BI-LAYER INPAINTING FOR NOVEL VIEW SYNTHESIS

Hwasup Lim, Yong Sun Kim, Seungkyu Lee, Ouk Choi, James D. K. Kim, and Changyeong Kim

Samsung Advanced Institute of Technology, Republic of Korea

ABSTRACT

We present a bi-layer inpainting method for synthesizing novel views from a single color image and its corresponding depth map under the exemplar-based inpainting framework. Unlike conventional image inpainting, the decisions of the inpainting direction and the sample regions are important to inpaint disoccluded regions which disclose hidden background regions in the new viewpoint. The proposed algorithm first labels boundaries along the disoccluded regions whether it belongs to the foreground or background objects, and then separates their surrounding regions into the foreground and background regions using the graph cut algorithm. The disoccluded regions are then filled from the background boundary with best-match patches taken from the background regions. As demonstrated in the experimental results, the proposed method recovers the disoccluded region with visually plausible quality.

Index Terms— Novel view synthesis, bi-layer segmentation, image inpainting, and depth image-based rendering

1. INTRODUCTION

Recently depth image-based view synthesis gains an increasing interest due to easy acquisition of depth images and needs of multi-view generation from limited view images. The most important issue raised in view synthesis is disclosure of occluded regions while viewpoint change.

Early work in view synthesis using depth images apply a Gaussian filter [1][2] to the depth image before 3D warping and then fill holes by averaging neighborhood pixels. These methods minimize possible disoccluded regions by alleviating the depth discontinuity. Excessive depth smoothing, however, produces distorted views along the boundaries with large depth discontinuity.

Filling large disoccluded regions seems to be solved by adopting the recent exemplar-based inpainting method [3]. This method cannot be directly applicable because the foreground boundary needs to be preserved during the inpainting process. Assumed that the boundary pixels with low depth on the horizontally opposite side of the disoccluded region belong to the background boundary, these pixels are copied next to the foreground boundary in the other side to prevent foreground expansion [4]. This approach also produces irregular foreground expansion in the disclosed regions enclosed by foreground regions in the both sides. Recently a pixel labeling method [5] is introduced to select candidate pixels to fill disoccluded region using depth, edge, and segmentation information with a discriminative probabilistic model. This method still provides no explicit distinction on foreground and background regions and has limitation on texture generation due to per-pixel synthesis.

This paper presents a novel algorithm for inpainting disoccluded regions under the exemplar-based inpainting framework [3] by separating the source region, the region to be sampled for filling the disoccluded region, into the foreground and background regions. Unlike conventional inpainting problems in [6][3] where all the remaining regions are used as source regions, the segmentation of the background and foreground region is important because the foreground region should be preserved during the inpainting process and the background region is mainly used for filling the disoccluded region.

To solve this problem, we first classify the pixels at occlusion boundary enclosing the disoccluded region into either the foreground or background boundary using the edge characteristics in the depth map, and then separates their surrounding regions into the foreground and background regions using the graph-cut optimization. The disoccluded region is then filled from the background boundary with best-match patches from the background region. The proposed framework is described in Fig. 1.

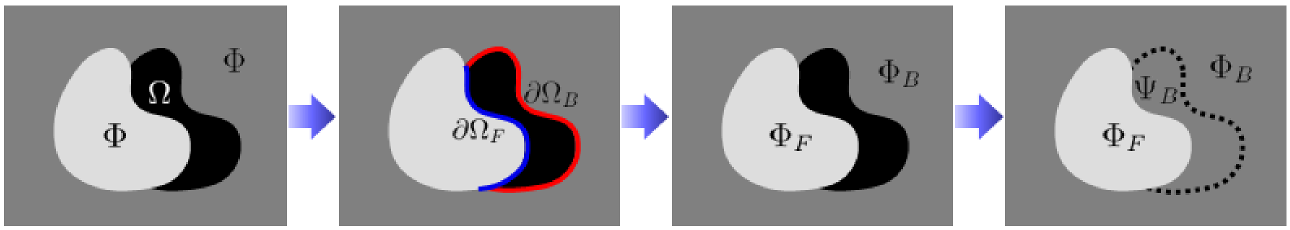
The paper is organized as follows. Section 2 presents the method of occlusion boundary labeling. Section 3 describes the method of bi-layer segmentation from occlusion boundary label. Section 4 explains bi-layer inpainting. Section 5 provides experimental results. Finally, Section 6 concludes this paper and presents future work.

