DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-036

A new hole filling method based on 3D geometric transformation for synthesized image

Hak Gu Kim and Yong Man Ro[1]; IVY Lab, Korea Advanced Institute of Science and Technology (KAIST); Republic of Korea

Abstract

This paper proposes a new hole filling method using 3D geometric transformation to make a synthesized image at virtual viewpoint. Disoccluded regions in virtual viewpoint could appear as hole regions in the synthesized image. The hole regions are supposed to be filled in a visually plausible manner. In most of previous works, hole regions were filled with translated patches (labels) extracted from non-hole region (source region). In many cases, however, it is difficult to find the most similar labels due to texture irregularity and perspective distortion in a wide view angle. In this paper, we propose a new hole filling method in which hole regions are filled with the best labels using 3D geometric transformation and associated nearest neighbor search. Experimental results show that the proposed method provides structurally consistent results with a high computational efficiency, which outperforms existing hole filling methods.

Introduction

There has been increasing interest of three-dimensional (3D) contents providing an enhanced viewing experience in multi-view imaging systems such as autostereoscopic display and free-viewpoint TV (FTV) [1], [2]. Sufficient 3D contents are needed to flourished 3D service. View synthesis could be one of helpful tools to enrich 3D contents, which generate virtual views at desired virtual viewpoints.

To generate virtual views with a given reference view, depth image based rendering (DIBR) is widely used. DIBR consists of two main processes, which are 3D warping and hole filling [3], [4]. In 3D warping process, a given reference view is warped to a desired virtual view with associated depth map. In the warped image at a virtual viewpoint, hole regions could occur because some regions occluded by foregrounds of the reference view could be disoccluded [5]-[7]. These hole regions are supposed to be filled with the best pixels or patches (labels) by hole filling methods [8]. The hole filling is known as a critical part that affects the quality of view synthesis.

Image inpainting method is frequently used to fill hole regions in the synthesized images [9]-[11]. Notably, unwanted distortions could arise in the synthesized image when existing image inpainting methods are directly applied to the hole filling in view synthesis. Without loss of generality, hole regions in the synthesized image come from the backgrounds occluded by the foregrounds of the reference view. Without considering the property of hole regions in the synthesized image, structural inconsistency in hole regions could arise, e.g., the absorption of foreground information into the background.

Local greedy-based hole filling methods have been proposed [12]-[14]. In [12] and [13], a depth information-based priority was combined with exemplar-based inpainting method for hole filling in view synthesis. Similarly, a local greedy-based hole filling method using structure tensor and depth information was proposed [14]. This method tried to extract structurally consistent candidate labels from background regions only. However, the local greedy algorithms could inevitably induce structural inconsistency because they did not consider neighboring regions. In recent years, to achieve structural consistency, global optimization-based hole filling methods have been proposed [15]-[18], which take into account its neighboring regions as well as hole regions. In [15], the global optimization-based image inpainting was extended for view synthesis by considering depth dissimilarity between candidate labels and target region to-be-filled. In [16]-[18], the human visual characteristics such as binocular symmetry and temporal flickering were taken into account in global optimization-based hole filling framework.

In the previous works mentioned above, translated labels from source region are copied and pasted to fill holes. In synthesized images with a wide angle, local textures could be significantly changed due to local scene shape variation such as perspective distortion or texture irregularity [19]. As a result, for wide angle synthesized image, translated labels could not be structurally consistent.

In this paper, we propose a new hole filling method based on 3D geometric transformation which aims to provide structurally consistent labels in wide angle synthesized images. For high quality synthesized images, hole regions are needed to be filled with the best visually plausible labels. To find these labels, which minimize dissimilarity with target region to-be-filled, we employ the 3D geometric transformation of labels by adjusting geometric parameters such as scale and 3D rotation factors in 3D space. With the 3D geometric transformation of candidate labels, it could allow structural consistency in the wide angle synthesized image because we consider 3D transformed labels as well as existing translated labels. In addition, the proposed hole filling method can effectively search candidate labels by nearest neighbor search. The best labels to fill hole region are likely to be existed around neighborhoods of hole region because natural images have a high spatial correlation [20], [21]. Based on this observation, we search for the neighboring regions of hole region. Experimental results show that the proposed hole filling method provides visually consistent results with high computational efficiency in wide angle synthesized image.

