our algorithm. The first three steps are categorized by depth map preprocessing and remaining parts are included in the view synthesis operation.

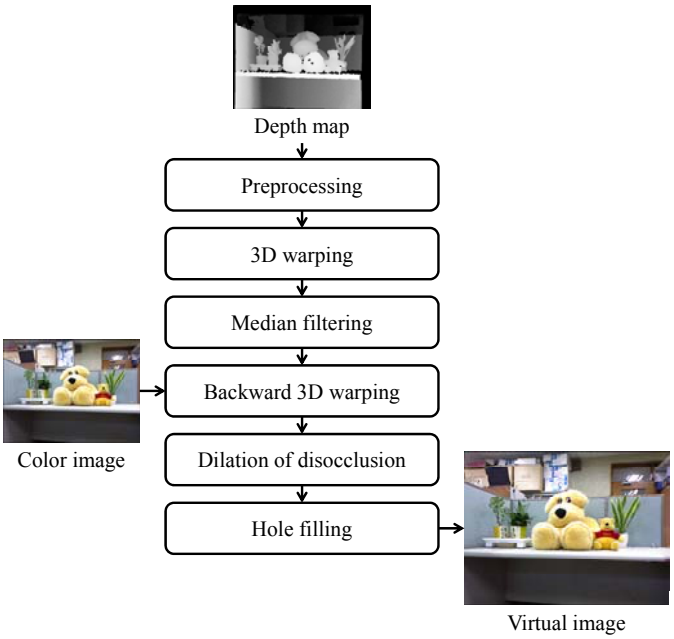
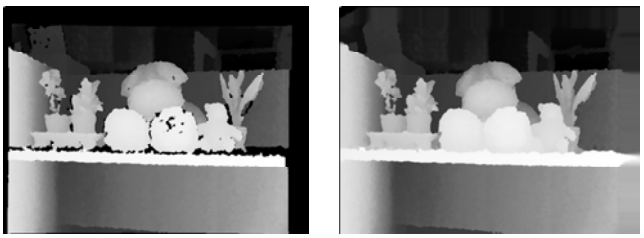



**Fig. 4.** Overall block diagram of the proposed method

### A. Depth Map Preprocessing

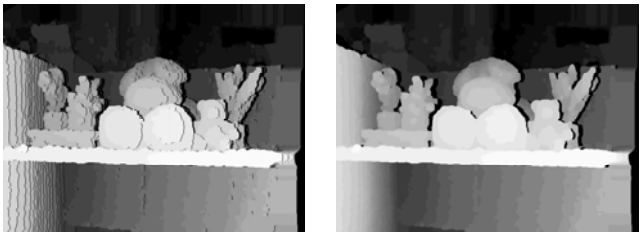
Since the position of the transmitter of infrared structured light and the receiver is different and there exist errors of the sensor itself, we obtain the depth map with some areas where the infrared sensor cannot retrieve the depths. Therefore, in the preprocess step, these areas are filled out by original image in-painting algorithm [9]. Figure 5 shows the depth preprocessing result using image in-painting algorithm.



(a) Original depth map (b) Preprocessed depth map

**Fig. 5.** Result of preprocessing

After the preprocessing, the 3D warping operation is performed using the depth map. During this step, the warped depth is truncated in integer value and as a result, the depth map includes truncation errors. These errors are easily removed by median filtering. Figure 6(a) shows the warped depth map and Fig. 6(b) shows the result of median filtering.



(a) warped depth map (b) result of median filtering

**Fig. 6.** Result of 3D warping

### B. Virtual View Synthesis

Using the warped depth map, the color image can be back-projected. It is computed by

$$I_{virtual}(x,y) = I_{original}\{x + D(x,y), y\} \quad (2)$$

where $D(x,y)$ represents the depth value at pixel position $(x,y)$. The back-projected color image is shown in Fig. 7. As shown in Fig. 7, most of pixels are filled but remaining holes near the object are found.

**Fig. 7.** Back-projected color image

In order to fill out those holes, we exploit the background in-painting operation. The in-painting algorithm first defines the region to be in-painted $\Omega$ and its boundary $\partial\Omega$ and the pixel $p$, the element of $\Omega$ is in-painted by its neighboring region $B_\varepsilon(p)$. In the proposed algorithm, we replace the boundaries facing the foreground with the corresponding background region located on the opposite side. This can be calculated by

$$p_{fg} \in \partial\Omega_{fg} \rightarrow p_{bg} \in \partial\Omega_{bg} \quad (3)$$

$$B_\varepsilon(p_{fg}) \rightarrow B_\varepsilon(p_{bg}) \quad (4)$$

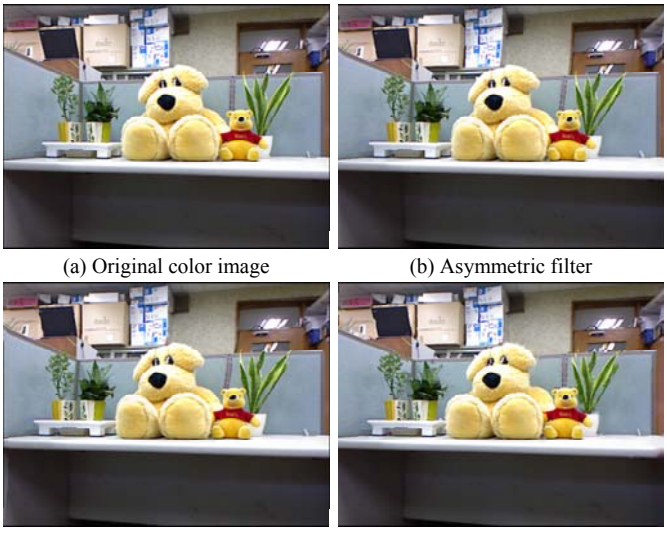
where *fg* and *bg* represent the foreground and the background, respectively.

