Discontinuity-adaptive Depth Map Filtering for 3D View Generation

Sang-Beom Lee
Gwangju Institute of Science and Technology (GIST)
261 Cheomdan-gwagiro,
Buk-gu, Gwangju
sblee@gist.ac.kr

Yo-Sung Ho
Gwangju Institute of Science and Technology (GIST)
261 Cheomdan-gwagiro,
Buk-gu, Gwangju
hoyo@gist.ac.kr

ABSTRACT

A depth image-based rendering (DIBR) technique is one of the rendering processes of virtual views with a color image and the corresponding depth map. The most important issue of DIBR is that the virtual view has no information at newly exposed areas, so called disocclusion. The general solution is to smooth the depth map using a Gaussian smoothing filter before 3D warping. However, the filtered depth map causes geometric distortion and the depth quality is seriously degraded. Therefore, we propose a new depth map filtering algorithm to solve the disocclusion problem while maintaining the depth quality. In order to preserve the visual quality of the virtual view, we smooth the depth map with further reduced deformation. After extracting object boundaries depending on the position of the virtual view, we apply a discontinuity-adaptive smoothing filter according to the distance of the object boundary and the amount of depth discontinuities. Finally, we obtain the depth map with higher quality compared to other methods. Experimental results showed that the disocclusion is efficiently removed and the visual quality of the virtual view is maintained.

Keywords

depth image-based rendering, three-dimensional television, discontinuity-adaptive filtering

1. 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, 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 video and the depth map streams [3][4]. We can deal with the depth map as 3D information of the real scene in the DIBR technique. The virtual view can be generated by following procedure. First, whole pixels 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 reprojected on the image plane of the virtual viewpoint. This procedure is called "3D warping" in the computer graphics literature [5].

The DIBR technique has several advantages compared to the conventional broadcasting system. The coding of a depth sequence is more efficient than that of a color sequence due to the low spatial and temporal variance of the depth map. This advantage of 3DTV system enables us to reduce the bandwidth required for transmission. Another advantage is that DIBR can synthesize any virtual views as if they were captured from several points of view.

Although the DIBR technique is suitable for 3DTV, it has some problems. The most significant problem of the DIBR technique is that when we synthesize the virtual view, we can see newly exposed areas, which are occluded in the original view but become visible in the virtual views. These areas are called disocclusion. The disocclusion is an annoying problem because the color image and the depth map cannot provide any information to naturally synthesize the virtual view. Therefore, the disocclusion should be filled out so that the virtual view 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, mirroring of background color, and preprocessing using a Gaussian smoothing filtering [3]. Recently, an asymmetric smoothing filtering is proposed for preprocessing [6][7]. This method reduces not only the disocclusion areas 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 [8][9]. Although we can reduce the deformation of the depth map by restricting the filtered areas, the depth quality is still unsatisfactory.

In this paper, we propose a new depth map filtering algorithm based on DIBR with further less degradation of a depth map. The main contribution of this paper is that we use the discontinuity-adaptive depth map filtering to reduce both disocclusion and geometric distortions. We apply an asymmetric smoothing filter to the depth map only nearby object boundaries. Especially, the filtered areas are predetermined by the depth discontinuity analysis.

2. DEPTH IMAGE-BASED RENDERING (DIBR) TECHNIQUES

The video and depth streams are used for synthesizing the virtual views in DIBR technique. The structure of DIBR technique is depicted in Figure 1.
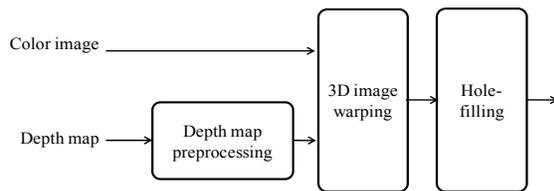
Figure 1. Depth image-based rendering technique

2.1 Depth Map Preprocessing

When we notice occlusion while synthesizing virtual views, we can easily resolve that problem: the points that are farther away from the camera position of the virtual viewpoint are substituted by the nearer points. However, when disocclusion is detected, the problem becomes more complex since there is no information about the disocclusion.

One of the solutions to remove the disocclusion is the layered depth image (LDI) [10]. LDI stores more than one pair of associated color and depth values for each pixel, with the number of layers typically depending on the scene complexity as well as the required synthesis quality. Although LDI seems to be a proper substitute, it has a main disadvantage that it is too much complicated for 3DTV broadcasting system.

