Bilateral Depth-Discontinuity Filter for Novel View Synthesis

Ismaël Daribo and Hideo Saito

Department of Information and Computer Science, Keio University,
3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan
{daribo, saito}@hvrl.ics.keio.ac.jp

***Abstract*—In this paper, a new filtering technique addresses the disocclusions problem issued from the depth image based rendering (DIBR) technique within 3DTV framework. An inherent problem with DIBR is to fill in the newly exposed areas (holes) caused by the image warping process. In opposition with multiview video (MVV) systems, such as free viewpoint television (FTV), where multiple reference views are used for recovering the disocclusions, we consider in this paper a 3DTV system based on a video-plus-depth sequence which provides only one reference view of the scene. To overcome this issue, disocclusion removal can be achieved by pre-processing the depth video and/or post-processing the warped image through hole-filling techniques. Specifically, we propose in this paper a pre-processing of the depth video based on a bilateral filtering according to the strength of the depth discontinuity. Experimental results are shown to illustrate the efficiency of the proposed method compared to the traditional methods.**

I. Introduction

The history of stereoscopy, stereoscopic imaging or three-dimensional (3D) imaging can be traced back to 1833 when Sir Charles Wheatstone created a mirror device that provides to the viewer the illusion of depth, in his description of the "Phenomena of Binocular Vision". Three-dimensional television (3DTV) has a long history, and over the years with the improvement of 3D technologies, more interest raised in 3DTV, which enables the depth perception of program entertainments without wearing special additional glasses.

A well suitable associated 3D video data representation is known as video-plus-depth that provides regular 2D videos enriched with their associated depth video. The 2D video provides the texture information, the color intensity, whereas the depth video represents the Z-distance per-pixel between the camera and a 3D point in the visual scene. For such purpose, a lot efforts have been performed for estimating depth information from multiple 2D video inputs. Furthermore, thanks to recent advances in semiconductor processes, it is also possible to directly capture depth video thereby using time-of-flight (TOF) camera [1], also known as depth camera.

Especially, depth image based rendering (DIBR) technique has been recognized as a promising tool for supporting advanced 3D video services required in 3DTV, by synthesizing some new novel views from the so-called video-plus-depth data representation. The most important problem in the generation of the novel views is to deal with the newly exposed areas, appearing as *holes* and denoted as *disocclusions*, which may be revealed in each novel images.

To deal with these disocclusions, one solution would be to rely on more complex multi-dimensional data representations, like layered depth image (LDI) data representation [2] that allows to store additional depth and color values for pixels that are occluded in the original view. This extra data provides the necessary information that is needed to fill in disoccluded areas in the rendered, novel views. However, this means increasing the overhead complexity of the system. On the other hand, disocclusions removal can be achieved by pre-processing the depth video in order to reduce the depth data discontinuities in a way that disocclusions decrease, followed by a post-processing of the warped image to replace the remained missing areas with some color informations.

In this paper a new filtering technique for depth image based rendering (DIBR) is proposed, allowing to reduce or completely remove the depth data discontinuities. This solution is based on a bilateral filter by taking into account the strength of the depth discontinuity.

The rest of the paper is organized as follows: in Section II, the stereoscopic rendering system is presented, followed by the disocclusion problem involved by the image warping process. Section III briefly reviews the bilateral filtering. Section IV addresses the problem of the recovery of disoccluded regions by pre-processing the depth video through the proposed bilateral filter. In Section V, the proposed system is experimentally evaluated, and finally, conclusions are drawn in Section VI.

II. Novel view synthesis

A. 3D image warping

This section describes the synthesis of a novel view through 3D image warping technique, one of the DIBR techniques [3], which enables the generation of the novel view. This includes a function for mapping points from the reference image plane to the targeted image plane.

Let us consider in this paper, a stereoscopic set-up with identical parallel cameras with known camera parameters (internal and external) (as illustrated in Fig. 1).

Conceptually, the 3D image warping process could be decomposed into two steps including a first back-projection of

MMSP'10, October 4-6, 2010, Saint-Malo, France.
978-1-4244-8112-5/10/$26.00 ©2010 IEEE.

