Depth-aided Exemplar-based Hole Filling for DIBR View Synthesis

Xuyuan Xu[1], Lai-Man Po[1], Chun-Ho Cheung[2] , Litong Feng[1], Ka-Ho Ng[1], Kwok-Wai Cheung[3]

1. Department of Electronic Engineering, City University of Hong Kong, Hong Kong SAR, China
2. Department of Information Systems, City University of Hong Kong, Hong Kong SAR, China
3. Department of Computer Science, Chu Hai College of Higher Education, Hong Kong SAR, China

ABSTRACT

Quality of synthesized view by Depth-Image-Based Rendering (DIBR) highly depends on hole filling, especially for synthesized view where large disocclusion exists. Thus, many hole filling methods are proposed to improve the synthesized view quality. Inpainting is the most popular approach to recover the disocclusion regions. However, the conventional inpainting either makes the disoccluded regions blurred via diffusion or propagates the foreground information to the disoclusion regions. Annoying artifacts are created in the view synthesis process. This paper proposes a depth-aided exemplar-based inpainting method for large disoclusion recovery. It consists of two processes, warped depth map hole filling and warped color image hole filling. Since depth map can be considered as a grey-scale image without texture, it is much easier to be filled. Disoccluded regions of color image are predicted based on its associated depth map information. Regions with texture lying around the background have higher priority to be filled than other regions and disoccluded regions are filled by propagating the background textures through the exemplar-based inpainting. Thus artifacts created by diffusion or using foreground information for prediction can be eliminated. Experimental results show texture can be recovered in large hole regions and the proposed method has better visual quality compared to existing methods.

Index Terms: DIBR, view synthesis, hole filling, inpainting

1. INTRODUCTION

Recent 3D technology has significant advancement. Advance display systems, such as autostereosopic display system and free-view point TV (FTV) [1,2], are under research and try to provide more realistic perception over the traditional stereoscopic systems. The efficient way for representation of 3D video is one of main problems. The high potential ways to represent the 3D format is video plus depth or multi-video plus multi-depth. It has high flexibility to render virtual views in specified position through DIBR, which is suitable for the application of advance display systems.

DIBR can be used to generate virtual views by warping from the reference views based on depth information. However, DIBR has intrinsic limitation with disocclusion arising in rendering process. It is due to the fact that areas covered by objects in the reference views may appear in the synthesized views. Newly exposed areas (termed 'disocclusions' or 'holes') do not have corresponding information about the reference view, their texture and depth attributes are uncertain. These holes, however, should be filled properly. Otherwise artifacts will appear on the disoccluded regions, especially in large disoccluded regions. It will make the user uncomfortable while viewing the rendered 3D scene. The artifacts can be classified as corona artifacts or hole filling artifacts. Corona artifacts are caused by misalignment of color and depth edges [3,10]. After warping, foreground color textures are treated as background information and are used to predict the disoccluded regions. The hole filling artifacts are caused by incorrect prediction in hole filling.

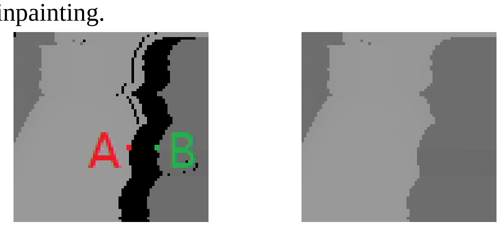
Many hole filling methods were proposed to solve the problem of hole filling artifacts. They fill holes based on the neighbor pixels by linear interpolation [4], extrapolation [4,5] or inpainting [6-9]. Interpolation and extrapolation are good when the hole is small. For large disocclusion, visual quality is degraded and texture of the background cannot be recovered. The inpainting process achieves better performance than interpolation and extrapolation, but results in much higher complexity. The inpainting method can be classified by two categories: structure-based inpainting [6] and exemplar-based inpainting [7- 9]. Structure-based inpainting fills the holes by propagating structures into the target regions via diffusion. The diffusion process makes the hole regions blurred and, thus, textures are lost. It is not, however, suitable for the view synthesis when the hole is large. Exemplar-based inpainting [7] propagates not only the structures but also the texture information into the target regions. It also avoids the blurring effects in the hole regions.

