INTER-VIEW CONSISTENT HOLE FILLING IN VIEW EXTRAPOLATION FOR MULTI-VIEW IMAGE GENERATION

*Soo Sung Yoon, Hosik Sohn, Yong Ju Jung, and Yong Man Ro**

Departments of EE, Korea Advanced Institute of Science and Technology (KAIST), Republic of Korea

ABSTRACT

This paper proposes a new inter-view consistent hole filling method in view extrapolation for multi-view image generation. In stereopsis, inter-view consistency regarding structure, color, and luminance is one of the crucial factors that affect the overall viewing quality of three-dimensional image contents. In particular, the inter-view inconsistency could induce visual stress on the human visual system. To ensure the inter-view consistency, the proposed method suggests a hole filling method in an order from the nearest to farthest view to the reference view by propagating the filled color information in the preceding view. In addition, a novel depth map filling method is incorporated to achieve the inter-view consistency. Experimental results show that the proposed method significantly improves the inter-view consistency for multi-view images and depth maps, compared to those of previous methods.

Index Terms— Inter-view consistency, view synthesis, multi-view image, hole filling.

1. INTRODUCTION

Three-dimensional (3D) displays have been re-emerged as one of the major technology trends. In the mass market, we now have witnessed a wide spread of 3D content services. In particular, multi-view (MV) autostereoscopic displays can immerse viewers in a unique viewing experience by presenting different multiple perspectives of a scene without wearing eye glasses [1]. However, due to the limitation of network bandwidth and storage capacities, multiple viewpoint images (*i.e.*, virtual view images) need to be generated from only fewer number of reference view images (*e.g.*, 2 or 3 views) [2].

Depth image-based rendering (DIBR) is one of the widely used techniques to synthesize virtual view images for autostereoscopic displays [1, 3]. The basic principle of DIBR methods is to generate virtual view images using projective geometry given from reference view images and their associated depth maps (DMs) [4]. One of the most challenges for DIBR is how to fill the holes caused by disocclusion region [3, 5]. It occurs when some backgrounds (BG) appear in the virtual view images, which are occluded by foreground (FG) objects in the reference views [3]. This becomes more critical for view extrapolation where only one reference view image is available [1].

For the hole filling, most of previous studies have focused on spatial consistency within each viewpoint image [1, 6, 7, 8]. In particular, considering that most of the hole (*i.e.*, disocclusion region) exists on the BG [3, 8], previous approaches have tried to fill in the hole by using region information of the BG to avoid penetrating the region information of the FG objects into the hole. By doing so, the spatial consistency could be retained in terms of image structure, color, and luminance between BG and FG objects in a synthesized virtual view. However, it should be pointed out that those previous methods could not always provide an inter-view consistency across different virtual views, because they focused on the spatial consistency within each viewpoint.

From the perceptual point of view for stereopsis, the *inter-view consistency* is one of the crucial factors that affect the overall viewing quality of 3D contents, as mentioned earlier studies [9, 10, 11]. In particular, the inter-view inconsistency could induce visual stress on the human visual system. The lack of inter-view consistency among virtual views could induce binocular rivalry, binocular suppression, and binocular superposition, *etc* [11, 12]. As a perceptual consequence, depth distortion and visual discomfort could be induced [11]. In order to provide satisfactory viewing quality for MV images, inter-view consistent hole filling is of great importance.

In this paper, we propose a new inter-view consistent hole filling method in view extrapolation for multi-view image generation. In particular, we devise a novel exemplar-based inpainting method that fills the hole by utilizing corresponding color information in an adjacent view and the filled depth information of the current view. This allows the achievement of inter-view consistency since the filled color information in the adjacent view is propagated across consecutive views. We further propose a novel depth map filling method that exploits the depth information of adjacent views for the inter-view consistency with reliable depth information. The filled DM with inter-view consistency is used to aid our inter-view consistent color image filling method. Experimental results have demonstrated that the proposed approach can provide much higher degree of inter-view consistency of MV images than those of previous methods.

The rest of the paper is organized as follows. In Section 2, we present the overall process of the proposed hole filling method. Section 3 presents the validation experiments that evaluate the performance of the proposed hole filling method. Finally, the conclusions are drawn in Section 4.