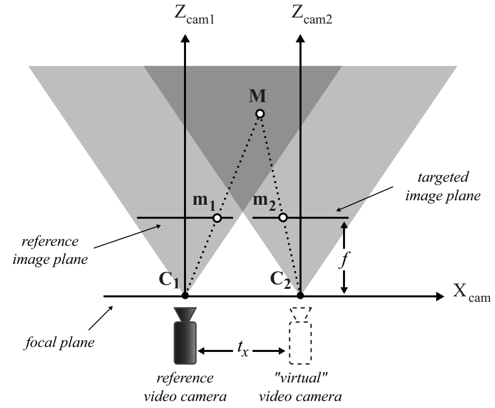


Fig. 1. Shift-sensor camera setup

the reference image into the 3D-world, followed by projecting the back-projected 3D scene onto the targeted image plane [3].

Considering the pixel location $\mathbf{m}_1 = (u_1, v_1)$, firstly a back-projection per-pixel is performed from the 2D reference camera image plane I_1 to the 3D-world coordinates. Then, a second projection from the 3D-world to the image plane I_2 of the novel targeted camera at pixel location $\mathbf{m}_2 = (u_2, v_2)$, and so on for each pixel location.

Then, the transformation that defines the new coordinates $\mathbf{m}_2$ in the novel view from the reference view at $\mathbf{m}_1$ according to the depth value $Z_{\mathbf{m}_1}$ can be expressed as a horizontal pixel displacement as follows:

$$\mathbf{m}_2 = \mathbf{m}_1 + \begin{pmatrix} \frac{f \cdot t_x}{Z_{s_1}} \\ 0 \end{pmatrix} \tag{1}$$

where t_x being the horizontal camera translation, and f being the focal length of the two video cameras.

B. The disocclusion problem

An inherent problem of the described 3D image warping algorithm is due to the fact that each pixel does not necessarily exist in both views. Consequently, due to the sharp discontinuities in the depth data (*i.e.* strong edges), the 3D image warping can expose areas of the scene that are occluded in the reference view and become visible in the second view. In Fig. 2, we can see the resulting warped picture from the 3D image warping process according to the camera set-up previously described. The disoccluded regions are colored in magenta.

Pre-processing the depth video allows to reduce the number and the size of the disoccluded regions, by meaning for example of smoothing, commonly operated with a Gaussian filter [4], [5]. Considering the fact that smoothing the whole depth video before 3D image warping damages more than simply applying a correction around the edges, a non-linear edge filter [6], a edge-distant dependent filter [7] and a discontinuity-based filter [8] have been proposed to reduce both disocclusions and filtering-induced distortions in the depth video.

More recently, bilateral filter [9] is used to enhance the depth video [10], [11] for his edge-preserving capability. Compared with conventional averaging or Gaussian-based filters, the bilateral filter operates in both spatial space and color intensity space, which results in preserving the sharp depth changes in conjunction with the intensity variation in color space, and in consistent boundaries between texture and depth images.

In this paper, we give more interest in the bilateral filter for his ability to consider multiple domains. We propose to extend this idea in the disocclusion domain to take into account the strength of the depth discontinuity. Let us first briefly review the bilateral filtering.

III. Bilateral filtering

A bilateral filter is an edge-preserving smoothing filter. Whereas many filters are convolutions in the spatial domain, a bilateral filter also operates in the intensity domain. Rather than simply replacing a pixel value with a weighted average of its neighbors, as for instance low-pass Gaussian filter does, the bilateral filter replaces a pixel value by a weighted average of its neighbors in both space and intensity domain. As a consequence, sharp edges are preserved by systematically excluding pixels across discontinuities from consideration.

The new depth value in the filtered depth map $\widetilde{Z}$ at the pixel location $\mathbf{s} = (u, v)$ is then defined by:

$$\widetilde{Z}_{\mathbf{s}} = \frac{1}{k(\mathbf{s})} \sum_{\mathbf{p} \in \Omega} f(\mathbf{p} - \mathbf{s}) g(Z_{\mathbf{p}} - Z_{\mathbf{s}}) Z_{\mathbf{p}}, \tag{2}$$

where Ω is the neighborhood around s under the convolution kernel, and $k(s)$ is a normalization term:

$$k(\mathbf{s}) = \sum_{\mathbf{p} \in \Omega} f(\mathbf{p} - \mathbf{s}) g(Z_{\mathbf{p}} - Z_{\mathbf{s}}). \tag{3}$$