2. OCCLUSION BOUNDARY LABELING

Given a pair of a color image and its corresponding depth map, a novel view can be synthesized by warping the color image using the depth map. An occlusion boundary is defined as non-occluded boundary pixels that circumscribe a disoccluded region.

Knowing the label of the occlusion boundary in the warped image, whether the foreground or background bound-

978-1-4577-1302-6/11/$26.00 ©2011 IEEE

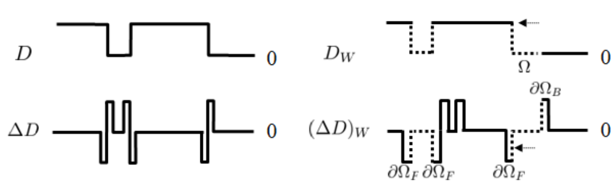


(a) (b) (c) (d)

Fig. 1. The proposed bi-layer inpainting procedure for novel view synthesis. (a) Image warping where Ω denotes the disoccluded region to be filled and Φ denotes non-disoccluded region to be sampled. (b) Occlusion boundary labeling where $\partial\Omega_F$ and $\partial\Omega_B$ represent the foreground and background boundary. (c) Bi-layer segmentation where Φ_F and Φ_B represent the foreground and background region. (d) Bi-layer exemplar-based inpainting where Ψ_B represents the filled region using the background region Φ_B.

ary, allows for certain strategies for filling the disoccluded region. In general the foreground boundary needs to be preserved under assumption of cardboard type foreground objects with no volume behind such that the disoccluded region should be filled from the background boundary with the background region. In real situation, the foreground boundary may need to be expanded according to rear volumetric information. We follow the cardboard assumption for the simplicity.

To label occlusion boundaries, previous methods consider relative depth difference of boundary pixels on the horizontally opposite side [4] or on the eight directions [7]. For example, the boundary pixels having the lowest depth is labeled as the background boundary. However, these assumptions are invalid when both sides of boundaries are the foreground boundary in real world and when unfortunately the lowest depth value is not present in the eight direction. The occlusion boundary label can also be determined using edge detection in the original depth map [8], associated with the gradient vector. Not all the detected edges become occlusion boundary since they depend on the user-given thresholds.



(a) (b)

Fig. 2. Occlusion boundary labeling. (a) Original depth map and its Laplacian map. (b) Warped depth map and warped Laplacian map for right view synthesis.

Based on the observation that the occlusion boundary is generated from splitting the depth discontinuity into the lower and higher side, the foreground boundary in the original depth map is the falling edge and the background boundary is the rising edge as shown in Fig. 2. It can be efficiently classified by the sign of the Laplacian of the depth map.

Let D be the input depth image and ΔD the Laplacian of the input depth image. D_W denotes the warped depth map using the input depth map D such that $D_W(p_x, p_y) = D(p_x - D_W(p_x, p_y), p_y)$ where p_x and p_y are the coordinates of a pixel p. $(\Delta D)_W$ denotes the warped Laplacian of the input depth map using D such that $(\Delta D)_W(p_x, p_y) = \Delta D(p_x - D_W(p_x, p_y), p_y)$. Let Ω the disoccluded region in the warped depth map D_W and $\partial\Omega$ the occlusion boundary.

The pixel, along the occlusion boundary, with a negative Laplacian value belongs to the foreground boundary and the positive belongs to the background boundary. The label function of pixel p is given as follows.

$$l(p) = \begin{cases} \partial\Omega_F & \text{if } (\Delta D)_W(p) < 0 \\ \partial\Omega_B & \text{if } (\Delta D)_W(p) > 0 \end{cases}, \quad \text{for } p \in \partial\Omega \tag{1}$$

where $\partial\Omega_F$ and $\partial\Omega_B$ denote the foreground and background boundary respectively. Note that zero Laplacian value represents no depth discontinuity and will not appear along the occlusion boundary. One can use the Laplacian of Gaussian to increase its support region and robustness against noises. Once the background boundary is known, a PDE-based image inpainting technique [9] can be utilized considering the background boundary as propagating front. It is well known that such an approach is appropriate only for filling small hole and suffers from blurring artifacts in large hole regions.

(a) (b)

Fig. 3. Occlusion boundary labeling result. (a) Original color image. (b) Warped color image with occlusion boundary label (blue = foreground boundary, red = background boundary).