Another solution is to preprocess the 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. Depth map preprocessing using smoothing filter has been studied by many researchers since the structure is simple and the synthesized image of the virtual viewpoint has a good visual quality in comparison with other algorithms.

Figure 2 shows various smoothing results for "Interview". As shown in Figure 2(b), applying simple smoothing filter fills out the disocclusion areas. However, the subjective quality of the virtual view is decreased in the case of the object including vertical edges since it causes a geometric distortion that the vertical edges of the synthesized image is bent. This problem gives the discomfort to viewers. From the Figure 2(b), we need to notice that the strength of filtering of a depth map in the horizontal direction should be less than that in the vertical direction. This is called asymmetric smoothing [6][7]. Figure 2(c) shows the asymmetric smoothing result.

The synthesized image after asymmetric smoothing of the depth map has a good subjective quality. However, there actually exist many errors since the depth map is seriously deformed. It is undesirable that we smooth out the whole depth map even if we can reduce the filtered areas through the prediction of the disocclusion areas. By aiming at this assumption, distance dependent filtering is proposed [8][9]. 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. Figure 3 shows the depth map after filtering. As shown in Figure 3(d), the deformation of the depth map is less than the depth map in Figure 3(c) and Figure 3(d). However, the geometric distortion still exists since this method exploits the symmetric filter as shown in Figure 2(d). Furthermore, since this method does not consider the strength of the depth discontinuity, the result of depth quality is still not sufficient.

(a) no preprocessing   (b) symmetric smoothing

(c) asymmetric smoothing   (d) distance dependent filtering
Figure 2. Smoothing results for "Interview"

(a) no preprocessing   (b) symmetric smoothing

(c) asymmetric smoothing   (d) distance dependent filtering
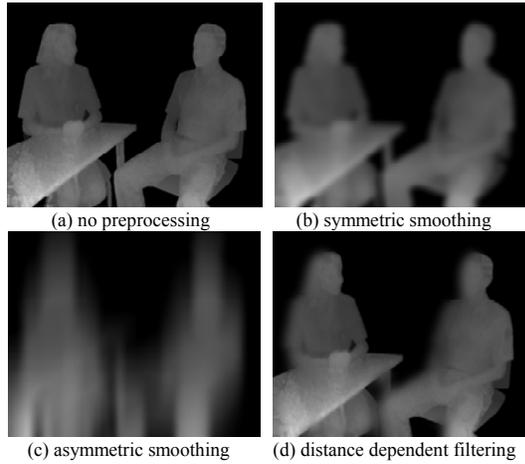
Figure 2. Smoothing results of depth map

2.2 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 view and the virtual 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 experimental results proved that the binocular perception performance is determined by only one view which is higher quality than the other view [11][12].

Figure 4 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.
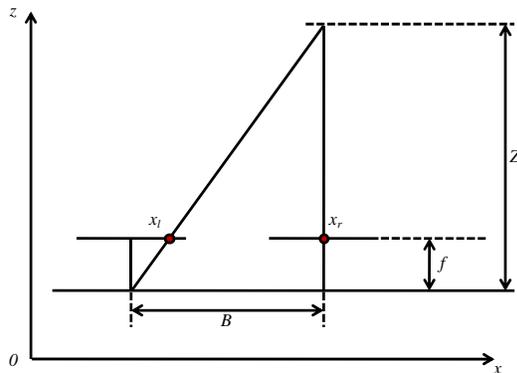
Figure 4. Relationship between disparity and depth

2.3 Hole-filling

After depth map preprocessing and 3D warping, most unknown regions of the virtual view 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.

3. PROPOSED DEPTH MAP FILTERING ALGORITHM

Figure 6 shows the block diagram of the proposed scheme. The basic framework is followed by the conventional algorithms but the remarkable difference is that the filtered regions are predicted by the depth discontinuity. Before filtering the depth map, we determine the filtered region by the depth edge detection and the depth discontinuity analysis. After 3D image warping and hole-filling, we obtain the virtual view. Each process will be described in detail.
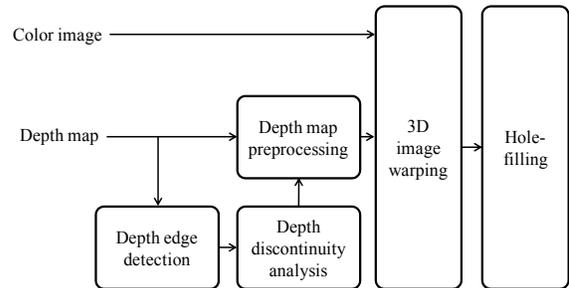
Figure 6. Block diagram of the proposed scheme