In practice, discrete Gaussian function is used for the spatial filter f in the spatial domain, and for the range filter g in the intensity domain as follows:

$$f(\mathbf{p} - \mathbf{s}) = e^{-\frac{d(\mathbf{p}-\mathbf{s})^2}{2\sigma_d^2}}, \tag{4}$$

$$\text{with} \quad d(\mathbf{p} - \mathbf{s}) = \|\mathbf{p} - \mathbf{s}\|_2 , \tag{5}$$

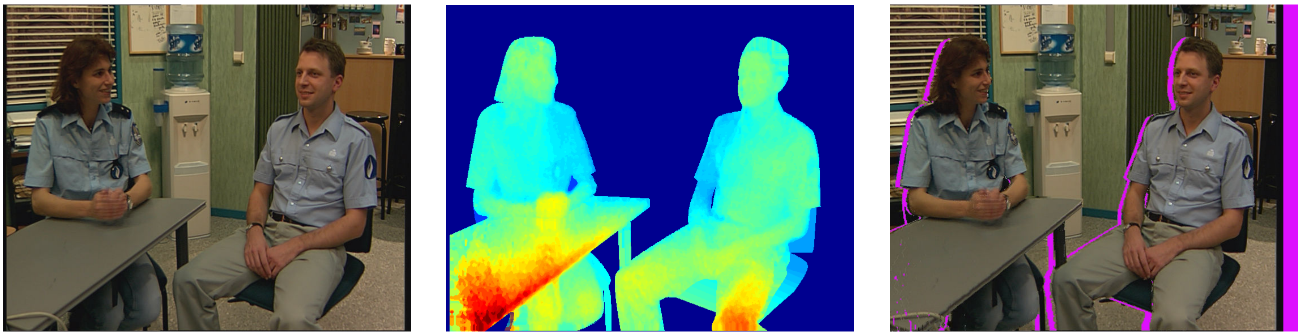
where $\|.\|_2$ is the Euclidean distance, and

$$g(Z_{\mathbf{p}} - Z_{\mathbf{s}}) = e^{-\frac{\delta(Z_{\mathbf{p}}-Z_{\mathbf{s}})^2}{2\sigma_r^2}}, \tag{6}$$

$$\text{with} \quad \delta(Z_{\mathbf{p}} - Z_{\mathbf{s}}) = |Z_{\mathbf{p}} - Z_{\mathbf{s}}| . \tag{7}$$

where σ_d and σ_r are the standard deviation of the spatial filter f and the range filter g, respectively. The extend of the filtering is then controlled by these two input parameters.

Therefore, a bilateral filter can be considered as a product of two Gaussian filters, where the value at a pixel location $\mathbf{s}$ is computed with a weight average of his neighbors by a spatial component that favors close pixels, and a range component that penalizes pixels with different intensity.



(a) reference image (b) depth map (c) warped image

Fig. 2. In magenta: newly exposed areas in the warped picture (from the ATTEST test sequence "Interview").

IV. Proposed depth map pre-processing

In order to address the disocclusion problem discussed in Section II-B, Gaussian-based filters have been proposed in the literature [4], [5], [6], [7], [8]. As we can see, in Fig. 3(a), the disocclusions have been totally removed from the novel view through an asymmetric Gaussian filter, but at the cost of high depth map degradation, resulting in decreasing the desired compelling depth perception on 3D display.

To overcome this issue, bilateral filter may be utilized for his abilities to better preserve depth informations. However, when considering only one reference view, the quality of the novel view rapidly decrease due to the disocclusions-filling-induced artifacts as we can see in Fig. 3(b). A trade-off has then to be find between preserving depth data details (for a compelling depth perception), and a high novel view synthesis quality.

In this paper, we give an attempt to combine adaptive approaches into bilateral filtering, whereby the disocclusions are removed by depth pre-processing, and in the meantime the compelling depth perception is preserved. We propose to extend the well-used Gaussian-based smoothing technique through a treatment inspired by the bilateral filter introduced by Tomasi *et al.* [9] described above, and a two-dimensional version of the discontinuity analysis of Lee *et al.* [8]. A new bilateral filtering method is then proposed that allows to take into account the strength of the depth discontinuity. Sharp discontinuities of interest in the depth video are then removed nearby disoccluded areas by considering disocclusion domain when adjusting weights of the bilateral filter. Other regions of the depth video are then filtered through the traditional bilateral filter, allowing to reduce for example coding-induced artifacts.

