DISOCCLUSION USING DEPTH RELIABILITY MAP FOR VIEW SYNTHESIS

Yi Lai, Xuguang Lan, Yuehu Liu, Nanning Zheng

AIAR, Xi'an Jiaotong University, Xi'an, China, 710049
{ylai, xglan, yhliu, nnzheng }@aiar.xjtu.edu.cn

ABSTRACT

In this paper, a novel disocclusion scheme is proposed based on the depth reliability maps for virtual view synthesis. One important aspect of the developed method is that the occlusion map is estimated with the depth reliability maps. Therefore, the holes in a virtual view image can be filled in by blending the two predict images based on the occlusion map. The experimental results demonstrate that the addressed method has better performance for view synthesis than the other three conventional methods through the assessment of image quality and algorithm cost.

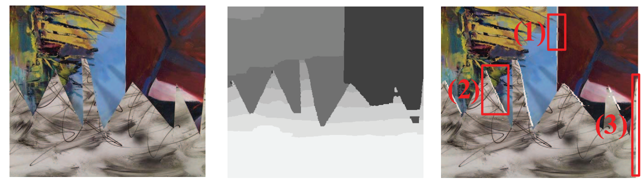
***Index Terms*—** Disocclusion, depth reliability map, view synthesis

1. INTRODUCTION

View synthesis is to generate an image that would be seen from a new viewpoint where no image is actually captured. It is a point technology for future free viewpoint television (FTV), 3D video compression, and other multi-view applications. Among a variety of techniques, depth-image based rendering (DIBR) is regarded as the most promising technology for rendering virtual views. The general DIBR process is depicted in Fig. 1. The most challenging problem for DIBR is how to deal with newly exposed areas (holes) appearing in the virtual view image.

The three kinds of holes possible in the synthesized image, which are separately denoted by (1), (2), and (3), are illustrated in Fig. 1 (c). (1) shows tearing, where a missing region appears in a oblique surface which is assigned to multiple depth layers; (2) demonstrates a region of true occlusion; (3) shows where occluded areas also occur when regions outside the field of view become visible [1]. Disocclusion is the attempt of recovering scene information of these holes by the visible parts in the reference view images.

To deal with these occlusion artifacts in the virtual view image, several schemes were reported. In [2], each occluded pixel was merged to an attached segment, and a pixel in a neighboring segment was selected to fill in the occluded region. And in [3], both the visible and hidden pixels were stored with a number of layers, so the holes areas could be diminished with the data from the hidden layers. The Bi-DIBR scheme separately warped two reference view images and then blended them to render an intermediate image [4]. Besides, an asymmetric Gaussian filter was used to smooth the sharp changes in depth map to reduce the artifacts [5]. A more complex inpainting technique was employed in [6].



(a) (b) (c)

Fig. 1. The general DIBR process for Sawtooth. (a) The reference image of view 2. (b) The depth map of (a). (c) The intermediate image synthesized by (a) and (b). The holes are shown in white. And the three kinds of holes possible in (c) are indicated by the red rectangle area, separately labeled by (1), (2), and (3).

Previous methods can eliminate some of the artifacts. However, none of them fills in the three kinds of holes simultaneously. In addition, the visual quality and the computation complexity can not be satisfied at the same time. In this paper, two contributions for the developed method are covered as following. First, an occlusion map can be produced based on the depth reliability maps. Second, with the weighted interpolation of the predict images based on the occlusion map, arbitrary view images along the baseline axis can be generated. As a result of our experiments, compared with other approaches, the virtual view image synthesized by the developed scheme achieves more satisfying image quality at an acceptable algorithm cost.

The rest of this paper is organized as follows. The architecture of the proposed method for view synthesis is treated in section 2. The occlusion map based view generation is addressed in section 3. Experimental results are shown in section 4. Finally, we conclude the whole paper in section 5.