Fig. 3 shows simple results of occlusion boundary labeling where the boundary pixels are accurately classified in the presence of complex structure. An additional benefit from the occlusion boundary labeling is that the ghost artifact common in depth-image based view synthesis, remaining foreground pixels on the background boundary due to depth inaccuracy, can be effectively removed by eroding the background boundary while preserving the foreground boundary.

3. FOREGROUND AND BACKGROUND SEGMENTATION

After labeling occlusion boundaries, the probability of the pixel that belongs to either the foreground or background region can be obtained from the depth distribution of pixels that belongs the foreground or background boundary. Given depth value of pixel p, we find the probability density function of each label using kernel density estimation as follows.

$$\begin{aligned} f_F(p) &= \frac{1}{n_F h} \sum_{p_i \in \partial\Omega_F}^{n_F} K\left(\frac{D_W(p) - D_W(p_i)}{h}\right) \\ f_B(p) &= \frac{1}{n_B h} \sum_{p_i \in \partial\Omega_B}^{n_B} K\left(\frac{D_W(p) - D_W(p_i)}{h}\right) \end{aligned} \tag{2}$$

where n_F and n_B denote the numbers of pixels belong to $\partial\Omega_F$ and $\partial\Omega_B$ respectively. K denotes a kernel function with a kernel width parameter h.

Using the probability of each label, we can build a Markov random field model under the assumption that the pixel belongs to either one of the two labels with the corresponding probabilities and the labels of adjacent pixels are prone to be same. Let Φ be the non-disoccluded region in D_W. Φ is to be separated into Φ_F and Φ_B, the foreground and background region respectively. The energy function is defined as

$$\begin{aligned} E(l) &= E_d(l) + E_s(l) \\ &= \sum_{p \in \Phi} U_p(l_p) + \lambda \sum_{\{p,q\} \in N} V_{p,q}(l_p, l_q) \end{aligned} \tag{3}$$

where $l_p \in \{\Phi_F, \Phi_B\}$ is the label of the pixel p in Φ and N represents the neighborhood system. The data cost $U_p(l_p)$ is computed from Eq. (2) and the smoothness cost $V_{p,q}(l_p, l_q)$ is computed from the label difference between neighborhood pixels. These costs are represented as follows:

$$U_p(l_p) = \begin{cases} 1 - \frac{f_F(p)}{f_F(p)+f_B(p)} & \text{if } l_p = \Phi_F \\ 1 - \frac{f_B(p)}{f_F(p)+f_B(p)} & \text{if } l_p = \Phi_B \end{cases} \tag{4}$$

$$V(l_p, l_q) = \begin{cases} 0 & \text{if } l_p = l_q \\ 1 & \text{if } l_p \neq l_q \end{cases} \tag{5}$$

The above MRF energy can be minimized via graph cuts [10]. For accuracy as well as efficiency, instead of using the whole image, Φ can be redefined as the bounded region that encompasses a single disoccluded region with a margin, which covers the search range used in exemplar-based inpainting. This tweak makes sense in that separated two disoccluded regions might produce different region labels for the same region.

Fig. 4 demonstrates the proposed bi-layer segmentation results where clear distinction between the foreground and background region are accomplished in both the depth map and color image.

(a) (b) (c)

Fig. 4. Foreground and background segmentation. (a) Warped color image (b) Warped depth map. (c) Segmented regions (blue = foreground, red = background).

4. EXEMPLAR-BASED OCCLUSION INPAINTING

Once the background region, the source region, is separated, the filling process in exemplar-inpainting method [3] is performed from the background boundary by finding best-match patches in the background region Φ_B. We adopt their patch priority computation for deciding the filling order around the occlusion boundary as described in Eq. (6). The difference compared with the original method is that the initial priority in the proposed method is set to high for the background boundary and zero for the foreground boundary to start from the background boundary. The modified priority equation is given as

$$P(p) = C(p)H(p) \tag{6}$$

$$C(p) = \frac{\sum_{q \in \Psi_{B_p} \bigcap \Phi_B} C(q)}{|\Psi_{B_p}|}, \quad H(p) = \frac{|\nabla I_{\mathbf{p}}^{\perp} \cdot n_p|}{\alpha}$$

where p denotes the patch center and $C(p)$ denotes the confidence term and $H(p)$ denotes the data term. Refer to [3] for details. For the initialization, the confidence term $C(\mathbf{p})$ is set to $C(p) = 0\ \forall p \in \Omega \bigcup \Phi_F$ and $C(p) = 1\ \forall p \in \Phi_B$. Ψ_p is set to $\Psi_p = 0\ \forall p \in \Phi_F$ not to include the foreground region.

