COMPARISON BETWEEN BLUR TRANSFER AND BLUR RE-GENERATION IN DEPTH IMAGE BASED RENDERING

Norishige Fukushima[1], *Naoki Kodera*[1], *Yutaka Ishibashi*[1], *and Masayuki Tanimoto*[2]

[1]Graduate School of Engineering, Nagoya Institute of Technology
[2]Nagoya Industrial Science Research Institute

ABSTRACT

An issue of object boundary reconstruction is important in depth image based rendering. For natural rendering, pixel values in an object boundary should be gradually changed from the background color to the foreground color. There are two approaches for this problem; one is blur regeneration by using alpha matting, and the other is blur transfer by using foreground biased dilation. The former requires three additional steps, i.e. foreground, background texture synthesis, and alpha channel estimation. The latter requires only one step, which is dilation for a depth map. To validate the effectiveness of these methods, we compare the matting method with the dilation method. In addition, we modify the dilation method to improve the quality of the synthesis view. Experimental results show that the improved method reaches the state-of-the-arts of the matting method, and its additional computational cost within 4 ms.

Index Terms **—** view synthesis, depth image based rendering, alpha matting, dilation

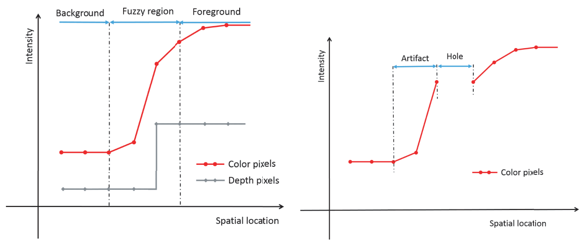
1. INTRODUCTION

Recently, free viewpoint image rendering has attracted increasing attention [1]. Free viewpoint images are expected as new 3D media. One of the rendering methods is depth image based rendering (DIBR) [2, 3], which synthesizes free viewpoint images from input images and associated depth maps. Usually, blurred regions around object boundaries have mixed foreground/background color into the other sides, and the regions can become artifacts in the synthesized view.

The problem is mainly caused by edge misalignment of object boundaries between a color image and a corresponding depth map. Object edges in the color image always contain transitional pixels, even if the object is in focus due to sampling and aliasing. The object edges in the associated depth map are basically located within the fuzzy regions of the corresponding color image, and the edges of the depth map do not have blurs (Fig. 1 (a)). After 3D warping, the blurred regions of natural images are split (Fig. 1 (b)), and the region become various artifacts.

There are 3 types of blur treatment; One is blur erasing type. Mori *et al.* [2] define the spilled regions by a hard threshold, and then erase the RGB image in the regions. Next, the erased regions are interpolated by using the other side view. Zinger *et al.* [3] proposed an extended method of [2] that adaptively erodes the spilled regions, and then eroded areas are interpolated by inpainting [4] with depth information. The erosion process absorbs the whole pixel information in the area, even if the pixel has useful information, such as fractional foreground and background colors. To save

This research was partially supported by JSPS KAKENHI Grant Number 25289117 and Hoso Bunka Foundation Grant.