* Corresponding author (ymro@ee.kaist.ac.kr)

978-1-4799-5751-4/14/$31.00 ©2014 IEEE

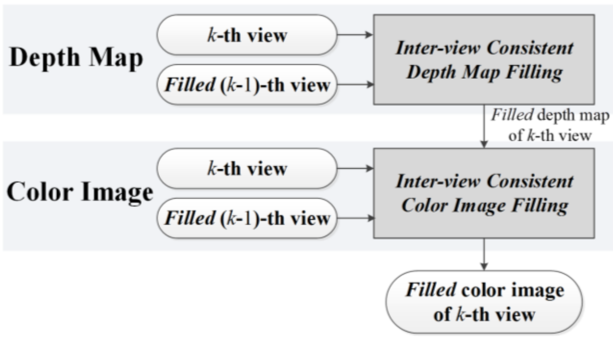


Fig. 1. Overall framework of the proposed method. Note that *k*-th view and (*k*-1)-th view are the target and adjacent views respectively.

2. PROPOSED HOLE FILLING METHOD

Given a color image of the reference view and its associated depth map (DM), virtual view images and DMs are generated by 3D warping [3]. In order to fill the hole in the virtual view images while providing the inter-view consistency, the proposed method fills the hole in an order from the nearest to the farthest views to the reference view so that filled color information in the adjacent view is propagated across consecutive views. (*e.g.*, from (*k*-1)-th view to *k*-th view in Fig. 1). Let two consecutive virtual views denoted as (*k*-1)-th and *k*-th views. The (*k*-1)-th virtual view is closer to the reference view than the *k*-th view.

The framework of the proposed method consists of two modules; inter-view consistent depth map filling and inter-view consistent color image filling, as illustrated in Fig. 1. We firstly perform the inter-view consistent depth map filling to aid the inter-view consistency in the color image filling. In this process, the filled DM of the adjacent view ((*k*-1)-th view) is utilized to make the filled DM of the target view (*k*-th view). Subsequently, the inter-view consistent color image filling method is applied with the filled DM of the target view and filled color image of the adjacent view. In the following subsections, we present the details of two modules of the proposed method.

2.1. Inter-view Consistent Depth Map Filling

For obtaining the inter-view consistent multi-view (MV) images, it is very essential to generate the inter-view consistent DMs. In this paper, we propose a novel line-wise depth map filling by using the adjacent view. The inter-view depth consistency is achieved by referring the filled depth map of the adjacent view.

In conventional line-wise DM filling methods [1], the hole is filled with depth values of neighboring backgrounds (BG) within an individual target view. However, the depth values that need to be filled do not always exist within the individual view (see patches in the target view in Fig. 2(a)). Fig. 2(a) shows the depth maps of adjacent and target views. As seen in Fig. 2(a), the depth values for the target view could be available in the adjacent view.

The novel line-wise filling method proposed in this paper is applicable to the case when the depth information does not

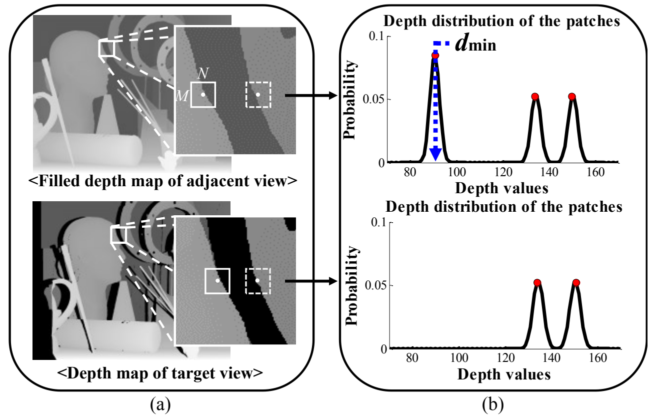


Fig. 2. The selection process of the depth value to fill the hole in the proposed depth map filling method. (a) Depth maps of adjacent and target views with boundary patches, which centers are located at the boundary of the hole. (b) Depth distributions of the boundary patches in both views in (a). Note that the boundary patches in the adjacent view correspond to the patches in the target view. The corresponding patches are obtained by using the depth values of the centers of the two patches. The red dots in (b) represent the local maxima. The black color in (a) represents the hole.

exist in the target view. To that end, we analyze the depth distributions of the local patches around the boundary of the hole in the target view and the corresponding boundary in the adjacent view. In particular, in order to find the depth value for filling the hole in the target view, the proposed DM filling method utilizes the local maxima in the depth distributions of local patches of the adjacent and target views.