The reminder of this paper is organized as follows. In Section 2, we present the proposed hole filling method based on 3D geometric transformation. In Section 3, we describe experiments to evaluate the performance of the proposed hole filling method for the synthesized image. Finally, the conclusions are drawn in Section 4.

Proposed method

In this paper, we apply 3D geometric transformation based hole filling into a Markov random field (MRF) framework. Figure 1 illustrates the MRF framework for hole filling. Let **I** denote the

[1] Corresponding author (ymro@kaist.ac.kr)

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-036

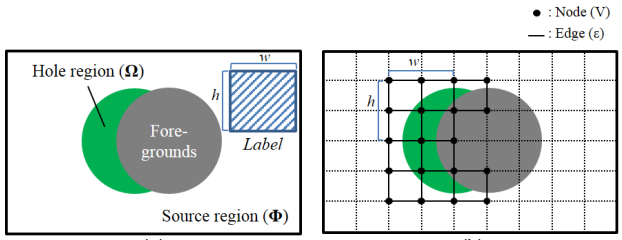


Figure 1. *Illustration of MRF framework for the hole filling. (a) Hole region (**Ω**) and source region (**Φ**) in the warped image. (b) Nodes and edges of the MRF for the hole filling.*

warped image at virtual viewpoint. **I'** denotes the geometrically transformed image (see Figure 2). As seen in Figure 1(a), the warped image (**I**) at virtual viewpoint has hole region (green region, denoted as **Ω**). Hole region represents disocclusion region on the warped image. Source region (**Φ**) means non-hole region except for foregrounds. Labels in Figure 1(a) mean small rectangle patches (size of $w \times h$) which are collected from source region in order to fill hole region. Figure 1(b) shows the nodes and edges of the MRF in hole filling framework. Notably, the nodes indicate the sampled lattice image points around hole region. The edges indicate the connections between the nearest neighboring nodes (e.g., 4 neighboring nodes).

3D Geometric Transformation based Hole Filling in Global Optimization

In this section, we present a new 3D geometric transformation based hole filling method in global optimization. The main idea of the proposed method is to minimize the cost between target region (i.e., small rectangular including source region and hole region) and transformed labels in global optimization framework. For this purpose, we design a novel MRF energy function as

$$F(\hat{x}) = \sum_{p \in V} E_p^{Data}(\mathbf{t}_p, \mathbf{c}_p) + \sum_{(p,q) \in \varepsilon} E_{p,q}^{Smooth}(\mathbf{c}_p, \mathbf{c}_q), \qquad (1)$$

where E_p^{Data} and $E_{p,q}^{Smooth}$ represent the data cost and smoothness cost for structural consistency, respectively. Note that p and q indicate node p and its neighboring node q, respectively. $\mathbf{t}_p = [t_p^x, t_p^y]^{\mathrm{T}}$ and $\mathbf{c}_p = [c_p^x, c_p^y]^{\mathrm{T}}$ are the center pixel position of the target region centered at node p and candidate label to be placed at node p, respectively. $\mathbf{c}_q = [c_q^x, c_q^y]^{\mathrm{T}}$ is the center pixel position of candidate label to be placed at node q. These candidate labels are 3D transformed labels in this paper.

The previous hole filling methods have considered pure translated labels. Instead, we consider transformed labels in 3D space. The transformed labels are found in collecting labels of geometrically transformed image (**I'**).

Let $\mathbf{h} = [h^x, h^y]^{\mathrm{T}}$ and $\tilde{\mathbf{h}} = [h^x, h^y, 1]^{\mathrm{T}}$ denote vector forms of a pixel point (h^x, h^y) in the warped image (**I**) and its homogeneous representation, respectively. Also, let $\mathbf{u} = [u^x, u^y]^{\mathrm{T}}$ and $\tilde{\mathbf{u}} = [u^x, u^y, 1]^{\mathrm{T}}$ denote vector forms of a pixel point (u^x, u^y) in the geometric transformed image (**I'**) and its homogeneous representation, respectively. Geometrically transformed image is generated by projection of world coordinates (i.e., 3D points) of **h**