Most of the inpainting methods for the disocclusion filling are structure-based inpainting and focus on small disocclusion recovery. They use two views color images with corresponding depth map to generate the virtual view in between. Size of hole can be dramatically reduced and it is much easier to recover common disoccluded regions. But to generate the virtual view that is not lying in between two existing views, annoying artifacts exist in the large hole regions because of the blurring effect. To tackle this blurring effect, exemplar-based inpaintings using depth information are proposed by Luo [8] and Daribo [9]. However, background textures could not be propagated to the hole regions since the hole regions lying in the foreground sometimes is filled before hole regions lying in the background and foreground pixels may be used for prediction. Annoying artifacts are created in synthesized views. In hole filling process, background information is preferred to fill the disoccluded regions. A depth-aided exemplar-based inpainting is proposed in this paper and aims at recovering both background structures and texture information in large disoccluded regions by considering the depth information.

This paper is organized as follows: an overview of depth-aided exemplar-based hole filling is introduced in Section 2. The details of warped depth map filling and warped color image filling are described in section 3 and 4 respectively. Experimental results are shown in section 5 and finally a conclusion is drawn in section 6.

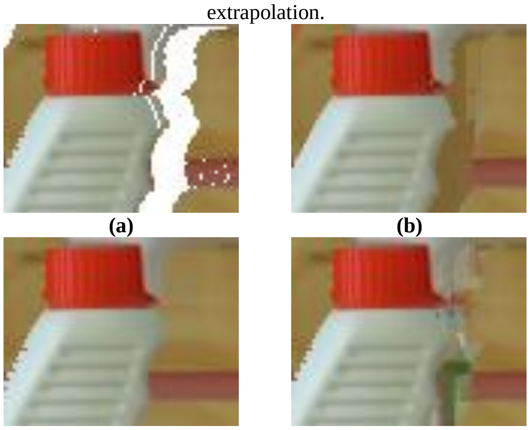
2. OVERVIEW

Disoccluded regions are usually the background regions and need to be filled properly to avoid the annoying artifacts. To recover the structures and texture information using background information are the two purposes of the hole filling. Depth-aided exemplar-based hole filling contains two procedures: (1) warped depth map hole filling, and (2) warped color image hole filling. Before hole filling, color and depth images are warped together. Then disoccluded regions in the warped depth map are filled first using background information. Since depth map is gray-scale non-texture plane, it is much easier to fill the disoccluded regions. After the warped depth map is filled, color image is filled by depth-aid exemplar-based inpainting.



(a) **(b)**

Fig. 1: (a) Warped depth map, (b) Filled depth map using depth based extrapolation.



(a) **(b)** **(c)** **(d)**

Fig. 2: (a) Warped image, (b) Holes filled by sequence order, (c) Holes filled by proposed order, and (d) Exemplar-based inpainting.

3. WARPED DEPTH MAP FILLING

In the warping process, positions of holes appear between the foreground and background objects. One example is shown in Fig. 1. Depth based extrapolation is used to fill the disoccluded regions. The disoccluded regions are created by relative disparity between foreground and background objects, like pixels A and B in Fig. 1(a). Pixel A is the neighboring pixel to Pixel B before warping process. So the disoccluded regions between A and B have higher probability to be filled using the background information that is the minimum value of A and B.

The depth based extrapolations are respectively described by Equation (1) and (2):

$$D[n] = \min(D[m_l], D[m_r]) \,, \tag{1}$$

$$d(W^{-1}(m_l), W^{-1}(m_r)) = 1 \,, \tag{2}$$

where D is the warped depth image and W^{-1} describes the inverse 3D warping function; m_l and m_r are the positions of foreground and background pixels lying in the contour of hole; $d(a, b)$ denotes the Euclidean distance, and n is the pixel location from m_l to m_r. Fig. 1(b) shows the filled depth map using the depth based extrapolation.