A. Discontinuity analysis

By considering two neighbor pixels $\mathbf{s}_1$ and $\mathbf{p}_1$ having a sharp depth discontinuity in the reference image, the disocclusion length ψ in the novel view can be computed as the Euclidean distance between the two warped pixels s_2 and p_2 in the novel view as follows:

$$\psi\left(\frac{1}{Z_{\mathbf{p}}} - \frac{1}{Z_{\mathbf{s}}}\right) = \|\mathbf{p}_2 - \mathbf{s}_2\|_2 \times H(\|\mathbf{p}_2 - \mathbf{s}_2\|_1) \qquad (8)$$

where H is the Heaviside step function whose value is zero for negative argument and one for positive argument. H allows to distinguish the side of the discontinuity. As illustrated in Fig. 2(c), the discontinuities of interest are located at the left side. $\|.\|_1$ and $\|.\|_2$ being the Minkowski distance [1] of order 1 and 2, respectively.

After solving by using Eq. (1), Eq. (8) ca be updated as:

$$\psi\left(\frac{1}{Z_{\mathbf{p}}} - \frac{1}{Z_{\mathbf{s}}}\right) = \left\|(\mathbf{p}_1 - \mathbf{s}_1) + t_x \cdot f \cdot \begin{pmatrix} 1/Z_{\mathbf{p}} - 1/Z_{\mathbf{s}} \\ 0 \end{pmatrix}\right\|_2 \times H(\|\mathbf{p}_1 - \mathbf{s}_1\|_1) \qquad (9)$$

B. Depth map bilateral filtering

In order to take into account the strength of the depth discontinuity, we propose to replace in Eq. (2) the edge-stopping function g by a discontinuity-smoothing function h that will increase the influence of depth pixels nearby disocclusions. The proposed discontinuity-based bilateral filter can then be expressed as follows:

$$\widetilde{Z}_{\mathbf{s}} = \frac{1}{k'(\mathbf{s})} \sum_{\mathbf{p}\in\Omega} f(\mathbf{p} - \mathbf{s}) h\left(\frac{1}{Z_{\mathbf{p}}} - \frac{1}{Z_{\mathbf{s}}}\right) Z_{\mathbf{p}}, \qquad (10)$$

where Ω is the neighborhood around $\mathbf{s}$ under the convolution kernel, and $k'(\mathbf{s})$ is the updated normalization term:

$$k'(\mathbf{s}) = \sum_{\mathbf{p}\in\Omega} f(\mathbf{p} - \mathbf{s}) h\left(\frac{1}{Z_{\mathbf{p}}} - \frac{1}{Z_{\mathbf{s}}}\right). \qquad (11)$$

The discontinuity-smoothing function h is expressed as follows:

$$h\left(\frac{1}{Z_{\mathbf{p}}} - \frac{1}{Z_{\mathbf{s}}}\right) = \begin{cases} e^{-\frac{\psi(1/Z_{\mathbf{p}} - 1/Z_{\mathbf{s}})^2}{2\sigma_r^2}}, & \text{if} \psi > \psi_{\min} \\ g(Z_{\mathbf{p}} - Z_{\mathbf{s}}), & \text{otherwise} \end{cases} \qquad (12)$$

with ψ defined in Eq. (9). $\psi_{\min}$ being the minimum disocclusion length which enables the smoothing of the discontinuities.

[1] For two 2D points (u_1, v_1) and (u_2, v_2) the Minkowski distance of order k is defined as: $\left(|u_1 - u_2|^k - |v_1 - v_2|^k\right)^{1/k}$. Euclidean distance being the Minkowski distance of order 2.

Therefore, the discontinuity-smoothing function h will enforces a smoothing of the depth video nearby disoccluded areas, and a strict preservation of the depth edges in non-disoccluded areas.

V. Experimental results

For the experiments, we have considered the Advanced Three-dimensional Television System Technologies (ATTEST) video-plus-depth sequence "Interview" (720×576, 25fps) [12]. The experimental parameters for the camera are t_x=48 mm for the horizontal camera translation and f=200 mm for the focal length, and concerning the bilateral filtering process the parameters σ_d and σ_r have been chosen respectively to 20 and 11.