3.1 Depth Edge Detection

Disocclusion areas are detected near object boundaries with sharp depth discontinuity. Moreover, these areas are altered according to the position of the virtual viewpoint. Figure 7 shows the different disocclusion areas. For example, when the virtual camera is located at the left side of the original camera, we can see the disocclusion areas near the left boundaries of the object as shown in Figure7(a).

(a) virtual viewpoint (left)   (b) virtual viewpoint (right)

Figure 7. Different disocclusion areas

By aiming at this assumption, we can reduce the filtered area. In the proposed method, the object boundaries are extracted from the depth map first. Then, the smoothing is performed nearby the boundaries.

3.2 Depth Discontinuity Analysis

The distance dependent smoothing assumes that the filter is only applied nearby the object boundaries so that the deformation of the depth map is reduced. However, the filtered region is determined without considering the strength of depth

discontinuity. For example, if there are various depth discontinuities in the depth map, it is undesirable to apply filter only considering the distance since the disocclusion areas are related with the strength of the depth discontinuity. Therefore, if we analyze the amount of depth discontinuity of object boundaries and then determine the filtered areas, we can minimize the deformation of the depth map.

Suppose that the neighboring two pixels (x_{r1}, y_{r1}) and (x_{r2}, y_{r2}) of the original right view have the sharp depth discontinuity. Then, the length of the disocclusion D at the virtual left view can be calculated as follows:

$$D = x_{l_2} - x_{l_1} \qquad (2)$$

By using Eq. 1, Eq. 2 can also be defined by

$$D = x_{r_2} + \frac{fB}{Z_2} - \left(x_{r_1} + \frac{fB}{Z_1} \right) = 1 + fB \left(\frac{1}{Z_2} - \frac{1}{Z_1} \right) \qquad (3)$$

Here, $x_{r2} - x_{r1}$ is equal to 1. We notice that the length of the disocclusion of the virtual view can be predicted by the depth discontinuity from Eq. 3. In other words, we can apply smoothing filter with reduced deformation of the depth map by limiting the smoothed region.

3.3 Depth Map Preprocessing

In the proposed method, we utilize the asymmetric smoothing filter considering the depth discontinuity and the 4-neighbor distance from the point to the object boundary. According to the distance, the original depth map and the filtered depth map are blended.

The filtered depth at (x,y) in a depth map I is defined by

$$I_{filtered}(x, y) = \alpha(x, y) \cdot I_{original}(x, y) + \{1 - \alpha(x, y)\} \cdot I_{Gaussian}(x, y) \qquad (4)$$

$$\alpha(x+u, y+v) = \begin{cases} \dfrac{|x-u| + |y-v|}{D(x, y)} & \text{if } |x-u| + |y-v| < D(x, y) \\ 1 & \text{otherwise} \end{cases} \qquad (5)$$

where $D(x,y)$ represents the depth discontinuity at (x,y) and the range of u and v are $-D(x,y) \leq u \leq D(x,y)$ and $-3D(x,y) \leq v \leq 3D(x,y)$. Moreover, the window size is set to be three times larger than the variance of the filter. The reason why the window size in the vertical direction is three times larger than that in the horizontal direction is for the purpose of asymmetric smoothing. $I_{Gaussian}$ represents the filtered depth map by asymmetric smoothing filter. The filter is defined by

$$I_{Gaussian}(x, y) = \sum_{v=-3D(x,y)}^{3D(x,y)} \sum_{u=-D(x,y)}^{D(x,y)} I_{original}(x-u, y-v) \cdot G_{\sigma_u, \sigma_v}(u, v) \qquad (6)$$

$$G_{\sigma_x, \sigma_y}(x, y) = \frac{1}{2\pi \sigma_x \sigma_y} \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2} \right) \qquad (7)$$

Figure 8 shows the filtering results of the proposed scheme. As shown in Figure 8(a), we extract the edge of the depth map and analyze the strength of the depth discontinuity. Then, the filtered region is determined by the depth discontinuity and the

4-neighbor distance from the point to the object boundary as shown in Figure 8(b). We can obtain the smoothed depth map as shown in Figure 8(c). Figure 8(d) shows the remaining holes after 3D warping. Since the disocclusion areas are efficiently removed, we can fill out the remaining holes by interpolation using neighbor pixels.