4. WARPED COLOR IMAGE FILLING

The color image filling has two procedures: filling order estimation and searching best matched patch. Filling order estimation is used to decide the order of patches to be filled. The best matched patch searching is to find the most similar patch to describe the patch to be filled. Filling order is updated after each patch is filled. The highest priority patch is selected and is filled by the best matched patch. This process repeats until the complete hole regions are filled.

4.1 FILLING ORDER ESTIMATION

Filling order is crucial to non-parametric texture synthesis [7], especially in view synthesis process. The background texture in hole regions cannot be recovered if the filling order is not well organized, like in Fig. 2(b). It fills the hole using sequential order, from left to right, top to bottom, and proposed similarity function in section 4.2. Although through the proposed similarity function background pixels are enforced to predict the hole to reduce artifacts, the texture cannot be recovered yet.

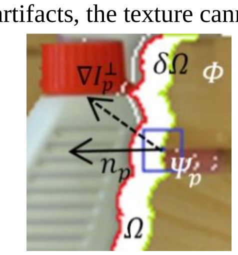


Fig. 3: Notation: Center of patch $\boldsymbol{\Psi_p}$ lies in the contour $\boldsymbol{\delta\Omega}$. $\boldsymbol{n_p}$ is the normal to $\boldsymbol{\delta\Omega}$ and $\boldsymbol{\nabla I_p^{\perp}}$ is the isophote at center of $\boldsymbol{\Psi_p}$. $\boldsymbol{\Phi}$ and $\boldsymbol{\Omega}$ denotes the non-hole regions and the hole regions, respectively. The green line and red line of $\boldsymbol{\delta\Omega}$ are the background and foreground pixels, respectively.

In contours of hole regions, background patches with texture and less hole regions should have the higher priority to be filled. If non-texture background patches or foreground patches are selected before texture background patches, texture in hole cannot be recovered like Fig. 2(b). The notation diagram of filling order is shown in Fig. 3. It is judged by four factors: 1) whether patch lies in the background region. 2) continuation of strong edge inside the patch, 3) confidence of patch which is described by the percentage of non-hole pixels inside the patch and 4) Consistent depth information of hole and non-hole of the patch. The priority function is defined as:

$$\mathrm{P(p)} = F(p) \cdot \frac{|\nabla I_p^{\perp} \cdot n_p|}{\alpha} \cdot \frac{N_{nh}}{N_{total}} \cdot \frac{\sum_{r\in\Omega\cap\Psi_p} \sum_{q\in\Phi\cap\Psi_p} e^{-\frac{d(D[r],D[q])^2}{2\sigma_d}}}{N_{nh}(N_{total}-N_{nh})} \,, \tag{3}$$

$$\text{where } F(p) = \begin{cases} 0, p \in B \\ 1, p \in F \end{cases} , \tag{4}$$

N_{nh} is the number of non-hole pixels, and N_{total} is the total number of pixels inside the patch; α is the normalized factors and σ_d is parameter used to adjust the slope of exponential function; B and F denote the sets of foreground and background pixels around the disoccluded regions.