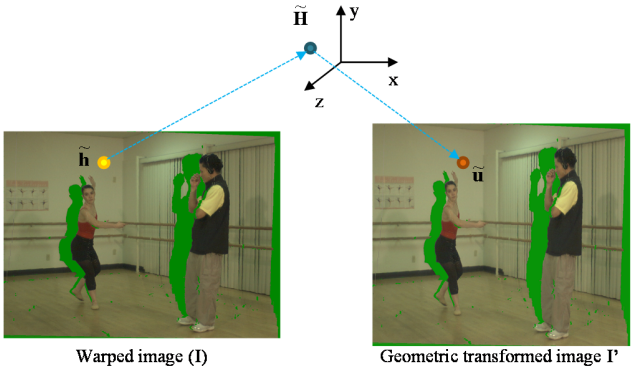


Figure 2. *Illustration of 3D geometric transformation of the warped image. For example, the geometric transformed image **I'** is obtained by setting the scale factor, s=1.2 and rotation angle about the y-axis, $\theta_y = 30°$.*

in 3D space. The homogeneous representation of the pixel points in **I'** can be written as

$$\tilde{\mathbf{u}} = s\mathbf{K}_t\mathbf{R}_t(\theta_x, \theta_y, \theta_z)^{-1}(\tilde{\mathbf{H}} - \mathbf{T}_t), \qquad (2)$$

where $\tilde{\mathbf{H}} = [H^x, H^y, H^z, 1]^{\mathrm{T}}$ is a homogeneous representation of world coordinate $\mathbf{H}=[H^x, H^y, H^z]^{\mathrm{T}}$ of image pixel point **h** in **I**. s is a non-zero scale factor. $\mathbf{K}_t$ and $\mathbf{T}_t$ are a given 3×3 intrinsic matrix and a 3×1 translation vector, respectively, for transformation. $\mathbf{R}_t(\theta_x,\theta_y,\theta_z)$ is a 3D rotation matrix in 3D space. It can be written as

$$\mathbf{R}_t(\theta_x,\theta_y,\theta_z) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x \\ 0 & \sin\theta_x & \cos\theta_x \end{bmatrix} \begin{bmatrix} \cos\theta_y & 0 & \sin\theta_y \\ 0 & 1 & 0 \\ -\sin\theta_y & 0 & \cos\theta_y \end{bmatrix} \begin{bmatrix} \cos\theta_z & -\sin\theta_z & 0 \\ \sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad (3)$$

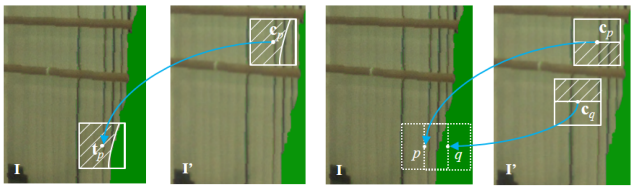
where θ_x, θ_y, and θ_z represent rotation angles about the x, y, and z-axis in 3D space, respectively. Each 3×3 matrix represents the rotation matrices by an angle θ_x, θ_y, *and* θ_z about the x, y, and z-axis, respectively.

By adjusting the scale factor and the rotation angles in 3D space, the geometric transformed image **I'** can be obtained. Figure 2 illustrates the proposed hole filling framework considering 3D geometric transformation for wide angle synthesized images. As shown in Figure 2, to find reliable labels which fit well in hole region of the synthesized image with wide angle, candidate labels are collected from 3D geometric transformed images. Consequently, transformed labels extracted from geometric transformed images as well as translated labels could be candidate labels in MRF energy function.

With 3D geometric transformation of the warped image, the proposed data cost term in Eq. (1) is defined as dis-similarity between the target region to-be-filled centered at node p and 3D geometrically transformed labels to be placed at node p. It can be written as

$$E_p^{Data}(\mathbf{t}_p, \mathbf{c}_p) = \sum_{\mathbf{d} \in \left[-\frac{w}{2} \times \frac{w}{2}\right] \times \left[-\frac{h}{2} \times \frac{h}{2}\right]} [1 - \mathbf{M}(\mathbf{t}_p + \mathbf{d})] \times [\mathbf{I}(\mathbf{t}_p + \mathbf{d}) - \mathbf{I}'(\mathbf{c}_p + \mathbf{d})]^2, \qquad (4)$$

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-036



(a) (b)

Figure 3. *Illustration of the proposed data cost and smoothness cost terms for spatial consistency. (a) The calculation of the data cost term. (b) The calculation of the smoothness cost term. The data term and smoothness cost term are calculated over the white diagonally-lined regions in (a) and (b), respectively.*

where **M** represents the binary hole mask, which is one for hole region and zero for source region.