(a) depth discontinuity   (b) filtered region

(c) filtered depth map   (d) remaining holes after filtering
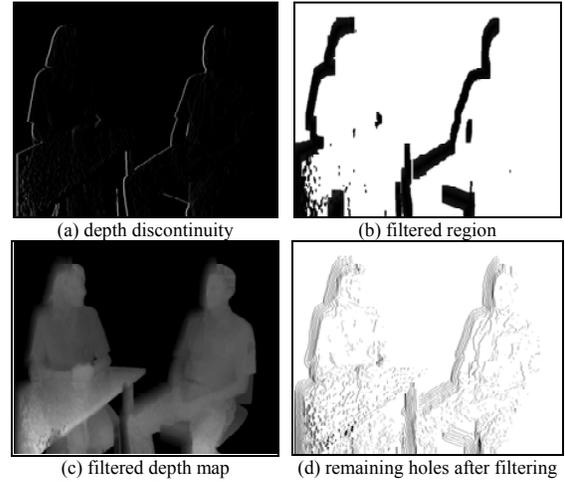
Figure 8. Depth filtering using depth discontinuity

(a) original color image   (b) original depth map

(c) asymmetric filter   (d) distance dependent filter

(e) proposed method
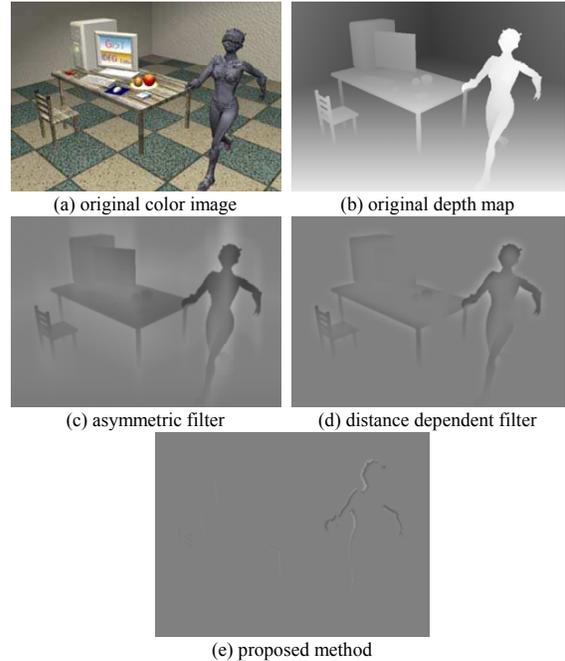
Figure 9. Difference image of depth maps

4. EXPERIMENTAL RESULTS

In order to evaluate the proposed scheme, we tested two pairs of sequences. One is "Interview" with 720×576 resolutions provided by Fraunhofer Heinrich-Hertz-Institut (HHI) in Germany. The other is "Sman" with 720×486 resolutions provided by Intelligent Design and Graphics Lab. at Gwangju

Institute of Science and Technology (GIST) in Korea. The "Interview" sequence is captured by the depth camera and "Sman" sequence is generated by computer graphics tool. The parameters for 3D view generation are set as follows: $B = 48$ mm for the distance between cameras and $f = 200$ mm for the focal length of the camera.

4.1 Depth Map Evaluation

For subjective evaluation of the depth map, we depicted the difference images of depth maps as shown in Figure 9. The difference images are shifted by +128 for better discrimination. We noticed that the proposed scheme reduced the filtered region compared with other algorithms since the smoothing filter is applied by the strength of depth discontinuity. As shown in Figure 9(c), the filtered region is variable along the object boundary. This aspect will be clearly seen in Figure 10.