2. ARCHITECTURE OF PROPOSED VIEW SYNTHESIS ALGORITHM

The block diagram of the presented view synthesis scheme is illustrated in Fig. 2. First, the left depth map and the right

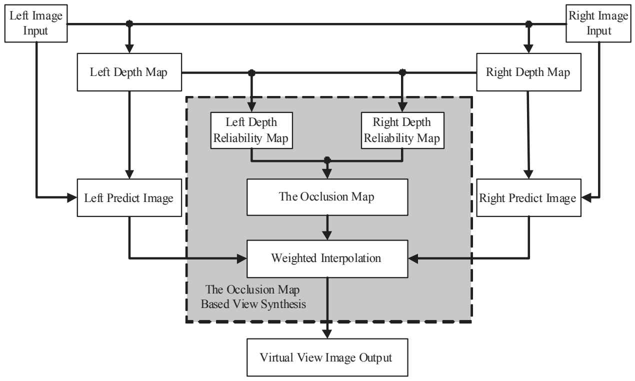


Fig. 2. The block diagram of the developed view synthesis algorithm.

one can be derived by a disparity estimation approach from the input reference images. Second, the left predict image and the right one can be calculated using the reference image with the associated depth map, respectively. In order to fill in the holes in a virtual view image to obtain better visual quality, we propose the occlusion map is derived by the depth reliability maps, which can be produced from the depth maps. At last, a weighted interpolation method is developed to generate a virtual view image with the occlusion map, the left and right predict images.

3. OCCLUSION MAP BASED VIEW SYNTHESIS

After finishing the estimation of depth maps from the left view image I_L and the right one I_R, the most crucial parts of diminishing the holes in a virtual view image I_P is to do two things. One is to find the position of the occluded pixels in I_P. The other is to decide how the holes are rendered efficiently. Therefore, the two key techniques for the addressed DIBR are discussed in this section.

3.1. Generating occlusion map

The pixels in I_P can be classified into three types: pixels visible in both I_L and I_R, pixels existing in one view but not in the other, and pixels invisible in both I_L and I_R. The number of the last kind of pixels is generally very small, so this type of holes can be eliminated with a median filter after synthesis. To render the other two kinds of pixels in I_P reasonably, the occlusion map based view generation scheme is proposed.

The depth of the occluded areas in the left view or right one can not be obtained correctly by a disparity estimation method. Hence, a reliability criterion of depth map can be used to as occlusion indicators. The reliability map is an efficient approach to measure the reliability of the depth map. Thus, the depth reliability map is proposed to distinguish occluded and visible areas.

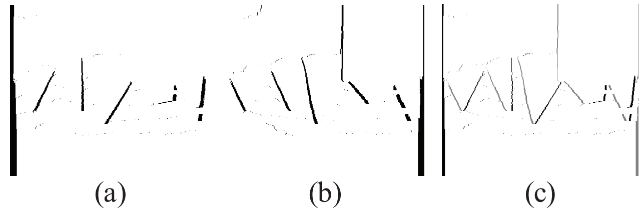


Fig. 3. The occlusion maps of Sawtooth. (a) The occlusion map of view 2. (b) The occlusion map of view 6. (c) The occlusion map derived from (a) and (b) in the virtual view (p = 0.50). The visible pixels are marked with white, and the occluded ones are marked with others.

First, the cross correspondence (CC) shows whether or not the acquired depth is valid by comparing the two depth maps. Under the rectified image assumption, point correspondences occur along a horizontal scan line. Here, the reliability map for the left view CC_L is taken as an example. The CC_L can be defined as follows [7]:

$$CC_L(x,y)=\begin{cases}1,\text{if } \| d^{L->R}(x,y)+d^{R->L}(x+d^{L->R}(x,y),y)\| \le 1\\ 0,\text{otherwise}\end{cases} \tag{1}$$

where $\|\bullet\|$ represents Euclidean norm. $d^{L->R}(x,y)$ denotes the disparity vector of the point (x,y) from left to right view, and $d^{R->L}(x,y)$ indicates the disparity vector of the point (x,y) from right to left view.