Figure 3(a) illustrates the calculation of the proposed data cost term. In Figure 3(a), the data cost term is calculated over white diagonally-lined regions in the target region centered at $\mathbf{t}_p$ (pixel position of node p) and the candidate label centered at $\mathbf{c}_p$.

The proposed smoothness cost term in Eq. (1) is defined as dis-similarity between the 3D geometrically transformed labels to be placed at node p and its neighboring node q for structural consistency in overlapped regions. The smoothness cost term can be written as

$$E_p^{Smooth}(\mathbf{c}_p,\mathbf{c}_q)=\sum_{q\in N(p)}\sum_{\mathbf{ds}\in O(p,q)}[\mathbf{I}(\mathbf{c}_p+\mathbf{ds})-\mathbf{I'}(\mathbf{c}_q+\mathbf{ds})]^2, \quad (5)$$

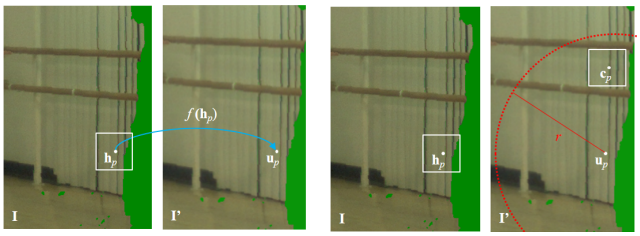
where $q\in N(p)$ represents 4 neighboring nodes of node p. $O(p,q)$ represents the overlapped region between two candidate labels to be placed at node p and q. $\mathbf{ds}=[ds^x,ds^y]^T$ indicates a shift variable vector within range of $O(p,q)$. The smoothness cost term indicates the structural consistency in the overlapped regions. The Figure 3(b) shows the proposed smoothness cost term for structural consistency.

The best labels are found by minimizing the proposed MRF energy function in synthesized image. Belief propagation (BP) is adopted as an optimization method.

Nearest Neighbor Search for efficient global optimization

In this section, we present nearest neighbor search (NNS) for hole filling to reduce computational cost of the proposed hole filling method effectively. In general, to find a global optimization solution by the conventional BP, heavy computational cost is required because all label combinations are calculated to find the best label minimizing MRF cost function. Further a lot of label combinations including transformed labels cause additional computational cost.

Without loss of generality, the best matching labels minimizing dis-similarity with target region would exist around neighboring regions of target region [20]. They are likely to be grouped into small spatial regions because natural images have a high spatial correlation [21]. From these observations, NNS could significantly reduce the computational cost by searching for reliable labels within the nearest neighbor regions of target region.



(a) (b)

Figure 4. *Illustration of the NNS for hole filling in the proposed method (a) Finding a corresponding point $\mathbf{u}_p$ of $\mathbf{h}_p$ in* **I'**. *(b) The red-dashed circle indicates the nearest neighbor region of $\mathbf{u}_p$ in* **I'**. *The candidate labels are collected in this region.*

Table 1. Dataset information and scenario for view synthesis

Datasets	Reference viewpoint	Virtual viewpoint	Frame (th)
Ballet	*Cam*4	*Cam*5	1
Breakdancer	*Cam*3	*Cam*4	1