Figure 10 shows the filtered regions of depth maps. The upper figures are results for "Interview" sequence and the lower figures are results for "Sman" sequence. The white areas in the figures are the deformed regions of depth map. As shown in Figure 10(c), the deformed region is apparently reduced in comparison with Figure 10(a) and Figure 10(b).
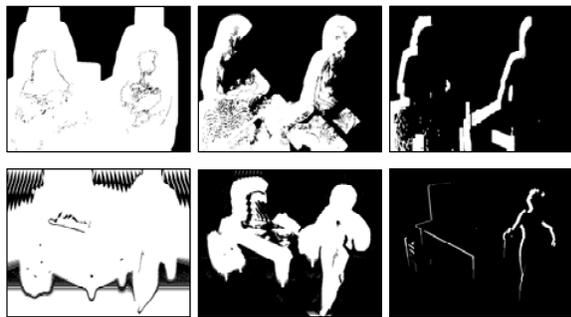
(a) asymmetric filter (b) distance dependent filter (c) proposed method

Figure 10. Filtered region of depth maps

For objective evaluation of depth quality, we performed PSNR comparison as other previous algorithms did. Figure 11 shows the PSNR comparison results for "Interview". Although asymmetric smoothing increased the subjective quality of the synthesized image, the depth quality is the lowest since asymmetric smoothing is applied to the whole depth map.
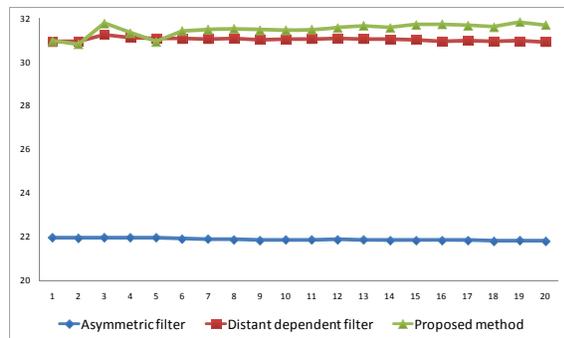
Figure 11. PSNR results for depth maps

On the other hand, the result of the proposed method outperformed the other algorithms.

4.2 Virtual View Evaluation

We compared the visual quality of the synthesized images of the virtual left view. For subjective evaluation, virtual view is illustrated in Figure 12. The black region in Figure 12(a) represents the disocclusion areas. From these results, the preprocessing efficiently removed the disocclusion areas. Especially, the proposed method naturally synthesized the virtual view. As shown in Figure 12(d), the proposed method gave a reasonable quality of the virtual view compared to other algorithms, even though the deformation of the depth map was reduced.

(a) no preprocessing   (b) asymmetric filter
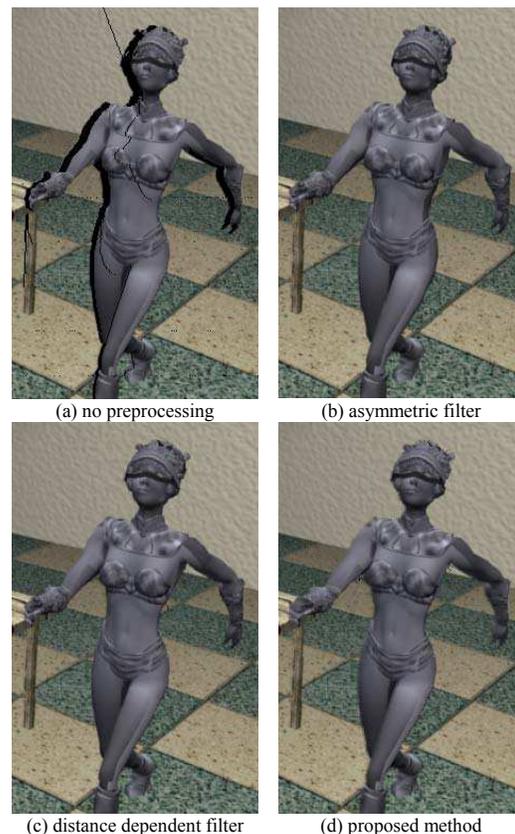
(c) distance dependent filter   (d) proposed method

Figure 12. Virtual view images

The synthesized images can also be evaluated by PSNR comparison [8]. The detailed comparison method is illustrated in Figure 13. First, by using the color image and the depth map, we synthesize the virtual view without depth map preprocessing. We regard this as an original image. Then, we synthesize the virtual view with depth map preprocessing and compare these two images in terms of PSNR. While calculating PSNR using these two images, we adopt only meaningful pixels except for those no information regions since the original image has a lot of disocclusion areas and holes. Therefore, we make a mask representing no information areas from the original image and calculate PSNR by using the original image and our result.
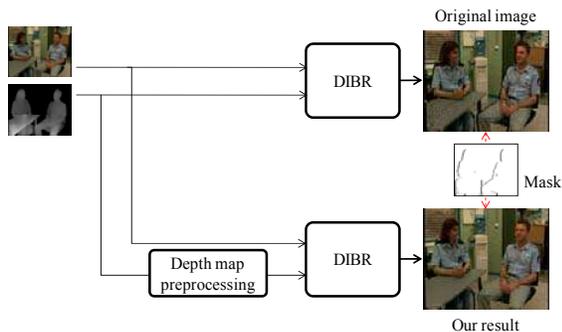
Figure 13. PSNR comparison method