Fig. 3 illustrates different results of the pre-processed depth video by applying an overall asymmetric Gaussian filtering, a bilateral filtering, and the proposed discontinuity-based bilateral filtering. One can see also the resulting warped image. In the case that depth pre-processing is not enough to suppress all the disocclusions, an inpaint-based hole-filling algorithm is applied afterwards to fill in the remained disocclusions. Especially, we performed the state-of-the-art inpainting algorithm developed by Bertalmio *et al.* [13] that uses the Navier-Stokes and fluid dynamics equations.

Let us consider the original depth map Z and the pre-processed depth map $\widetilde{Z}$, and also, the warped pictures I_{virt} and $\widetilde{I}_{\text{virt}}$ respectively according to Z and $\widetilde{Z}$.

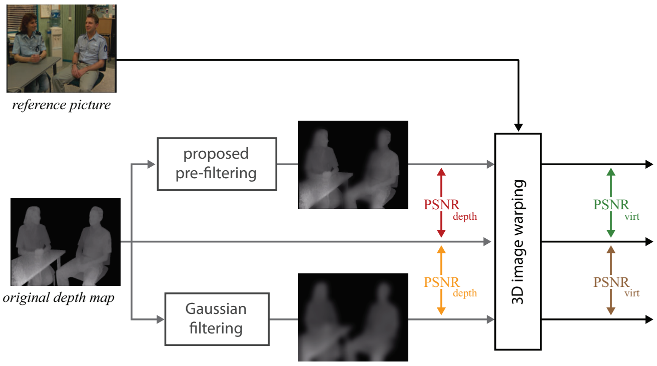


Fig. 4. Practical disposition of the two PSNR computations

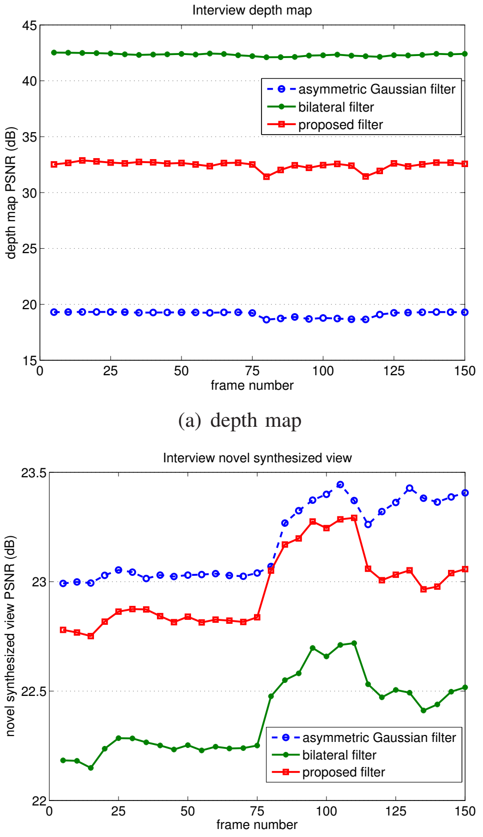
In order to measure the filtering-induced distortion in the depth map, and in the warped pictures, we defined two objective PSNR measurements before and after the 3D image warping (see Fig. 4) as follows:

$$\text{PSNR}_{\text{depth}} = \text{PSNR}(Z, \widetilde{Z}) \quad \text{and} \tag{13}$$

$$\text{PSNR}_{\text{virt}} = \text{PSNR}(I_{\text{virt}}, \widetilde{I}_{\text{virt}})_{\mathcal{D}\setminus\mathcal{O}\cup\widetilde{\mathcal{O}}} \tag{14}$$

where $\mathcal{D} \in \mathbb{N}^2$ is the discrete image support, $\mathcal{O}$ and $\widetilde{\mathcal{O}}$ being the occlusion image support of the 3D image warping using respectively Z and $\widetilde{Z}$.

$\text{PSNR}_{\text{depth}}$ is calculated between the original depth map and the filtered one. Hence, $\text{PSNR}_{\text{depth}}$ only considers coding artifacts in the depth map. However, it does not reflect the overall quality of the warped image. Then, $\text{PSNR}_{\text{virt}}$ is calculated between the warped image mapped with the decoded depth map and the original depth map. In order not to introduce in the $\text{PSNR}_{\text{virt}}$ measurement the warping-induced distortion, $\text{PSNR}_{\text{virt}}$ is computed only on the non-disoccluded areas $\mathcal{D}\setminus\mathcal{O}\cup\widetilde{\mathcal{O}}$.