Figure 4 illustrates the proposed NNS scheme for efficient global optimization-based hole filling. By searching the nearest neighbor regions in **I'**, the correspondence of target region to-be-filled is found. Let the nearest neighbor field (NNF) be defined as a function f [19], which is a distance of correspondences in **I** and **I'**. Given a pixel position $\mathbf{h}_p$ in **I** and its corresponding pixel position $\mathbf{u}_p$ in **I'**, $f(\mathbf{h}_p)$ can be simply calculated by $\mathbf{u}_p$ - $\mathbf{h}_p$. In generating 3D geometrically transformed image **I'**, distance values (i.e., distance between the pixel position $\mathbf{h}_p$ in **I** and its corresponding pixel position $\mathbf{u}_p$ in **I'**) are stored in an array. As a result, the corresponding point can be found simply by referring to the values of f as offsets without additional computational costs, as shown in Figure 4(a). Then, reliable labels are collected with high beliefs around corresponding point of the target region to-be-filled (see Figure 4(b)). Note that the belief of each label in BP indicates the probability of placing the candidate labels at the hole region. The nearest neighbor region is defined as a region in radius r. In our experiment, radius r is experimentally set to 200.

Consequently, we can considerably reduce computational cost while reducing hole filling quality loss. In addition, the proposed method can provide structural consistency since hole regions are likely filled with the best matching labels around the nearest neighbor regions.

Experiments and Results

To verify the proposed hole filling method, we performed experiments. In our experiments, two multi-view color-plus-depth images, which were Ballet and Breakdancer from Microsoft research [22] were used. They were wide angle views and had perspective distortions by circular camera arrangement.

Table 1 illustrates test datasets and view synthesis conditions in our experiment. Given color and the associated depth images, the warped images at virtual viewpoints were generated by a 3D warping in MPEG view synthesis reference software (VSRS) [7]. For performance comparisons in wide angle synthesized images, we filled hole regions by using three hole filling methods, which were Ahn's method [14], Habight's method [15], and the proposed method.

DOI: 10.2352/ISSN.2470-1173.2016.5.SDA-036

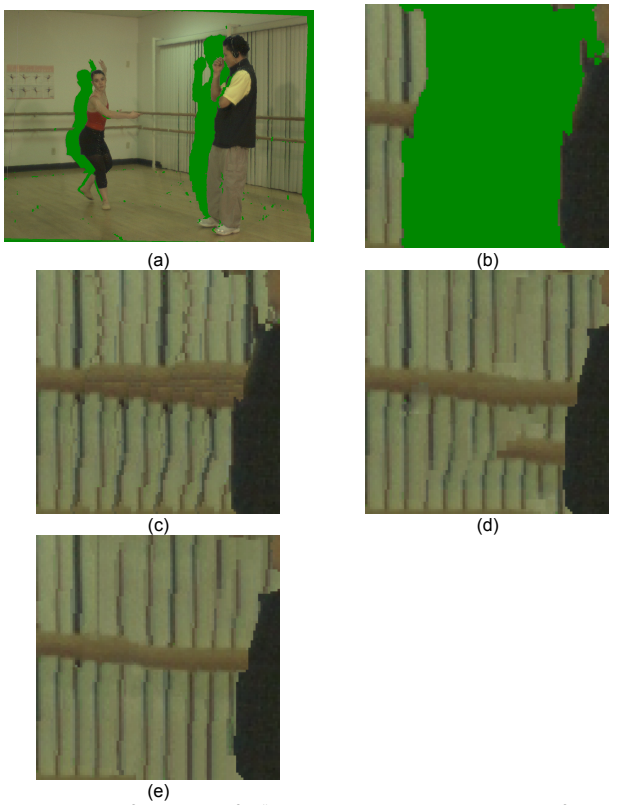


Figure 5. *Hole filling results for "Ballet". (a) Warped image. (b) Magnified parts of (a). (c) Ahn's method [14]. (d) Habigt's method [15]. (e) Proposed method.*

