HOLE FILLING WITH RANDOM WALKS USING OCCLUSION CONSTRAINTS IN VIEW SYNTHESIS

Sunghwan Choi, Bumsub Ham, Kwanghoon Sohn

Dept. of Electrical and Electronic Eng., Yonsei University, Seoul, Korea
khsohn@yonsei.ac.kr

ABSTRACT

In this paper, we propose a hole filling technique which coherently reconstructs the hole region during the view synthesis. The holes can be filled successfully in case that the virtual camera locates between real cameras by using interpolation. However, they cannot be handled in case that the virtual camera locates beyond the field of view of the real camera. We address this problem by jointly using image completion technique and random walks. First, occlusion constraint is imposed in order to guide the filling order. It is observed that the holes occur in a similar pattern because of the geometric characteristic of the camera configuration. This observation named vertical prior in this paper is also used to label each pixel on the fill front with foreground or background. Second, the probabilities estimated by random walks are utilized to find the patch candidates and to select the optimal patch. The experimental results show that the proposed method gives visually pleasing results over both interpolation and conventional image completion method.

***Index Terms*—** Hole filling, inpainting, view synthesis, random walks

1. INTRODUCTION

3D technology has been making rapid progress. While most of this remains at the experimental level, it is no longer tomorrow's technology. Some technology such as freeview rendering is widely used in the entertainment industry, e.g., game, movie, and sports broadcasting. Furthermore, 3D TV has reached commercialization. In these days, a user does not want to take on a passive attitude toward getting the information any more. Judging from this paradigm, freeview TV (FTV) meets the needs of a user in that it enables a user to change viewpoint freely as if she or he is there [1]. Therefore, FTV along with 3D TV will be a next generation TV system. It is an active system in contrast to the conventional TV characterized by the word "passivity" which means that a user only sees a fixed view point. However, most existing contents have been produced for the purpose of stereoscopic display. To adopt these contents to auto-stereoscopic or freeview system, it is needed to generate virtual views from given stereo pairs with associated depth information.

Depth image based rendering (DIBR) is one of the key technologies for synthesizing virtual views. The associated disparity map is used to reconstruct corresponding 3D layout from stereo pairs. It is inevitable that regions occluded or cannot be covered by existing real cameras appear as the virtual camera moves. We call these regions as holes in this paper. The points concealed by the foreground object would have been visible after view synthesis. It is important to fill in these hole regions since it affects the quality of the synthesized virtual views.

Many researchers have made an effort to fill the hole regions seamlessly. Interpolation is one of the classic examples [2]. It has low complexity, and handles the hole successfully when the hole is small. It, however, has some drawbacks when the hole becomes large. Some artifacts such as blurring appear in the synthesized view. Inpainting (or image completion) is an alternative tool in order to infer missing information. The information inferred is semantic in contrast to the interpolation. It generally can be classified into two categories: structure-based inpainting and exemplar-based inpainting. Structure-based inpainting infers missing information by propagating the structure information from visible regions to regions to be filled. It is not appropriate to view synthesis since it is governed by diffusion or fluid PDEs [3] [4], which causes the hole to be blurred. Exemplar-based inpainting infers missing information by propagating the structure information as well as texture information, which prevents the hole to be filled from being blurred [5]. It consists of two procedures: filling order estimation and finding optimal patch. It is important to estimate the filling order since it significantly influences the quality of the output image. It is calculated by measuring the strength of isophotes (data term) and the amount of reliable information (confidence term) within a patch followed by finding optimal patch via sum of square difference (SSD). However, the priority in [5] is not appropriate to fill in the hole during view synthesis since it has no ability to handle depth ambiguities, e.g., the intersection of two perpendicular regions with different depth information. It is likely that foreground information might be penetrated into the occluded area unless the depth ambiguities are remedied appropriately. L. Wang et al proposed the hole filling technique so that the stereo pairs can be edited [6]. Although the depth ambiguities were resolved to some extent by using corresponding depth information, only foreground objects can be removed or edited.

In this paper, we propose a new method to fill in the hole which occurs during view synthesis. We address this problem by jointly using image completion technique, i.e., inpainting, and random walks. We present a new filling order scheme based on the occlusion constraint which models the characteristic of the procedure of view synthesis. It is observed that the holes occur in a similar pattern because of the geometric characteristic of the camera configuration. This observation is also used to label each pixel on the fill front with foreground or background. We consider the procedure of patch finding as labeling problems, which can be solved by using many inference algorithms such as belief propagation and random walks [7]. Optimal patch is selected according to the probability estimated by random walks in contrast to the conventional method, which makes the algorithm robust to the size and the shape of the aggregation window.