We use weighted sum of squared difference with the weight function $w(p) = k|\nabla I|$ to find patches that provide better similarity along edge continuation. Poisson image editing technique [11] is then used to alleviate seam artifact due to patch copying. The filling process ends when no pixel belongs to Ω.

5. EXPERIMENTAL RESULTS

The proposed method was tested on Middlebury data set [12], which provides 7 views indexed 0 to 6 and where views 1 and

Fig. 5. Experimental results. The first row is the results by [4]. The second row is our results. The third row is the original image of view 3

Year	Previous method[4]	Proposed method
2003 (2 sets)	28.86	29.14
2005 (6 sets)	29.28	30.91
2006 (21 sets)	31.98	33.96

Table 1. Visual quality comparison (Ave. PSNR(dB))

5 have their depth maps as well. We used the color image and depth map in view 1 to synthesize view 3. The search range was set to 100 and the patch size was set to 15. The 3x3 Laplacian operator was used to label the occlusion boundary and the Gaussian kernel with $h = 10$ was used to estimate the probability distribution in Eq. (2). The background boundary with 2 pixel distance was removed to prevent the ghost artifact.

As shown in Fig.5, the proposed method qualitatively outperforms the existing method, especially in the regions where complex overlapping of the foreground and background objects occurs. The visual quality drastically increases because no foreground texture is introduced into the inpainted regions. Table 1 shows the comparison with the existing method [4]. To our best knowledge, no other quantitative result was reported on novel view synthesis using the same data set.

The proposed method can be easily extendable to the intermediate view synthesis. The occlusion boundary labeling can be performed referring each depth map and thereafter the remaining processes can be completed without loss of generality. The amount of the disoccluded region decreases due to the projection of other view images.

6. CONCLUSION

The paper emphasizes the selection of the source region in the exemplar-based inpainting framework for depth-based view synthesis because the continuity from the background pattern should prolong from the background region while preserving the foreground region. We proposed a novel way to segment the source region using the occlusion boundary label under the Markov random field framework. As demonstrated in the several experimental results, the proposed bi-layer inpainting method produces more plausible results. Further work will include more sophisticated strategies for handling the occlusion boundary label to prevent cardboard effect and multi-layer inpainting where a bi-layer cannot represent complicated structure due to the overlapping by multiple foreground and background regions.

7. REFERENCES

[1] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV," in *Proc. of SPIE*, 2004, vol. 5291, pp. 93–104.

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

[3] A. Criminisi, P. Perez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Trans. on Image Proc.*, vol. 13, no. 9, pp. 1200–1212, 2004.

[4] K.-J. Oh, S. Yea, and Y.-S. Ho, "Hole filling method using depth based in-painting for view synthesis in free viewpoint television and 3-d video," in *Proc. of PCS*, 2009, pp. 233–236.

[5] L. C. Tran, C. J. Pal, and T. Q. Nguyen, "View synthesis based on conditional random fields and graph cuts.," in *Proc. of ICIP*, 2010, pp. 433–436.

[6] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proc. of SIGGRAPH*, 2000, pp. 417–424.

[7] S. Zinger, L. Do, and P. H. N. de With, "Free-viewpoint depth image based rendering," *Journal of VCIR*, pp. 533–541, 2010.

[8] K. Muller, A. Smolic, K. Dix, P. Kauff, and T. Wiegand, "Reliability-based generation and view synthesis in layered depth video," in *IEEE Workshop on MMSP*, 2008, pp. 34 – 39.

[9] M. M. Oliveira, B. Bowen, R. Mckenna, and Y. s. Chang, "Fast digital image inpainting," in *Proc. of VIIP*, 2001, pp. 261–266.

[10] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. on PAMI*, vol. 23, pp. 1222–1239, 2001.

[11] P. P.érez, M. Gangnet, and A. Blake, "Poisson image editing," *ACM Trans. of Graphics*, pp. 313–318, 2003.

[12] Middlebury, "http:// vision.middlebury.edu/ stereo," .