(a) Color and depth value (b) Color pixels after 3D warping

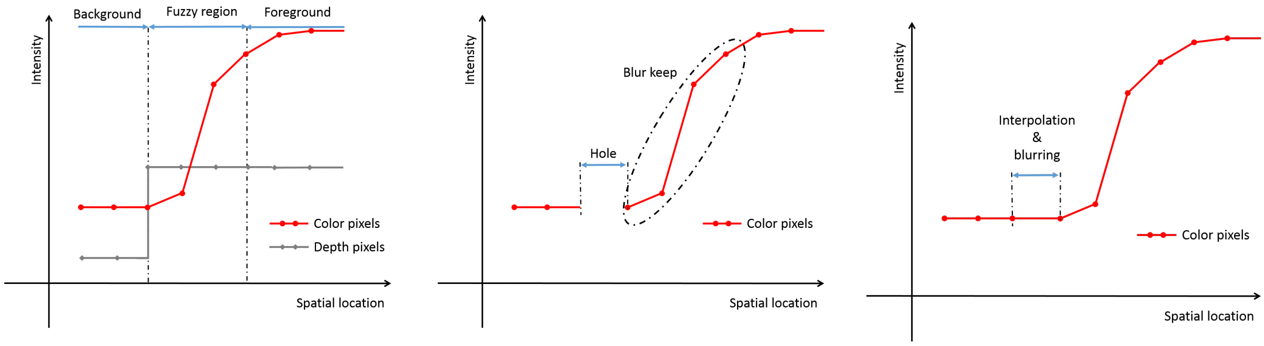
Figure 1. Misalignment of object boundaries between color image and depth map.

the erasing regions, we cannot avoid remaining blurred region or spilled region. In addition, these methods cannot reconstruct blur the object boundaries. To overcome these problems, following two types are used.

The other type is blur regeneration type [6, 7, 8], which use alpha matting [5]. Alpha matting can split blurred regions into foreground and background regions, and the matting based synthesis regenerates the blurred regions by rendering the split regions individually. Zitnick *et al.* [6] propose a method that computes matting information from all depth discontinuities, and use Bayesian matting [9]. Chan *et al.* [7] proposed a method which decomposes object boundaries by segmenting and tracking objects. These methods can perform accurate matting, but require much computational cost. In [8], the authors propose a method that computes alpha matting using guided filter [10], and evaluate synthesized image signals with/without the matting method including the conventional matting method. The method is computationally effective than the conventional matting method. In addition, the method has higher PSNR than the conventional accurate matting method. The matting-based view synthesis methods are effective, but the method requires 3 times DIBR processing than the non-matting-based method; foreground warping background warping and alpha mask warping.

Another approach is blur transfer type, which uses foreground biased warping [11]. The method just dilates input depth maps to shift blur from the reference view to the synthesized view. The method does not generate accurate blurred regions, but the rendered view smoothly connected between foreground and background. But subtle error exists in the rendered views.

For better boundary treatment, we compare the blur regeneration type with the blur transfer type. In addition, we will show that the blur transfer type can be dramatically improved by using a slight additional processing. The organization of this paper as follows. In Sect. 2, we show comparison methods and propose a blur transfer method. Sect. 3 compares the blur transfer type and the blur re-generation type. Finally, we conclude this paper in Sect. 4.



(a) Depth map dilating (b) After 3D warping (c) Hole interpolation and blurring

Figure 2. Color image and depth map conditions of each process.

2. VIEW SYNTHESIS METHODS

The traditional view synthesis [2] of DIBR involves three major steps: (1) depth map preprocessing, (2) 3D image warping, and (3) blending and hole-interpolating. For object boundary treatment, we review two types of the extended view synthesis methods and show an improved method (See Fig. 3).

2.1. blur re-generation type

The first one is the blur re-generation type, which generates blurred region uses alpha matting. The alpha matting splits a blur image into three images, which are foreground images, background images, and alpha masks. Each fore/back-ground image has not blur. In collaboration with the matting, the view synthesis performs 3 times DIBR processing of the step (1) to step (3) for the fore/back/alpha images. Then, a blending process of the foreground and background images with alpha mask is also required as an additional step. Blurs are regenerated in the last step of the alpha blending. We use the state-of-the-arts of blur regeneration method [8], which is computationally efficient and has high rendering quality. The bottleneck of the type is the matting process and 3 times DIBR processes (Fig. 3 upper).

2.2. Blur transfer type

The other is blur transfer type, which renders the blurred region bound with a foreground region at once. To bind the blurred region with the foreground region, the method adds a simple process before the step (1). That is a dilation filter for inputting left and right depth maps [11]. After the process, foreground depth values can cover the almost fuzzy/burred regions (Fig. 2 (a)). Warping with the dilated depth maps, the blurred regions at the boundary transfer directly to the target viewpoint, and artifacts are not generated as shown in Fig 2 (b). After the warping, each small hole is interpolated by the nearest background value (Fig. 2 (c)).