This paper is organized as follows. Motivation and overview are covered in section 2. The occlusion constraint as well as the probabilistic framework via random walks is explained in section 3. In section 4, we present the experimental results, and compare

(a) (b) (c)

Fig. 1. The visible artifacts caused by depth ambiguities: (a) initial image. (b) Image completion [5]. (c) Image completion [5] with the occlusion constraint.

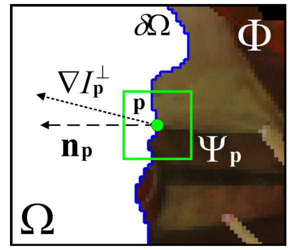


Fig. 2. Notation diagram. Φ denotes the known region of the entire image while Ω indicates the holes. The region in $\Psi_{\mathbf{p}} \cap \Phi$ and $\Psi_{\mathbf{p}} \cap \Omega$ are called source region and target region, respectively. $\perp$ denotes the orthogonal operator. $\mathbf{n}_{\mathbf{p}}$ is a unit vector orthogonal to the front $\delta\Omega$ in the point $\mathbf{p}$. I denotes the entire image.

the proposed method with conventional image completion methods. Finally, we conclude with a brief summary in section 5.

2. MOTIVATION AND OVERVIEW

2.1. Priority formulation for filling order in view synthesis

In inpainting, the strength of isophotes are only considered when the priority of filling order is computed, which causes the foreground texture to be penetrated into the occluded region. Therefore, the quality of the result might be degraded as well as the visual coherence of the result is diminished unless depth ambiguities are appropriately remedied as shown in Fig. 1.

It is observed that the hole occurs in a similar pattern due to the camera configuration, e.g., the hole may exists along the object in the vertical direction in parallel camera configuration. In addition, the distribution of disparity value in the contour along the hole shows a similar pattern. It is hence natural that the contour at the foreground side has large disparity values. We utilize this vertical prior to guide the filling order so that the depth ambiguities are handled successfully as will be detailed in section 3.1.

2.2. Optimal patch finding

The neighborhood region around a newly exposed area is likely to be the candidates for holes since most of holes lie on the continuation of neighborhood region. This characteristic can be modeled as a labeling problem. Given a set of points on the contour of the hole, the points belong to the foreground are labeled as "visible seed" in accordance with the vertical prior while the remaining points are labeled as "invisible seed". This is exactly coincided with random walks theory [7]. The labeling problem can be solved by random walks by which the Dirichlet integral can be minimized. Hence, the probability computed by the weights existing between nodes is assigned to each unseeded (unlabeled) points by serving both foreground points and background points as seeds. For this coincidence in the problem-specific domain, we adopt random walks theory to find patch candidates in the probabilistic framework as will be discussed in section 3.2. Notations are shown in Fig. 2.

3. PROPOSED METHOD

3.1. Occlusion constraint

Hole filling process proceeds through a best-first filling strategy which depends on the priority assigned each patch along the fill front. Higher priority would be assigned if the following conditions are satisfied: 1) the pixels in the patch lie on the continuation of strong edges. 2) They are surrounded by the pixels having the high confidence. 3) The pixels in the patch have low depth variation. Each of these conditions for priority is called data term ($D(\mathbf{p})$), confidence term ($C(\mathbf{p})$), and occlusion constraint term ($Z(\mathbf{p})$), respectively. The priority is defined as follows:

$$P(\mathbf{p}) = C(\mathbf{p}) D(\mathbf{p}) Z(\mathbf{p}), \tag{1}$$

where

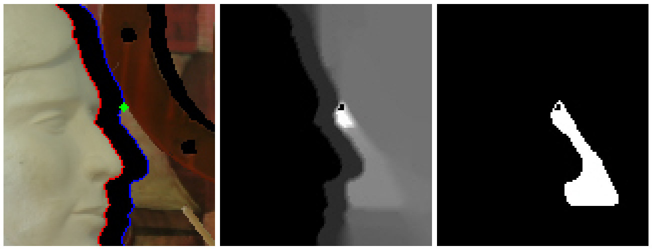
$$\begin{aligned} C(\mathbf{p}) &= \frac{\sum_{\mathbf{q} \in \Psi_{\mathbf{p}} \cap \Phi} C(\mathbf{q})}{|\Psi_{\mathbf{p}}|} \quad D(\mathbf{p}) = \frac{|\nabla I_{\mathbf{p}}^{\perp} \cdot \mathbf{n}_{\mathbf{p}}|}{\alpha} \\ Z(\mathbf{p}) &= L(\mathbf{p}) \left\{ 1 - \frac{\max_{\mathbf{q}, \mathbf{r} \in \Psi_{\mathbf{p}} \cap \Phi} d(\mathbf{q}, \mathbf{r})}{\beta} \right\}. \end{aligned} \tag{2}$$

$|\Psi_{\mathbf{p}}|$ represents the area of $\Psi_{\mathbf{p}}$. α and β are normalization factors. $d(\mathbf{q}, \mathbf{r})$ represents the difference between the disparity value between $\mathbf{q}$ and $\mathbf{r}$. $L(\mathbf{p})$ is a binary function, which indicates the label of $\mathbf{p}$ yielding 0 (or 1) if $\mathbf{p}$ is in the foreground points (or background points).