To detect the local maxima, we estimate the depth distributions of the two boundary patches (*i.e.*, local patches) from the depth histograms of the each adjacent and target view. In this paper, the kernel density estimation method [13] is used to estimate the depth distributions. In the case where the number of local maxima for the boundary patches in the target view is smaller than that of the adjacent view (*e.g.*, in Fig. 2(b)), we regard that the depth values to fill the hole are not available in the target view. In this case, the minimum depth value among the local maxima in the adjacent view (*i.e.*, $d_{\min}$ in Fig. 2(b)) is selected to fill the horizontal line in the hole of the target view. This is in line with the fact that the holes mainly exist on the BG [3]. In the other case (*i.e.*, the number of local maxima in the target view is larger than that of the adjacent view), the minimum depth value among the local maxima in the target view is selected to fill the line. When the hole does have any correspondence, the filled depth values of each line in the adjacent view are used to fill the hole in the target view.

Note that, for each line in the hole, the boundary patches are defined as rectangular patches which centers are located at the outer boundary of the hole. That is, the boundary patches are centered at the left- and right-most pixels on each horizontal line of the hole (see Fig. 2(a)).

2.2. Inter-view Consistent Color Image Filling

In this section, we present a novel exemplar-based inpainting method to achieve the inter-view consistency. The main idea is to design an improved matching cost that takes the inter-view consistency into account. By doing so, we expect to preserve the inter-view consistency as well as the spatial consistency.

The proposed color image filling method fills the hole in a consecutive order from the nearest to the farthest views to the reference view. Therefore, in general, there are overlapping holes, *i.e.*, the holes that have correspondences between the target view and the adjacent view (refer to the regions between yellow dashed lines in the two neighboring views in Fig. 3). For the inter-view consistency, it is essential to ensure the consistency of such overlapping holes.

As such, the best matching candidate patch is selected by considering 1) the inter-view consistency between the overlapping holes in the adjacent view and 2) the spatial consistency between the source region and hole. The cost function to find the best matching candidate patch $\mathbf{\Psi}_{c'}$ is defined as follows:

$$\mathbf{\Psi}_{c'} = \arg\min_{\mathbf{\Psi}_{c'} \in \mathbf{\Phi}} \{\omega_1 \cdot M(\mathbf{\Psi}_{t,h}, \mathbf{\Psi}_{c,h}) + \omega_2 \cdot M(\mathbf{\Psi}_{t,s}, \mathbf{\Psi}_{c,s})\}, \quad (1)$$

where $\mathbf{\Psi}_{t,h}$ and $\mathbf{\Psi}_{t,s}$ denote the hole and source regions within the *target* patch, respectively. $\mathbf{\Psi}_{c,h}$ and $\mathbf{\Psi}_{c,s}$ denote the regions corresponding to the hole and source regions within the *candidate* patch, respectively. And $M(\cdot, \cdot)$ is a matching criterion between color information of the patches, defined by the sum of squared differences. In the cost function, color information of each pixel of the hole within the target patch (*i.e.*, $\mathbf{\Psi}_{t,h}$) is obtained by using the filled depth values of the target view in Section 2.1 and the already filled color values of the adjacent view (see the red regions in Fig. 3(a)). The first and second terms in the cost function are related with the inter-view and spatial consistency, respectively. ω_1 and ω_2 are the weights for the inter-view and spatial consistency, respectively.

Note that, for the non-overlapping holes (*i.e.*, the holes that do not have correspondences between the target view and the adjacent view), the proposed method considers only the spatial consistency term in Eq. (1). That is, the weight for the inter-view consistency ω_1 is set to zero. Also, prior to filling the hole, the hole area is slightly dilated (*e.g.*, two pixels) to reduce the boundary artifacts around the hole caused by mismatches between color and associated depth values [3, 14].

The filling order and search area for candidate patches are critical factors in MV image generation [6, 15]. Notably, the hole in the virtual view mostly appears on the boundaries between the foreground (FG) object and their neighboring BG [3, 6]. Furthermore, holes mostly occur on the BG regions occluded by a foreground object. This implies that holes have higher coherence with pixels of the neighboring BG rather than the FG objects. In this paper, the exemplar-based inpainting is enhanced such that the hole placed near a BG is filled with higher priority and only surrounding BG is regarded as source regions [1, 6, 8]. This enables the holes to be filled with only BG information.