4.2 SEARCHING BEST MATCHED PATCH

To search the best matched patch for the color image hole filling, a depth-aided similarity function C_d is used to measure the similarity. The functions are illustrated below:

$$C_d(\Psi_{p1},\Psi_{p2}) = M_{nh}(\Psi_{p1},\Psi_{p2}) \cdot M_h(\Psi_{p1},\Psi_{p2}), \quad (5)$$

$$M_{nh}(\Psi_{p1},\Psi_{p2}) = \frac{1}{N_{nh}} \sum_{r\in\Psi_{p1}} \sum_{q\in\Psi_{p2}} e^{-\frac{d(D[r],D[q])^2}{2\sigma_d}} \cdot e^{-\frac{d(C[r],C[q])^2}{2\sigma_c}} \cdot H(r) \cdot H(q), \quad (6)$$

$$M_h(\Psi_{p1},\Psi_{p2}) = \frac{1}{N_{total}-N_{nh}} \sum_{r\in\Psi_{p1}} \sum_{q\in\Psi_{p2}} e^{-\frac{d(D[r],D[q])^2}{2\sigma_d}} \cdot \tilde{H}(r) \cdot \tilde{H}(q), \quad (7)$$

$$H(x) = \begin{cases} 1, x \in \Phi \\ 0, x \in \Omega \end{cases}, \quad (8)$$

$$\tilde{H}(x) = 1 - H(x), \quad (9)$$

where D is the filled warped depth and C is the warped color image. σ_d and σ_c are two thresholds used to adjust the slope of non-linear exponential function. The non-linear exponential function (Gaussian model) is used to measure the similarity, which results in a value of the range from zero to one, i.e. it normalizes the color and depth cost. Patch with smaller differences in both color and depth image is more preferred for prediction, which can be described by non-linear exponentiation functions of depth and color information. They have the characteristic that small difference results in much higher similarity value and large difference lead to lower similarity value. The similarity term (C_d) includes two parts, cost of non-hole region (M_{nh}) and hole region (M_h). The non-hole region is measured by both color and depth information while the hole region is only measured by the depth information due to the lack of color information. By considering the depth information, the best matched patch has a very similar property as the patch to be filled, both in the hole region and non-hole region with texture and depth information. The search is performed inside the searching windows $\boldsymbol{W}$ ($\pm$32 pixels) to find the best matched patch with highest similarity cost that can be used to predict the hole regions.

5. EXPERIMENTAL RESULTS

The Middlebury datasets [10] with ground truth depth map are used for experiments. View 1 is used to synthesize view 3 to demonstrate the ability to recover large hole regions. Before the view synthesis, a foreground biased depth map refinement [11] is used to correct the misalignment of depth map and color image. After warping, fast inpainting [4] of structure-based inpainting, exemplar-based inpainting [5], depth-aided image inpainting [9] and the proposed depth-aid exemplar-based inpainting are used to fill the holes and results are shown in Fig. 4, Fig. 5 and Fig. 6. Patch size is 9×9, σ_d and σ_c are set to 50 and 500, respectively.

The fast inpainting of structure-based inpainting approach can recover the structure of the disoccluded regions, but performance degrades when hole size goes larger, as shown in Fig. 4(d), Fig. 5(d) and Fig. 6(d). The diffusion causes the blurring effect inside the hole region and results in annoying artifacts.