## IV. EXPERIMENTAL RESULTS

We have evaluated the proposed algorithm with two aspects:

visual quality and computational time. The resolution of the color image and the depth map is 640×480. The parameters for stereo view generation are set by $B$ = 48 mm for the distance between cameras and $f$ = 200 mm for the focal length of the camera.

Figure 8 shows the experimental results of view synthesis. Figure 8(a) shows the original color image and Fig. 8(b), Fig. 8(c), and Fig. 8(d) represents the synthesis results of asymmetric filter, discontinuity-adaptive filter, and the proposed algorithm, respectively. As shown in Fig. 8(d), remaining holes are naturally removed compared to other methods since the proposed algorithm conducted the background-based image in-painting operation.



(a) Original color image (b) Asymmetric filter

(c) Discontinuity-adaptive filter (d) Proposed algorithm

**Fig. 8.** Results of view synthesis

Figure 9 shows the enlarged figures of Fig. 8. As shown in Fig. 9(a) and Fig. 9(b), there still remains the geometric errors in background. However, even though the proposed algorithm performed relatively unnatural hole-filling, it never deformed the depth map at all and caused geometric errors.

(a) Asymmetric filter (c) Discontinuity filter (c) Proposed algorithm

**Fig. 9.** Result of view synthesis

Table 1 shows the computational time of each process. From those results, the proposed system enabled the real-time processing up to 18.87 fps. Without any techniques for real-time processing, such as GPU programming or fast algorithms, stereo video was easily generated in real-time.

**Table 1.** Computational time

| Process | Computational time (ms) |
|---|---|
| Depth preprocessing | 12.00 |
| 3D warping | 11.00 |
| View synthesis | 5.00 |
| Background in-painting | 25.00 |
| **Total** | **53.00 (18.87 fps)** |

## V. CONCLUSION

In this paper, we have proposed a stereo view generation algorithm using the Kinect depth camera. The proposed scheme focused on the natural view synthesis and the real-time implementation. Therefore, we performed the depth preprocessing and view synthesis. The depth map is preprocessed by several image processing techniques and the synthesized image is obtained by background-based image in-painting operation. From the experimental results, we noticed that the proposed algorithm generated the natural stereo images and the entire process is implemented in nearly real-time.

## ACKNOWLEDGMENT

This research was supported in part by MKE under the ITRC support program supervised by NIPA (NIPA-2011-(C1090-1111-0003)).

## REFERENCES

[1] G. Riva, F. Davide, W. A. Ijsselsteijn, Being There: Concepts, Effects and Measurement of User Presence in Synthetic Environments, Ios Press, Amsterdam, Netherlands, 2003.

[2] A. Redert, M. O. Beeck, C. Fehn, W. IJsselsteijn, M. Pollefeys, L. Van Gool, E. Ofek, I. Sexton, P. Surman, "ATTEST: Advanced Three-dimensional Television System Techniques," *in Proc. of International Symposium on 3D Data Processing*, pp. 313-319, June 2002.

[3] C. Fehn, "Depth-Image-Based Rendering (DIBR), Compression and Transmission for a New Approach on 3-D TV," *in Proc. of SPIE Conf. Stereoscopic Displays and Virtual Reality Systems*, vol. 5291, pp. 93-104, Jan. 2004.

[4] W. R. Mark, L. McMillan, G. Bishop, "Post-Rendering 3D Warping," *in Proc. of Symposium on Interactive 3D Graphics*, pp. 7-16, April 1997.

[5] L. Zhang, W. J. Tam, "Stereoscopic Image Generation Based on Depth Images for 3D TV," *IEEE Trans. on Broadcasting*, vol. 51, pp. 191-199, June 2005.

[6] S. Lee and Y. Ho, "Discontinuity-adaptive Depth Map Filtering for 3D View Generation," *in Proc. of Immersive Telecommunications*, pp. T8(1-6), May 2009.

[7] L. Stelmach, W. Tam, D. Meegan, A. Vincent, P. Corriveau, "Human Perception of Mismatched Stereoscopic 3D Inputs," *in Proc. of International Conference on Image Processing*, vol. 1, pp. 5–8, Sept. 2000.

[8] PrimeSense, http://www.primesense.com/?p=487.

[9] A. Telea, "An Image Inpainting Technique based on The Fast Marching Method," *Journal Graphics Tools*, vol. 9, pp. 25-36, May 2004.