Improved blur transfer method: Our improved method adds a simple process to the DIBR step (3) of the blur transfer type. For smoothing signals between the interpolated region and the blur transferred region, we add a smoothing process for the boundary (Fig. 2 (c)). We first make a boundary mask between the foreground and background objects in the rendering view. For making the mask, we use the Canny filter for the generated depth map of the rendering view. Usually, we can obtain the generated depth from the intermediate process of DIBR. After generating the mask, we smooth the masked region by using the Gaussian filter

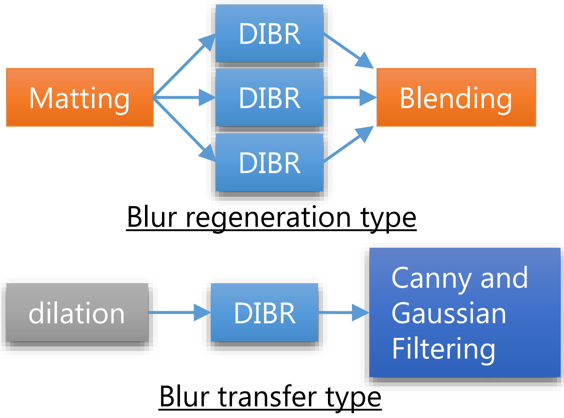


Figure 3. Two types of view synthesis for boundary treatment.

($\sigma = 0.3$). The slight blurring dramatically improves the quality of the rendering images.

3. EXPERIMENTAL RESULT

In the experiments, we evaluate the basic method [2], the matting based method [8], the dilation based [11] method, and its extension proposed in this paper. We use the Middlebury's data sets [12], and use left and right images as view synthesis anchors, and center images as just reference images for PSNR (peak signal-to-noise ratio) evaluation[1]. The input depth maps are the ground truth provided by the data sets. The code can be downloaded[2].

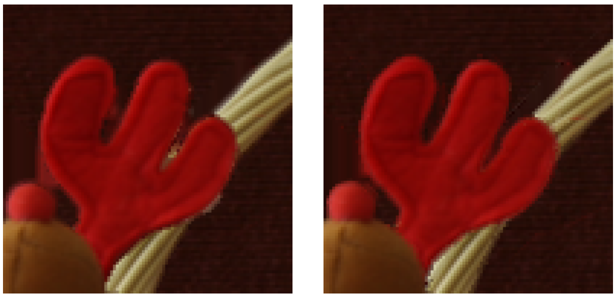
Table 1 shows the results of various datasets and their averaged values. In Table 1, the matting based method [8] has the highest PSNR in the average score. The second best is our proposed method. The difference between these methods are 0.31 dB in the averaged PSNR. In addition, the proposed blur transfer method has about 0.61 dB higher than the conventional method of the blur transfer type, and the additional computational time is within 4 ms. The additional steps are 2 dilations for input depth maps ($<$ 1ms), Canny filter for depth map on the synthesized image ($<$ 3ms), and Gaussian filter with a small kernel ($<$ 1ms). On the contrary, the matting method takes 3 times DIBR processing

[1] We also use SSIM (structural similarity) for the evaluation, however, differences among the competitive methods become small. Evaluation method for blurred region is our future work.

[2] http://nma.web.nitech.ac.jp/fukushima/research/viewsynthesis.html.

(a) Basic [2] (b) Matting [8] (c) Proposed (dilation + Gaussian)

Figure 4. Zoomed synthesized view of various methods.



(a) Proposed method (b) Matting method

Figure 5. Zoomed synthesized view (Reindeer).

(a) Proposed method (b) Matting method

Figure 6. Zoomed synthesized view (Bowling1).