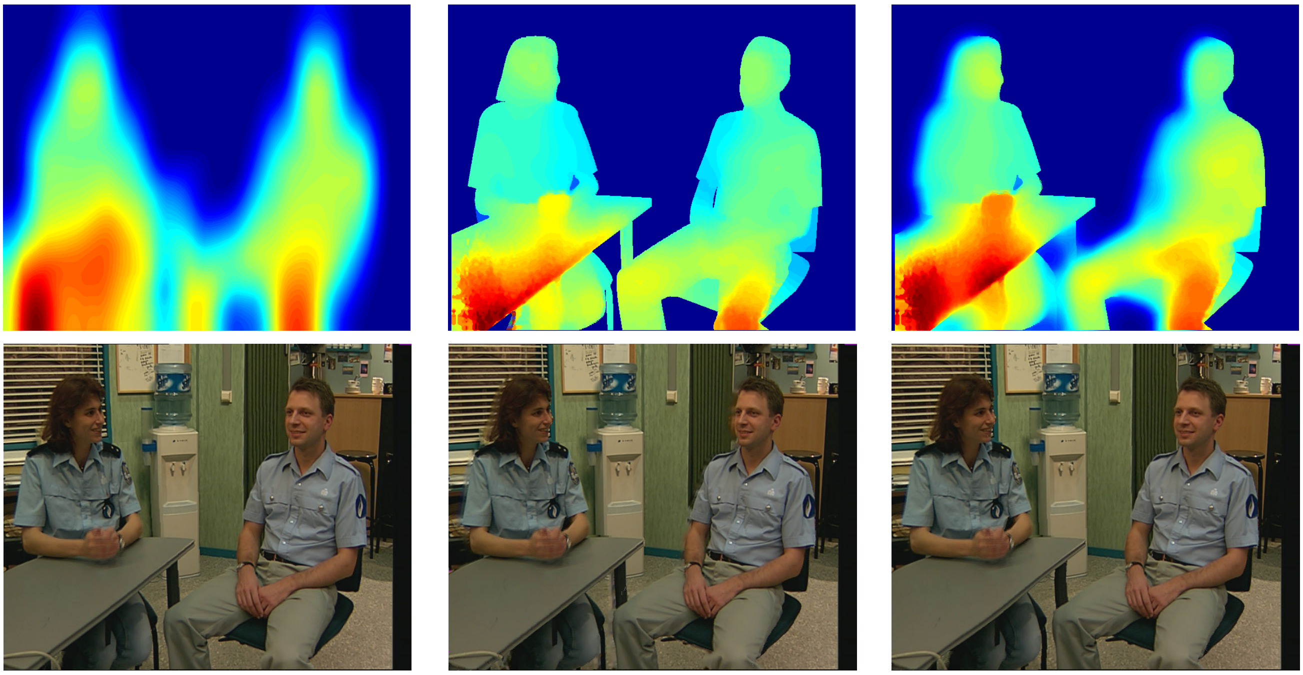
(a) depth map

(b) synthesized novel view

Fig. 5. Objective PSNR results.

Although the good visual quality of the warped image that used an asymmetric Gaussian filtering, the depth map suffers from a high distortion as illustrated in Fig. 5(a). On the other hand, indeed bilateral filtering has the benefits to better preserve the depth discontinuity details, but at the cost of a low disocclusion removals, which require then the utilization of hole-filling techniques, and thus, the introduction of interpolation-induced distortion in the warped image (see Fig. 5(b)).

Our proposed filtering has the benefits to exploit both advantages of these previous approaches by smoothing the required depth discontinuities and preserving the other discontinuities, and in the meantime providing a smooth representation of the depth video. Moreover by also smoothing homogeneous areas, like traditional bilateral filter do, the proposed filtering is able to denoise the depth video, *e.g.* removing coding-induced artifacts. As illustrated in Fig. 5, our proposed method combines adaptive Gaussian-based techniques in bilateral filtering as a trade-off between the preservation of the depth map details, and the quality of the novel view.