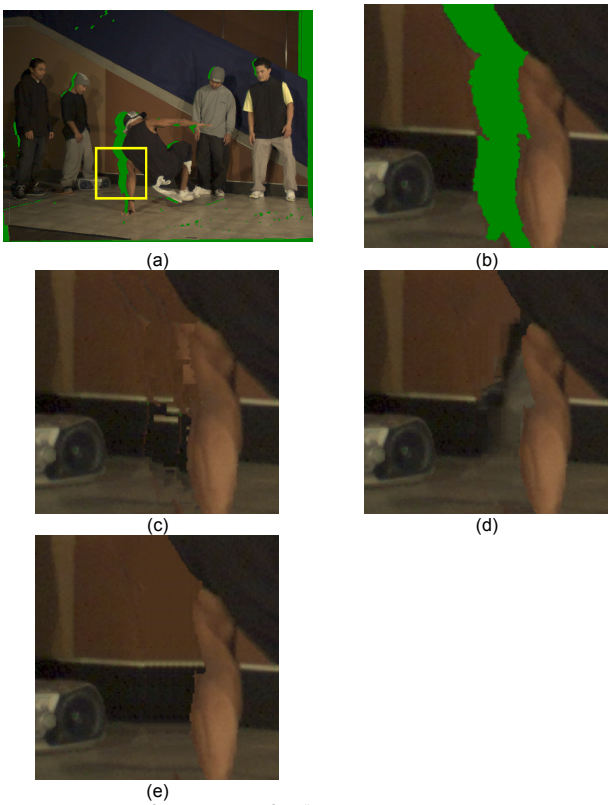


Figure 6. *Hole filling results for "Breakdancer". (a) Warped image. (b) Magnified parts of (a). (c) Ahn's method [14]. (d) Habigt's method [5]. (e) Proposed method.*

The size of label was 9×9 (i.e., $w = h = 9$). In our experiments, search range was $[-\pi/6, \pi/6]$ for each rotation and [0.8, 1.2] for scale for the proposed hole filling method.

Figure 5 and Figure 6 show visual results of three hole filling methods for "Ballet" and "Breakdancer", respectively. In these figures, the first and second row present warped image and magnified parts of the warped image, respectively. Figure 5(c), (d) and Figure 6(c), (d) show hole filling results of the existing hole filling methods. As shown these figures, structurally inconsistency in hole region could be seen. The existing hole filling methods had perspective distortion or texture irregularity. They failed to maintain structural consistency in the synthesized image. On the other hand, Figure 5(e) and Figure 6(e), which have been obtained by the proposed hole filling method, show structurally consistent results in wide angle synthesized image.

To evaluate computational efficiency of the proposed method using NNS, we measured computational gain of the proposed method. We calculate computational gain by measuring the ratio of run time of BPs with and without the proposed NNS scheme.

Table 2 shows the computational gain of the proposed method. As shown in Table 2, the proposed NNS scheme in BP obtained about 50 times more computational efficiency. In particular, the proposed hole filling method provided visually consistent results in wide angle synthesized images.

Table 2. Computational gain for view synthesis

	Ballet	Breakdancer
Computational gain	× 50.12	× 49.23

Conclusions

In hole filling for a wide angle view synthesis, it is difficult to find the best matching labels due to texture irregularity and perspective distortions. To cope with this difficulty, this paper proposed a novel hole filling method based on 3D geometric transform for a wide angle synthesized image. By collecting the candidate labels from 3D geometrically transformed images, we could find the best matching labels in global optimization framework. In addition, a computational efficiency could be achieved by nearest neighbor search. Experimental results demonstrated that the proposed hole filling method provided structure-consistent results in wide angle synthesized image with a low computational cost.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No.2015R1A2A2A01005724).

References

[1] I. P. Howard and B. J. Rogers, Seeing in Depth, I Porteous: Ontario, 2002.

[2] N. A. Dodgson, “Multi-view autostereoscopic 3D display,” presented at the Stanford Workshop 3-D Imaging, Cambridge, U.K., 2011.

[3] L. Zhang, C. Vazquez, and S. Knorr, “3D-TV content creation: Automatic 2d-to-3d video conversion,” IEEE Trans. Broadcast., Jun. 2011.

[4] L. Tran, C. Pal, and T. Nguyen, “View synthesis based on conditional random fields and graph cuts,” in Proc. IEEE Int’l Conf. on Image Processing (ICIP), pp. 433-436, Sep. 2010.

[5] P. Ndjiki-Nya, M. Koppel, D. Doshkov, H. Lakshman, P. Merkle, K. Muller, and T. Wiegand, “Depth image-based rendering with advanced texture synthesis for 3D video,” IEEE Trans. Multimedia, vol. 13, no. 3, pp. 453-465, Jun. 2011.

[6] C. Fehn, “Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3DTV,” in Proc. SPIE Stereoscopic Displays and Virtual Reality Systems, 5291, pp. 93–104, May. 2004.