(15 ms), a matting processing (55 ms), and a blending processing ($<$ 2ms). The blur regeneration type takes over 100 ms for rendering. Therefore the proposed method is $\times 5$ faster than the state-of-the arts with only losing 0.31 dB PSNR. Figure 4 shows the zoomed synthesized images of the "Teddy". We can observe contour artifacts in the basic method as shown in Fig 4 (a). The proposed method (dilation + Gaussian) can soften the contour artifacts, and the result looks like the rendering result of the matting method (Figs. 4 (b), (c)).

The proposed method has a limitation. If we have complex background color or background depth structure, we cannot render well (See Figs. 5,6.). Figure 5 shows the zoomed synthesized images at the "Reindeer" of the proposed and the matting method.

In the synthesized image of the proposed method, the blurs around the object boundary are contaminated with the color of the middle positional object. This reason is that the proposed method just dilates the foreground depth value, thus we cannot ignore the object in the middle position. As a result, the textures of the object in the middle position and the foreground object are mixed. Figure 6 shows the zoomed synthesized images at the "Bowling1" sequence of the proposed and the matting method. In this sequence, the proposed method has extremely lower PSNR than the matting method. We can observe the degraded region in the synthesized image of the proposed method. The reason is that the input depth maps are not dilated enough. In this case, the object edges in dilated depth maps exist in the foreground region or still blurred region. Thus, after 3D warping, blur around object boundaries in color image spills background color in the synthesized image, and the occluded area is interpolated by using around color information that comprehend foreground color or mixed color.

4. CONCLUSION

We improved a blur transfer type of depth image based rendering, and compared with a blur re-generation method. The results showed that the proposed method improves the subjective quality at object boundaries, and has 0.61 dB improvement in PSNR objectively. In addition, the proposed method reaches the state-of-the-arts of the blur re-generation method, which uses the alpha matting, and the difference of PSNR is 0.31 dB. In addition, the computational cost is about $\times 5$ effective.

In our future works, instead of using the ground truth depth maps, we use estimated depth maps, which have accurate boundaries, by using the semi-automatic method [13] or the boundary refinement methods [14]. Furthermore, we will consider the effects of coding distortions in the view synthesis process [15, 16].

5. REFERENCES

[1] A. Smolic, K. Mueller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, and T. Wiegand, "3D Video and Free Viewpoint Video-Technologies, Applications and MPEG Standards," in *Proc. IEEE International Conference on Multimedia and Expo (ICME)*, pp. 2161-2164, July 2006.

[2] Y. Mori, N. Fukushima, T. Yendo, T. Fujii, and M. Tanimoto, "View Generation with 3D Warping Using Depth Information for FTV," *Signal Processing: Image Communication*, 24 (1-2), pp. 65-72, Jan. 2009.

Table 1. PSNR (dB) of various images. The 4th column of the Prop. use additional dilation operation and Gaussian filtering for the basic method.

Data	Basic [2]	Matt. [8]	Max [11]	Prop.
Average	**35.62**	**38.85**	**37.93**	**38.54**
TeddyH	32.74	35.03	34.97	35.13
ConesH	29.70	31.07	29.24	31.40
Aloe	33.77	36.27	35.52	36.18
Art	30.70	35.05	33.62	34.89
Baby1	37.31	40.21	38.33	38.90
Baby2	36.75	38.75	38.47	38.79
Baby3	34.49	37.12	36.45	36.96
Books	33.45	37.06	36.52	36.73
Bowling1	34.42	39.72	35.46	35.93
Bowling2	33.68	36.74	34.32	34.61
Cloth1	43.15	43.30	43.17	43.44
Cloth2	36.48	42.18	41.27	42.24
Cloth3	37.74	40.33	40.58	41.39
Cloth4	35.35	37.38	37.04	37.52
Dolls	33.25	37.96	37.35	38.59
Flowerpots	31.29	31.86	31.43	31.69
Lampshade1	38.91	43.32	42.46	42.91
Lampshade2	39.76	44.16	42.46	43.02
Midd1	34.33	37.35	36.87	37.48
Midd2	33.42	38.71	36.40	37.28
Reindeer	34.37	38.14	36.75	38.15
Laundry	29.00	36.31	35.57	36.14
Moebius	35.70	40.18	39.89	40.67
Monopoly	35.65	37.92	37.52	37.69
Plastic	37.32	44.52	41.71	42.82
Rocks1	38.70	40.64	40.87	40.98
Rocks2	38.06	40.61	40.52	40.78
Wood1	43.21	44.90	45.21	45.40
Wood2	40.31	40.95	40.45	40.72
Venus	35.69	37.79	37.54	37.83