Second, a left reliability map Ω_L can be formulated by

$$\Omega_L=\{(x,y)\,|\,CC_L(x,y)=1\}. \tag{2}$$

This means that Ω_L is the set of pixels in I_L visible in both I_L and I_R. Hence, the set O_L in which pixels are visible in I_L but not in I_R is calculated as follows:

$$O_L=\overline{\Omega_L}=\{(x,y)\,|\,CC_L(x,y)=0\}. \tag{3}$$

In the similar way, the set O_R in which pixels are visible in I_R but not in I_L can also be calculated. The O_L and O_R for Sawtooth [8] are illustrated in Fig. 3 (a) and (b), respectively.

The virtual view image can be interpolated at arbitrary point along the baseline axis between the left and right view images. We define the parameter $p=0$ as the position of I_L, and $p=1$ as the position of I_R. Therefore, any $0<p<1$ is a valid position of an intermediate view. As a consequence, the occlusion map $O_P(x,y)$ corresponding to a position p can be computed by

$$O_P(x,y)=\{(x,y)\,|\,O_{P\mathrm{I}}(x,y)\cup O_{P\mathrm{II}}(x,y)\cup O_{P\mathrm{III}}(x,y)\} \tag{4}$$

where $O_{PI}(x,y)$, $O_{PII}(x,y)$, and $O_{PIII}(x,y)$ are separately defined as follows:

$$\begin{cases} O_{PI}(x,y) = \{(x,y) \mid (x + pd^{L->R}(x,y), y) \in O_L(x,y)\} \\ O_{PII}(x,y) = \{(x,y) \mid (x + pd^{L->R}(x,y), y) \in \overline{O_L} \\ \qquad \text{and } (x + (1-p)d^{R->L}(x,y), y) \in \overline{O_R}\} \\ O_{PIII}(x,y) = \{(x,y) \mid (x + (1-p)d^{R->L}(x,y), y) \in O_R\}. \end{cases} \tag{5}$$

where O_{PI} denotes the set of the pixels in I_P visible in I_L but not in I_R, O_{PII} represents the one visible in both I_L and I_R, and O_{PIII} indicates the one visible in I_R but not in I_L.

The set O_{PI}, O_{PII}, and O_{PIII} for Sawtooth are demonstrated in Fig.3 (c). The pixels belonging to O_{PI} are marked with black, the pixels belonging to O_{PII} are marked with white, and the pixels belonging to O_{PIII} are marked with gray.

3.2. Weighted interpolation based view synthesis

In order to fill in the holes, the occlusion map based view interpolation method is covered. First, the pixel locations are horizontally shifted by the disparity of the pixels, which are separately scaled by p or $1-p$ for the left and right view. Thus, the pixels are mapped from I_L and I_R into the virtual view. The two predict images, I_{PL} and I_{PR}, corresponding to I_L and I_R, are produced, respectively.

Second, the virtual view image I_P is blended from I_{PL} and I_{PR} using the interpolation weights ω_L and ω_R, so that $I_P = \omega_L I_{PL} + \omega_R I_{PR}$. The weight ω_R is set to be $1-\omega_L$ in order to realize the condition $\omega_L + \omega_R = 1$.

As for the set O_{PI}, the virtual image information is preferred from the left view image I_L, for the set O_{PII}, normal interpolation from both images is conducted, and for the set O_{PIII}, the right view image I_R is taken with greater preference. Hence, the following formula is introduced to calculate the weight ω_L:

$0 \le p \le 0.5$:

$$\omega_L(x,y) = \begin{cases} 1, \text{ if } (x,y) \in O_{PI} \\ 1-p, \text{ if } (x,y) \in O_{PII} \\ 1-2p, \text{ if } (x,y) \in O_{PIII} \end{cases}. \tag{6}$$

$p > 0.5$:

$$\omega_L(x,y) = \begin{cases} 2-2p, \text{ if } (x,y) \in O_{PI} \\ 1-p, \text{ if } (x,y) \in O_{PII} \\ 0, \text{ if } (x,y) \in O_{PIII} \end{cases}. \tag{7}$$

Therefore, the virtual view image with high quality can be generated by the weighted interpolation based on the occlusion map.