[7] O. Stankiewicz, K. Wegner, M. Tanimoto, M. Domanski, "Enhanced view synthesis reference software (VSRS) for Free-viewpoint Television," ISO/IEC JTC1/SC29/WG11 MPEG2013/M31520, 2013.

[8] K.-J. Oh, S. Yea, and Y.-S. Ho, “Hole filling method using depth based inpainting for view synthesis in free viewpoint television and 3-D video,” in Proc. Picture Coding Symp., Chicago, IL, May 2009.

[9] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, “Image inpainting,” in Proc. ACM SIGGRAPH, 2000.

[10] A. Criminisi, P. Pérez, and K. Toyama, “Region filling and object removal by exemplar-based image inpainting,” IEEE Trans. Image Process., vol. 13, no. 9, pp. 1200–1212, Sep. 2004.

[11] N. Komodakis and G. Tziritas, “Image completion using efficient belief propagation via priority scheduling and dynamic pruning”, IEEE Trans. Image Process., vol. 16, no. 11, pp. 2649–2661, Nov. 2007.

[12] I. Daribo and H. Saito, “A novel inpainting-based layered depth video for 3DTV,” IEEE Trans. Broadcast., vol. 57, no. 2, pp. 533-541, Jun. 2011.

[13] I. Daribo and B. Pesquet-Popescu, “Depth-aided image inpainting for novel view synthesis,” in Proc. IEEE Int. Workshop Multimedia Signal Process., Oct. 2010.

[14] I. Ahn and C. Kim, “A novel depth-based virtual view synthesis method for free viewpoint video,” IEEE Trans. Broadcast., vol. 59, no. 4, pp. 614-626, Dec. 2013.

[15] J. Habigt and Diepold, “Image completion for view synthesis using Markov random fields and efficient belief propagation”, in Proc. IEEE Int’l Conf. on Image Processing (ICIP), pp. 2131-2134, Sept. 15-18, 2013.

[16] H. G. Kim, Y. J. Jung, S. S. Yoon, and Y. M. Ro, “Multi-view stereo image synthesis using binocular symmetry based global optimization,” in *Proc. SPIE*, vol. 9391, pp.93910X, 2015.

[17] H. G. Kim, Y. J. Jung, S. S. Yoon, and Y. M. Ro, “Temporally consistent hole filling method based on global optimization with label propagation for 3D video”, in Proc. IEEE Int’l Conf. on Image Processing (ICIP), 2015.

[18] H. G. Kim and Y. M. Ro, “Multi-view stereoscopic video hole filling considering spatio-temporal consistency and binocular symmetry for synthesized 3D video,” IEEE Trans. Circuits Syst. Video Technol., 2015 (accepted).

[19] Jia-Bin Huang, Sing Bing Kang, Narendra Ahuja, and Johannes Kopf, “Image completion using planar structure guidance,” ACM Trans. Graph., vol. 33, no. 4, pp. 129:1–129:10, 2014.

[20] C Barnes, E. Shechtman, A. Finkelstein, D. B. Goldman, “PatchMatch: A randomized correspondence algorithm for structural image editing,” ACM Trans. Graphics, vol. 28, no. 3, Aug. 2009.

[21] V. Kober, M. Mozerov, and J. Alvarez-Borrego "Nonlinear filters with spatially-connected neighborhoods," Optical Engineering, vol. 40, no. 6, pp. 971-983, Jan. 2001.

[22] C. L. 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, no. 3, pp. 600–608, Aug. 2004.

Author Biography

Hak Gu Kim received the B.S. and M.S. degree from Inha University, Incheon, South Korea, in 2012 and 2014, respectively. He is currently working toward the Ph.D. degree at Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea. His research interests include 3D image/video processing, human 3D perception, and visual quality assessment.

Yong Man Ro received Ph.D. degrees from KAIST. He was a researcher at Columbia University and a research fellow at the UC, Berkeley. He is currently a professor and the chair of signals and systems group of the school of electrical engineering in KAIST. His research interests are image processing, 3-D video processing, computer vision, visual recognition. Dr. Ro received the young investigator finalist award of ISMRM. He served as an associate editor for IEEE SPL.