[3] S. Zinger, L. Do, and P.H.N. de With, "Free Viewpoint Depth Image Based Rendering," *Journal of Visual Communication and Image Representation, 21(5-6)*, pp. 533-541, July 2010.

[4] A. Telea, "An Image Inpainting Technique Based on The Fast Marching Method," *Journal of Graphics Tools*, 9(1), pp. 23-34, 2004.

[5] J. Wang and M. F. Cohen, "Image and Video Matting: A Survey," *Foundations and Trends in Computer Graphics and Vision*, 3(2), pp. 97-175, Jan. 2007.

[6] C. Zitnick, S.B. Kang, M. Uyttendaele, S, Winder, and R. Szeliski, "High-quality Video View Interpolation Using A Layered Representation," *ACM Trans. Graphics*, vol. 23 (3), Aug. 2004.

[7] S.C. Chan, Z.F. Gan, K.T. Ng, Ka-Leung Ho, and H.Y. Shum, "An Object- Based Approach to Image/Video-Based Synthesis and Processing for 3-D and Multiview Televisions," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 19 (6), June 2009.

[8] N. Kodera, N, Fukushima, and Y. Ishibashi, "Filter Based Alpha Matting for Depth Image Based Rendering," in *Proc. IEEE Visual Communications and Image Processing (VCIP)*, Nov. 2013.

[9] Y.Y. Chuang, B. Curless, D. H. Salesin, and R. Szeliski, "A Bayesian Approach to Digital Matting," in *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, 2001.

[10] K. He, J. Sun, and X. Tang, "Guided Image Filtering," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1397-1409, June 2013.

[11] X. Xu, L.-M. Po, K.-W. Cheung, K.-Ho Ng, K.-M. Wong, and C.-W. Ting, gA Foreground Biased Depth Map Refinement Method for DIBR View Synthesis," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp.805-808, Mar. 2012.

[12] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-frame Stereo Correspondence Algorithms," *International Journal of Computer Vision*, vol. 47, no. 1-3, pp. 7-42, Apr.-June 2002.

[13] M.O. Wildeboer, N. Fukushima, T. Yendo, M.P. Tehrani, T. Fujii, and M. Tanimoto, "A Semi-automatic Multi-view Depth Estimation Method," in *Proc. SPIE Visual Communications and Image Processing (VCIP)*, 2010.

[14] T. Matsuo, N. Fukushima, and Y. Ishibashi, "Weighted Joint Bilateral Filter with Slope Depth Compensation Filter for Depth Map Refinement," in *Proc. International Conference on Computer Vision Theory and Applications (VISAPP)*, Feb. 2013.

[15] N. Fukushima, T. Inoue, and Y. Ishibashi, "Removing Depth map Coding Distortion by Using Post Filter Set," in *Proc. IEEE International Conference on Multimedia and Expo (ICME)*, June 2013.

[16] T. Inoue, N. Fukushima, and Y. Ishibashi, " Non-essentiality of Correlation between Image and Depth Map in Free Viewpoint Image Coding: Accurate Depth Map Case," in *Proc. 3DTV-CON*, July 2014.