The filling order of the proposed method is defined as follows:

$$P(p) = C(p)^{\alpha} \cdot D(p)^{\beta} \cdot E(p)^{\gamma}, \quad (2)$$

where *C(p)* is the confidence term that indicates the reliability of the current patch, and *D(p)* is the data term that gives a special priority to the isophote direction [6]. *E*(*p*) is obtained by inversion of the DM. It allows for giving higher priority to the

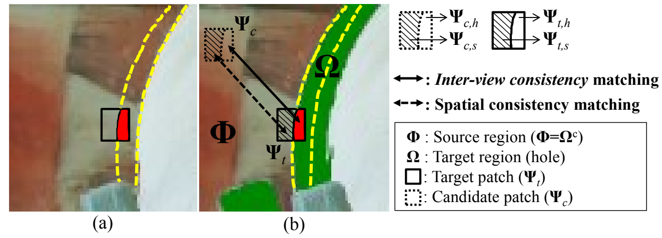


Fig. 3. Illustration of the proposed color image filling. (a) Filled color image of adjacent view. (b) Color image of target view. Note that the regions between the yellow dashed lines represent the overlapping hole. Also, color information of the hole within the target patch (*i.e.*, $\mathbf{\Psi}_{t,h}$) is obtained by using the filled depth values of target view and the filled color information of the adjacent view (refer to the red regions in the figure).

boundary of the hole near the BG since the inverted depth values in BG are higher than those in the FG objects. In addition, in order to avoid matching patches from the FG object, all of the candidate patches are selected from the BG only. In particular, a patch is regarded as the candidate patch only if all of depth values in a candidate patch are lower than $d_t+\varepsilon$. Here, d_t is the minimum depth value among the local maxima of a target patch. And ε is a tolerance parameter.

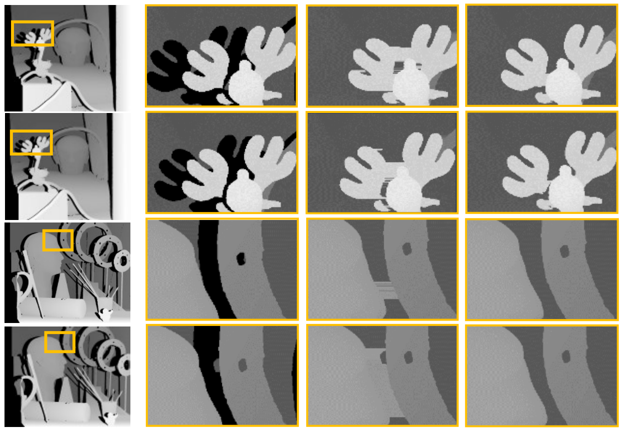
3. EXPERIMENTS AND RESULTS

In this section, experiments were performed to verify the effectiveness of the proposed method. The datasets used for the experiment were collected from the Middlebury datasets [16]. In the experiment, we used “Reindeer”, “Laundry”, “Art”, and “Dolls” images because they had various textures and depth levels. The spatial resolutions of the images are 1342×1110, 1342×1110, 1390×1110, and 1390×1110 pixels, respectively.

Given the reference image and depth map (DM), four different virtual views were extrapolated in the left direction by the 3D warping technique in the MPEG view synthesis reference software (VSRS version3.5) [17]. The Middlebury datasets that we used consisted of seven different view images for each image data. Among the view images, we selected *View*5 as a reference view for each image data. And *View*4, *View*3, *View*2, and *View*1 were the virtual views that we generated. *View*4 was the nearest view from the reference view.

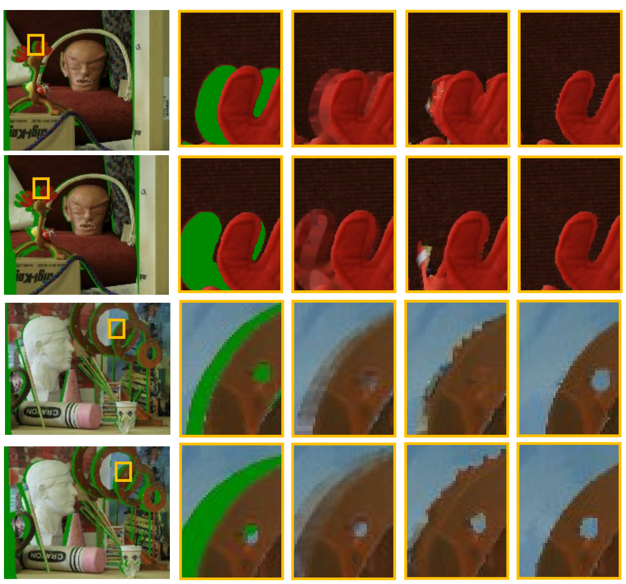
We used a size of 9×9 pixels patch for the color image and DM filling. Note that, in the DM filling, we used a relatively small patch size to preserve the local characteristics [18]. The weighting factors ω_1 and ω_2 in Eq. (1) were respectively selected as 0.9 and 0.1 to give a more weight on the inter-view consistency. Also, the parameters α, β, and γ in Eq. (2) were empirically set to 0.5, 0.2, and 7, respectively. We normalized the *C(p)*, *D(p)*, and *E(p)* in Eq. (2) with values ranging from 0 to 255 to give the same effects on the filling order.