$C(\mathbf{p})$ indicates how many valid pixels are in the source region, which is updated in a recursive manner as filling proceeds, and $D(\mathbf{p})$ stands for the strength of isophotes such as edges [5]. $Z(\mathbf{p})$ prefers homogeneous depth areas along the background contour in order to assign higher priority so that patches lying on the intersection of the foreground and the background have lower priority. It is because that the patch lying on the intersection may have some portion of the foreground, which causes mismatching errors in the exemplar searching procedure. Depth ambiguities are handled by $L(\mathbf{p})$ which assigns zero priority to the point along the foreground contour. Fig. 1 shows the superiority of the occlusion constraint. The result shows that artifacts caused by depth ambiguities are remedied by using the occlusion constraint.

3.2. Random walks for finding optimal patch

To deal with problem of finding patch candidates as labeling problem in the probabilistic framework, we introduce patch-based (undirected) graph representation $G = (V, E)$ for random walks formulation. V is the set of all the point $\mathbf{p} \in I$, and E is the set of weighted edges. First, each point is classified into the appropriate label for serving as an initial seed. All the left-most points in the corresponding contour in the hole are assigned to the label "visible seed" if a virtual camera is placed on the right side of a real camera. The points in the source region are labeled as "invisible seed". Note



(a) (b) (c)

Fig. 3. Patch candidates via random walks: (a) initial seeds (b) The probability of a random walker starting from an unlabeled node. (c) The patch candidates.

that this process is based on vertical prior. Then, the points in $\delta\Omega$ are assigned to the initial seeds by corresponding labels as shown in Fig. 3 (a). The boundary along the object (red line) is served as visible seed ($s = 0$) while all the points in source region (green dot) is served as invisible seed ($s = 1$). Defining weights existing between nodes is needed to solve this labeling problem with random walks. The weight is defined as follows.

$$w_{i\,j} = e^{-\sigma S(\mathbf{p}_i, \mathbf{p}_j)} M(\mathbf{p}_i, \mathbf{p}_j) + \varepsilon, \quad (3)$$

where σ is a weighting factor which prefers the smaller difference among the patches having similar pixel values. ε is a small positive constant, which guarantees the stability. $M(\mathbf{p}_i, \mathbf{p}_j)$ is a binary function which yields 1 (or 0) if the shape of the source region of both $\mathbf{p}_i$ and $\mathbf{p}_j$, i.e., $\Psi_{\mathbf{p}_i} \cap \Phi$ and $\Psi_{\mathbf{p}_j} \cap \Phi$, are the same (or different). $S(\mathbf{p}_i, \mathbf{p}_j)$ represents the patch dissimilarity cost between $\mathbf{p}_i$ and $\mathbf{p}_j$ as follows.

$$S(\mathbf{p}, \mathbf{q}) = \frac{\sum_{\mathbf{r} \in \psi_{\mathbf{p}}} \sum_{\mathbf{s} \in \psi_{\mathbf{q}}} \| \mathbf{r} - \mathbf{s} \|}{\eta}, \quad (4)$$

where $\|\mathbf{r} - \mathbf{s}\|$ means the Euclidean distance between $\mathbf{r}$ and $\mathbf{s}$. η is a normalizing factor. From the weight as in Eq. (3), the Laplacian matrix is built as follows.

$$L_{ij} = \begin{cases} d_i & if \quad i = j, \\ -w_{ij} & if \quad \mathbf{p}_i \text{ and } \mathbf{p}_j \text{ are adjacent nodes,} \\ 0 & if \quad \text{otherwise,} \end{cases} \quad (5)$$

where d_i is the degree of node i defined as $d_i = \sum w_{ij}$ for all edges e_{ij} incident on $\mathbf{p}_i$. L can be decomposed into four submatrices by ordering seed nodes first and unseeded nodes second as follows:

$$L = \begin{bmatrix} L_M & B \\ B^T & L_U \end{bmatrix}, \quad (6)$$

where L_M and L_U correspond to the weights of seeded nodes and unseeded nodes, respectively. Therefore, solving the Dirichlet problem is equivalent to finding the roots of the following algebraic equation.