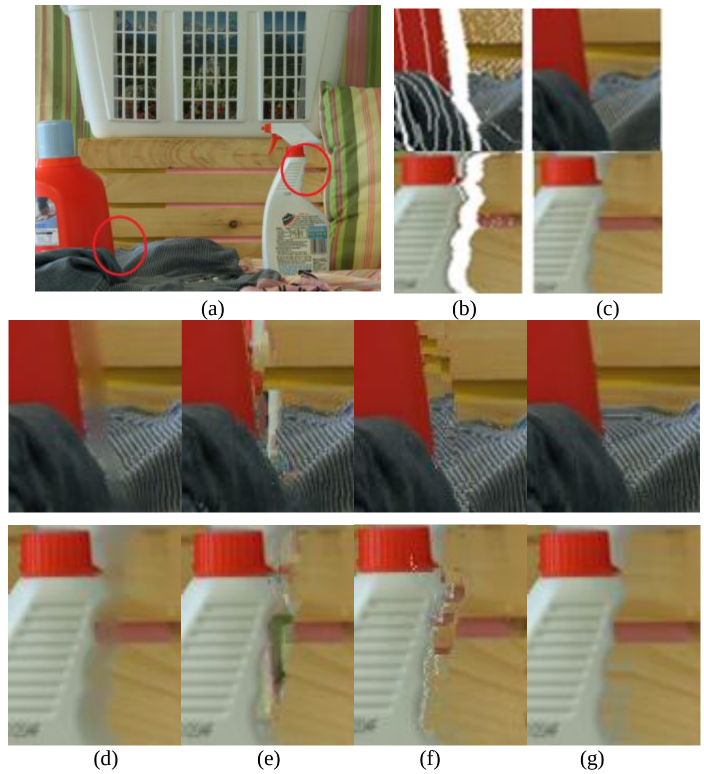


Fig. 4 Results of Laundry: (a) Input view 1 image, (b) Warped hole image, (c) Original view 3, (d) inpainting, (e) exemplar-based inpainting, (f) depth-aided image inpainting, and (g) Proposed method.

Fig. 5 Results of Moebius: (a) Input view 1 image, (b) Warped hole image, (c) Original view 3, (d) inpainting, (e) exemplar-based inpainting, (f) depth-aided image inpainting, and (g) Proposed method.

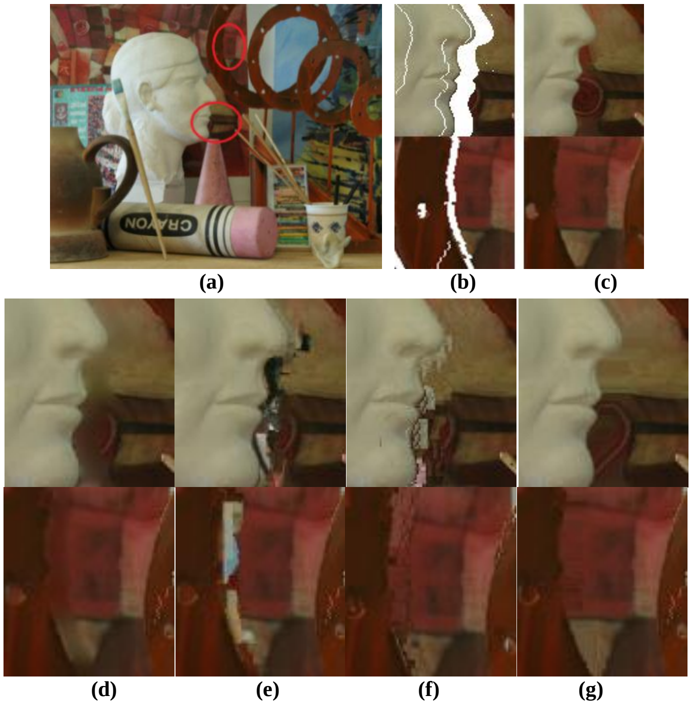


Fig. 6 Results of Art: (a) Input view 1 image, (b) Warped hole image, (c) Original view 3, (d) inpainting, (e) exemplar-based inpainting, (f) depth-aided image inpainting, and (g) Proposed method.

The exemplar-based inpainting can recover the clear background structure and texture. However, it is not designed for the disoccluded region recovery of the 3D view synthesis. It targets to remove foreground objects and tries to reconstruct the removal region using background region. Depth information is not considered in both filling order selection and the search for the best matched patch. SSD is used to measure the similarity. Patch with lower similarity could have lower SSD cost, and results are shown in of Fig. 4(e), Fig. 5(e) and Fig. 6(e). In addition, it searches all the patches inside the image to get the best matched patch, which increases the computational complexity.

Although depth consistent is considered in filling order of depth-aided image inpainting, patch lying in the foreground of the disoccluded regions can be filled before patch lying in the foreground. Thus texture in the disoclusion cannot be recovered. Since depth information is used in hole filling, foreground information could be reduced to propagate to the disoccluded regon as Fig. 4(f), Fig. 5(f) and Fig. 6(f).