(a) asymmetric Gaussian pre-filtering (b) bilateral pre-filtering (c) proposed pre-filtering

Fig. 3. Different depth map pre-processing, and the associated synthesized novel view. ((top row) pre-processed depth map, (bottom row) novel view).

VI. Conclusion

In this paper, we have addressed the problem of recovering disoccluded regions by pre-processing the depth video based on a bilateral filtering, well known for his ability to operates in different domains. The proposed filter has inherited the bilateral filter benefits, and in the meantime take into consideration the strength of the depth discontinuities. As a consequence, the depth map is then smoothed, and allows an edge preservation in non-disoccluded regions, and a discontinuity removal nearby disoccluded regions. Experimental results show that the visual quality of the warped image is improved, and in meantime the depth discontinuities are preserved in non-disoccluded areas. A trade-off has then be reached in both depth map and novel view domains that allows to better preserve the depth perception on 3D display and in the meantime enhance the quality of the synthesized novel view. Future studies will concentrate on temporal coherence in order to ensure spatial and temporal stability.

VII. Acknowledgments

This work is partially supported by National Institute of Information and Communications Technology (NICT), Japan.

References

[1] T. Oggier, M. Lehmann, R. M. Rolf Kaufmann M.S., P. Metzler, G. Lang, F. Lustenberger, and N. Blanc, "An all-solid-state optical range camera for 3D real-time imaging with sub-centimeter depth resolution (swissranger)," in *SPIE, conference on optical system design*, 2003.

[2] J. W. Shade, S. J. Gortler, L.-W. He, and R. Szeliski, "Layered depth images," *Computer Graphics*, vol. 32, no. Annual Conference Series, pp. 231–242, Jul. 1998. [Online]. Available: http://grail.cs.washington.edu/projects/ldi/

[3] L. McMillan, Jr., "An image-based approach to three-dimensional computer graphics," Ph.D. dissertation, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, 1997.

[4] W. J. Tam, G. Alain, L. Zhang, T. Martin, and R. Renaud, "Smoothing depth maps for improved steroscopic image quality," in *Three-Dimensional TV, Video, and Display III*, B. Javidi and F. Okano, Eds., vol. 5599, 2004, pp. 162–172.

[5] L. Zhang and W. Tam, "Stereoscopic image generation based on depth images for 3D TV," *IEEE Transactions on Broadcasting*, vol. 51, no. 2, pp. 191–199, Jun. 2005.

[6] W.-Y. Chen, Y.-L. Chang, S.-F. Lin, L.-F. Ding, and L.-G. Chen, "Efficient depth image based rendering with edge dependent depth filter and interpolation," in *Proc. of the IEEE International Conference on Multimedia and Expo (ICME)*, 6-8 2005, pp. 1314–1317.

[7] I. Daribo, C. Tillier, and B. Pesquet-Popescu, "Distance dependent depth filtering in 3D warping for 3DTV," in *Proc. of the IEEE Workshop on Multimedia Signal Processing (MMSP)*, Crete, Greece, Oct. 2007, pp. 312–315.

[8] S.-B. Lee and Y.-S. Ho, "Discontinuity-adaptive depth map filtering for 3D view generation," in *Proc. of the 2nd International Conference on Immersive Telecommunications (IMMERSCOM)*. ICST, Brussels, Belgium, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2009, pp. 1–6.

[9] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. of th 6th International Conference on Computer Vision (ICCV)*, 1998, pp. 839–846.

[10] C.-M. Cheng, S.-J. Lin, S.-H. Lai, and J.-C. Yang, "Improved novel view synthesis from depth image with large baseline," in *Proc. of the 19th International Conference on Pattern Recognition (ICPR)*, 2008, pp. 1–4.

[11] O. Gangwal and R.-P. Berretty, "Depth map post-processing for 3D-TV," in *Digest of Technical Papers International Conference on Consumer Electronics (ICCE)*, Jan. 2009, pp. 1–2.

[12] C. Fehn, K. Schüür, I. Feldmann, P. Kauff, and A. Smolic, "Distribution of ATTEST test sequences for EE4 in MPEG 3DAV," ISO/IEC JTC1/SC29/WG11, M9219 doc., Dec. 2002.

[13] M. Bertalmio, A. Bertozzi, and G. Sapiro, "Navier-stokes, fluid dynamics, and image and video inpainting," in *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2001, pp. I–355–I–362 vol.1.