4. SIMULATION RESULTS AND ANALYSIS

To confirm the performance of our proposed view synthesis method, several experiments are conducted for four datasets from the Middlebury database [8], [9]. Each of the datasets contains 9 rectified stereo images, labeled view 0-8. Using view 2 and 6, we generate the intermediate 3 views (e.g. p = 0.25, 0.50, and 0.75). And the synthesized images are compared with the truth ones provided in the datasets. The tests are implemented with the Visual C++ 6.0 and performed on an Intel Core2 Duo 2.2-GHz computer that has 1GB memory. We compare the PSNR performance, subjective results, and computing time complexity of the developed algorithm with other algorithms, including the Bi-DIBR approach [4], the Gaussian filtering scheme [5], and the inpainting technique [6].

As shown in Fig. 4, the proposed method has the best PSNR performance compared with the other approaches. For Sawtooth, the PSNR value of the presented scheme is 1.69 dB superior to the Bi-DIBR, 3.71 dB to the Gaussian, and 1.68 dB to the inpainting. It is reasonable for two reasons. First, the occluded areas in the virtual view image can be detected efficiently by the depth reliability maps. Second, based on the occlusion map, the holes are filled in with the information from the left and right view images, so that a higher performance can be obtained than the other three methods. The similar improvements for the other three datasets are also illustrated in Fig. 4. Since Cones and Teddy have a bigger disparity, and the estimated depth maps of them are poor, the PSNR value of them is lower than that of Sawtooth and Venus.

Fig. 5 demonstrates the ground truth image and several subjective results of Sawtooth in view 4 (p = 0.50). As depicted in Fig. 5, the virtual view image generated by the developed algorithm has better visual quality than the other three methods. However, there are several noticeable artifacts in the outputs using the Bi-DIBR method and the inpainting approach, as shown in Fig. 5 (b) and (d), respectively. And the intermediate image rendered by the Gaussian algorithm has some serious geometric distortion as depicted in Fig. 5 (c). These results confirm the presented view synthesis scheme is more efficient.

To test the complexity of the discussed arrpoaches, the average computing time for the algorithms is evaluated for the four tested datasets. The computing time complexity is shown in Table I. Because the Bi-DIBR technique has enough information for filling in the occluded areas, it has the lowest computation complexity. The Gaussian scheme involves time consuming operations to apply a big size filter window to smooth the depth map. Due to searching for matching block to obtain a satisfactory image quality, the

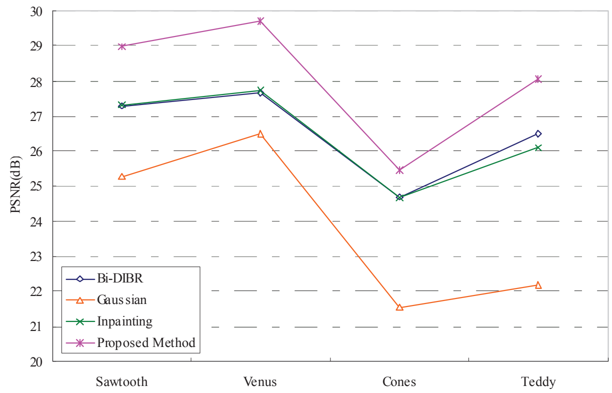


Fig. 4. Performance comparison of different virtual view synthesis methods for the four tested datasets, including Sawtooth, Venus, Cones, and Teddy.