$$L_U x^s = -B^T f^s, \quad (7)$$

where x^s is the vector to be calculated which represents the probability of each unlabeled patches that will be assigned label s. f^s is the vector with degree m as in Eq. (8), i.e., the number of seeded nodes.

$$f_i^s = L(\mathbf{p}_i). \quad (8)$$

The solution of the Eq. (7) gives the probabilities meaning matching confidence on each unlabeled patch. Hence, each unlabeled patch will have a probability from which the confidence of being whether “visible” or “invisible” can be inferred. Fig. 3 (b) shows the example. Intensity value represents the probability that a patch belongs to “invisible” seed, where higher intensity value corresponds to higher probability. By assigning each unseeded patch to the label according to the highest probability, the patch candidates can be obtained as shown in Fig. 3 (c).

Verification step is also performed to the candidates. The patch candidate labeled as “invisible” is canceled out if it contains the occluded region or the average disparity of the target region exceeds that of the source region. Then, the optimal patch is selected from the patch candidates via SSD. Note that the probabilities of the “invisible” patches are used to measure the similarity between costs.

Table 1. The quantitative evaluation results for Middlebury Stereo Data set

	Interpolation [2]		Inpainting [5]		Proposed method	
Data set	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Art	25.02	0.86	23.87	0.82	26.12	0.86
Books	24.59	0.91	24.61	0.92	25.19	0.93
Dolls	25.68	0.90	25.37	0.90	25.24	0.90
Laundry	26.86	0.90	26.38	0.89	28.39	0.92

4. EXPERIMENTAL RESULTS

We use the test sequence from [8] with ground truth depth maps. View 3 is synthesized from View 1 followed by filling the holes. σ and ε are set to 400 and 10^{-6}, respectively in all experiments. A default window size is 3×3. Fig. 4 shows the result of the proposed method. It shows that both texture and structure is propagated into regions to be filled well as shown in Fig. 4 (e). The boundary of the foreground object is well preserved while other methods produce noticeable artifacts such as blurring. Fig. 4 (d) shows that the mismatching patches are penetrated into the occluded areas, which results in destroying the boundary of the objects. This structural collapsed artifact is more fatal than blurring since it is easily discerned by a user. The woody texture is continuously propagated to the edge of the detergent container in ‘Laundry’ sequence. It shows the limitation of the exemplar-based inpainting. The proposed method, however, produces more semantic result over other methods especially in ‘Laundry’ sequence even though the reconstructed image is slightly different from the ground truth. We measure PSNR and SSIM index in order to evaluate the performance of the proposed method quantitatively as shown in Table 1. It shows that the performance of the interpolation is higher than that of inpainting even though the images synthesized by inpainting show more semantic results. It is since PSNR measures only the difference of pixel values on the entire image, blurred areas would have an advantage in PSNR.

5. CONCLUSION

In this paper, we have proposed a hole filling technique robust to large-scale occlusion, which results in reconstructing the holes coherently during view synthesis. The proposed method remedies the occluded regions by jointly using image completion technique and random walks framework. Occlusion constraint as well as vertical prior is imposed in order to guide the filling priority. Optimal patch is selected from the patch candidates according to the probabilities inferred by random walks followed by verification step. In future, we will extend the proposed method so that it would be applied to multiview video.

(a) (b) (c) (d) (e) (f)

Fig. 4. The experimental results of (from top to bottom) 'Dolls', 'Books', 'Art' and 'Laundry' sequences: (a) ground truth, (b) hole regions, (c) interpolation [2], (d) inpainting [5], (e) the proposed method, (f) cropped ground truth.

6. REFERENCES

[1] A. Kubota, A. Smolic, M. Magnor, M. Tanimoto, T. Chen, and C. Zhang, "Multiview Imaging and 3DTV," *IEEE Signal Processing Magazine*, Vol. 24, No. 6, pp. 10-21, Nov. 2007.

[2] S. E. Chen, and L. Williams, "View Interpolation for Image Synthesis," *ACM SIGGRAPH*, pp. 279-288, Aug. 1993.

[3] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image Inpainting," *ACM SIGGRAPH*, pp. 417-424, Jul. 2000.

[4] M. Bertalmio, A. L. Bertozzi, and G. Sapiro, "Navier-Stokes, Fluid Dynamics, and Image and Video Inpainting," *in Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 355-362, Dec. 2001.

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

[6] L. Wang, H. Jin, R. Yang, and M. Gong, "Stereoscopic Inpainting: Joint Color and Depth Completion from Stereo Images," *in Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Jun. 2008.

[7] L. Grady, "Random Walks for Image Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 28, No. 11, pp. 1768-1783, Nov. 2006.

[8] http://vision.middlebury.edu/stereo