In the proposed method, the depth map is filled before the color image. After the depth map is filled, disoccluded regions of color image are filled based on the filled warped depth image and their surrounding background texture information. The modified filling order estimation has higher priority to select the patch lying in the background boundary with consistent depth information. Since non-linear exponential functions are used for searching the best matched patch, prediction of disoccluded regions using the foreground pixels can be avoided compared with depth-aided image inpainting, which uses the SSD as cost function. The proposed method has the ability to recover both the structure and texture of the disoccluded regions without blurring effect or using the foreground information, as shown in Fig. 4(g), Fig. 5(g) and Fig. 6(g).

6. CONCLUSION

In this paper, a depth-aided exemplar-based inpainting is proposed to recover the structure and texture of the large disoccluded regions. It warped both depth map and color image. The warped depth map is easier to be filled and depth-aid extrapolation is used to fill it. After the depth map is filled, the color image is filled by depth-aided exemplar-based inpainting. Filling order is estimated according to whether patch lies in the background, continuation of strong edge, confidence of patch and consistent depth information inside the patch. Non-linear similarity functions with depth consideration are used to find out the best matched patch for hole filling. Experimental results show that the proposed method achieves better results without blurring effect and has the ability to propagate the background structure and texture to the disoccluded regions as compared with other hole filling methods.

7. REFERENCE

[1] T. Fujii and M. Tanimoto, "Free-viewpoint TV system based on ray-space representation," in Proc. SPIE ITCom, vol. 4864-22, pp. 175-189, Nov. 2002.

[2] L. Zhang and W.J. Tam, "Stereoscopic Image Generation Based on Depth Images for 3D TV," IEEE Trans. broadcast., vol. 51, no. 2, pp. 191-199, Jun. 2005.

[3] K. Müller, P. Merkle, T. Wiegand, "3-D Video Representation Using Depth Maps", IEEE Proceedings, vol.99, no.4, pp.643-656, Apr. 2011.

[4] C. Vázquez, W. J. Tam and F. Speranza, "Stereoscopic Imageing: Filling Disoccluded Areas in Depth Image-Based Rendering", in Proc. SPIE, Vol. 6392, 2006.

[5] L.M. Po, S. Zhang, X. Xu, Y. Zhu, "A new multidirectional extrapolation hole-filling method for depth-image-based rendering", Inter. Conf. Image Processing (ICIP), pp. 2637-2640, Sep. 2011.

[6] M. M. Oliveira, B. Bowen, R. McKenna, and Y.S. Chang , "Fast Digital Image Inpainting", Inter. Conf. Visualization, Image and Image Processing (ICVIIP), pp. 261-266, 2001.

[7] A. Criminisi, P. Pérez, and K. Toyama, "Object Removal by Exemplar-Based Inpainting", IEEE Trans. Image Processing, vol. 13, No.9, pp. 1200-1212, Sep. 2004.

[8] K. Luo, D. Li, Y. Feng, M. Zhang, "Depth-aided inpainting for disocclusion restoration of multi-view images using depth-image-based rendering", Journal of Zhejiang University SCIENCE A, vol 10, pp 1738-1749, Dec. 2009.

[9] I. Daribo, B. Pesquet-Popescu, "Depth-aided image inpainting for Novel View Synthesis", Inter. Conf. Multimedia Signal Processing (MMSP), pp. 167-170, Dec. 2010.

[10] D. Scharstein and C. Pal, "Learning conditional random fields for stereo", Inter. Conf. Computer Vision and Pattern Recognition (CVPR), pp.1-8, Jun. 2007.

[11] X. Xu, L.M. Po, K.W. Cheung, K.H. Ng, K.M. Wong and C. W. Ting, "A Foreground Biased Depth Map Refinement Method For DIBR View Synthesis", Inter. Conf. Acoustics, Speech, and Signal Processing (ICASSP), pp. 805-808, Mar. 2012.