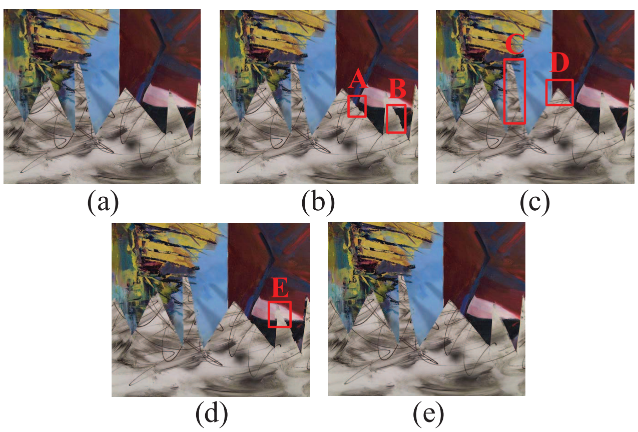


Fig. 5. The virtual view images generated by different methods in view 4 for Sawtooth. (a) Ground truth. (b) Bi-DIBR. (c) Gaussian. (d) Inpainting. (e) Proposed method. The artifacts are indicated by the red rectangle area, separately labeled by A, B, C, D, and E.

TABLE I

COMPUTING TIME COMPLEXITY

View Synthesis Method	Relative Time Cost
Bi-DIBR [4]	1
Gaussian [5]	1.39
Inpainting [6]	7.73
Proposed method	1.15

inpainting method has the highest computation complexity. The proposed algorithm has to employ extra operations to get the occlusion map, and hence demands more run-time. Comparing with the other three methods, the computing time complexity of the treated way is moderate.

5. CONCLUSIONS

In this paper, an efficient view synthesis scheme is developed to disocclude. The virtual view image is interpolated with the weights, which are calculated by the occlusion map based on the depth reliability maps. Therefore, the holes in the intermediate image can be filled in efficiently. Experiment results confirm the efficiency of the presented methods with only minor increase of the computing time complexity. In the future, more efficient disparity estimation is a very important research topic.

6. ACKNOWLEDGEMENT

This work was supported in part by the Program 973 No. 2012CB316400, key project of National High-Tech Program “863” of P.R.China under Grant No. 2009AA011709, and NSFC No.61175010 and 60805044.

7. REFERENCES

[1] J. Peason, P.L. Dragotti, and M. Brookes, “Accurate non-iterative depth layer extraction algorithm for image based rendering,” *ICASSP 2011*, Prague, Czech, pp. 901-904, May. 2011.

[2] L. Tran, C. Pal, and T. Nguyen, “View synthesis based on conditional random fields and graph cuts,” *ICIP 2010*, Hong Kong, pp. 433-436, Sept. 2010.

[3] J. Shade, S. Gortler, L. He, and R. Szeliski, “Layered depth map,” *SIGGRAPH 1998*, Orlando, USA, pp. 231–242, July. 1998.

[4] Z.W. Liu, P. An, S.X. Liu, and Z.Y. Zhang, “Arbitrary view generation based on DIBR,” *ISISPCS07*, Xiamen, China, pp. 168-171, Nov. 2007.

[5] L. Zhang, and W.J. Tam, “Stereoscopic image generation based on depth images for 3DTV,” *IEEE Trans. on Broadcasting*, vol. 51, no. 2, pp. 191-199, Jun. 2005.

[6] A .Criminisi, P. Perez, and K. Toyama, “Region filling and object removal by exemplar-based image inpainting,” *IEEE Trans. on Image Processing*, vol.13, no.9, pp. 1200-1212, Sept. 2004.

[7] J.H. Park, and H.W. Park, “A Mesh-Based Disparity Representation Method for View Interpolation and Stereo Image Compression,” *IEEE Trans. on Image Processing*, vol.15, no.7, pp. 1751-1762, July. 2006.

[8] 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.-Jun. 2002.

[9] D. Scharstein, and R. Szeliski, “High-accuracy stereo depth maps using structured light,” *CVPR 2003,* Madison, WI, pp. 195-202, Jun. 2003.