From the comparison method shown in Figure 13, we obtained PSNR results for "Interview" as depicted in Figure 14. The PSNR of the proposed method improved compared to other algorithms as shown in Figure 14. As a result, the proposed method using discontinuity-adaptive filter gave us a better performance of both the depth quality and the visual quality of the virtual view.
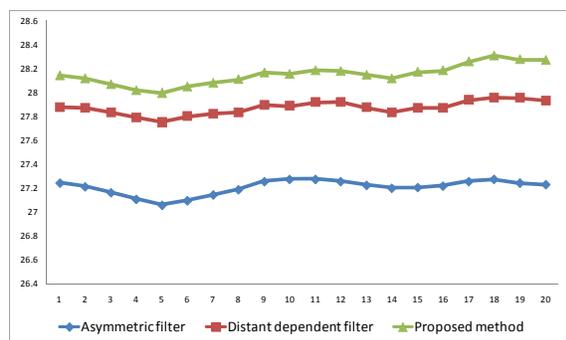
Figure 14. PSNR results for virtual views

5. CONCLUSIONS

In this paper, we have proposed a new depth map filtering algorithm for 3DTV. The proposed scheme focused on the minimization of the filtered region. Therefore, we filtered the depth map using a Gaussian smoothing filter with further reduced deformation. The edge detection was performed and the filtered region was determined by the amount of depth discontinuity and the distance between a point and the object boundary. From the experimental results, we noticed that the minimization of a filtered region improved both the depth quality and the virtual view quality. Finally, we could synthesize the natural virtual view and the results could be efficiently used for 3DTV system.

6. ACKNOWLEDGEMENTS

7. 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, Jun. 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] C. Fehn, K. Hopf, Q. Quante, "Key Technologies for an Advanced 3D-TV System," in Proc. of SPIE Three-Dimensional TV, Video and Display, pp. 66-80, Oct. 2004.

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

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

[7] W. J. Tam, L. Zhang, "Non-uniform Smoothing of Depth Maps before Image-based Rendering," in Proc. of ITCOM'04, vol. 5599, pp. 173-183, Oct. 2004.

[8] W.-Y. Chen, Y.-L. Chang, S.-F. Lin, L.-F. Ding, L.-G. Chen, "Efficient Depth Image Based Rendering with Edge Dependent Depth Filter and Interpolation," in Proc. of IEEE International Conference on Multimedia and Expo, pp. 1314-1317, July 2005.

[9] I. Daribo, C. Tillier, B. Pesquet-Popescu, "Distance Dependent Depth Filtering in 3D Warping for 3DTV," in Proc. of Multimedia Signal Processing, pp. 312-315, Oct. 2007.

[10] J. Shade, S. Gortler, L.-W. He, R. Szeliski, "Layered Depth Image," in Proc. of SIGGRAPH'98, pp. 231-242, Jul. 1998.

[11] 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, Sep. 2000.

[12] P. Seuntiens, L. Meesters, W. Ijsselsteijn, "Perceived Quality of Compressed Stereoscopic Images: Effects of Symmetric and Asymmetric JPEG Coding and Camera Separation," ACM Transactions on Applied Perception, vol. 3, no. 2, pp. 95–109, Jan. 2006.

[13] S.Y. Kim, S.B. Lee, Y.S. Ho, "Three-dimensional Natural Video System based on Layered Representation of Depth Maps," IEEE Trans. on Consumer Electronics, vol. 52, pp. 1035-1042, Aug. 2006.

[14] ZCam[TM] add-on, http://www.3dvsystems.com, 2005.

[15] C.L. Zitnick, S.B. Kang, M. Uyttendaele, S. Winder, R. Szeliski, "High-quality Video View Interpolation Using a Layered Representation," in Proc. of SIGGRAPH'04, pp. 600-608, Aug. 2004.