For the DM filling, we compared the result of the proposed DM filling method with that of Ndjiki-Nya’s method [1]. As seen from Fig. 4(c), the DM results of [1] showed inter-view inconsistency as well as distortion of depth values. Due to the inter-view inconsistency, unnatural depth perception and visual discomfort could be induced. In contrast, the proposed DM filling method showed inter-view consistency and had no distortion of depth values as seen in Fig. 4(d).



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

Fig. 4. Depth map filling results. (a) Depth map with holes. (b) Magnified of (a). (c) Ndjiki-Nya's depth map filling method [1]. (d) Proposed depth map filling method. The black color in (a) and (b) represents the hole. The first and third rows are *View*2 for each image data. Also, the second and fourth rows are *View*1. The first two rows and the others are 'Reindeer' and 'Art', respectively.



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

Fig. 5. Color image filling results. (a) Color images with holes. (b) Magnified of (a). (c) MPEG VSRS [17]. (d) Criminisi's method [19]. (e) Proposed color image filling method. The green color in (a) and (b) represents the hole. The first and third rows are *View*2 for each image data. Also, the second and fourth rows are *View*1. The first two rows and the other rows are "Reindeer" and "Art", respectively.

Furthermore, in order to compare the color image filling results, MPEG VSRS [17] and Criminisi's method [19] were used as competitive methods. As seen from Figs. 5(c)-(d), the inter-view inconsistency was observed in the previous methods (*i.e.*, inconsistency between the 1st and 2nd rows, and 3rd and 4th rows of Figs. 5(c)-(d)). On the other hand, in the proposed method, the consistency between the virtual views was achieved as shown in Fig. 5(e). Namely, mismatches between views did not arise when filling the virtual view images.

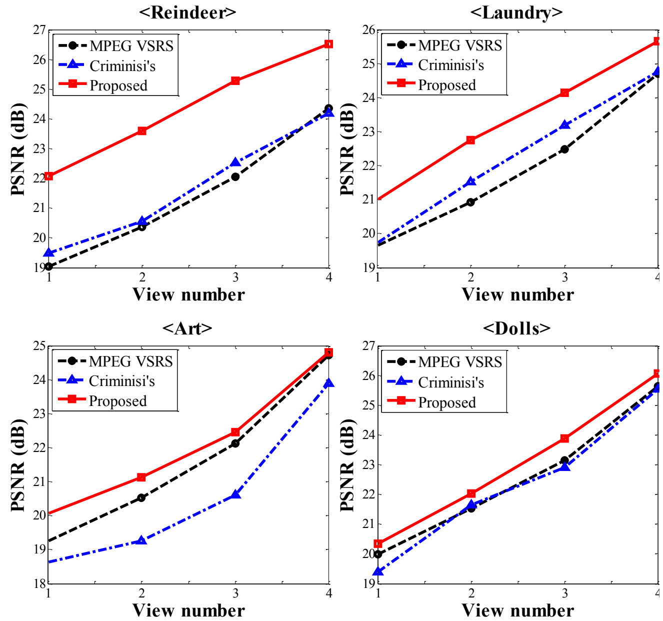


Fig. 6. PSNR comparisons of the proposed color image filling method with previous methods over views. The graphs from top-left to down-right correspond to "Reindeer", "Laundry", "Art", and "Dolls", respectively.

Table 1. Average PSNR values (dB)

Method	Reindeer	Laundry	Art	Dolls
MPEG VSRS [17]	21.45	21.94	21.66	22.57
Criminisi's [19]	21.69	22.30	20.60	22.38
Proposed	**24.36**	**23.39**	**22.11**	**23.09**

Finally, in order to evaluate the effectiveness of the proposed method quantitatively, we computed the peak signal-to-noise ratio (PSNR) for the results from Fig. 5(c)-(e) with the original images. As seen from Fig. 6 and Table 1, the proposed method outperformed the previous methods. Overall, the results showed that the proposed method could effectively fill the hole in terms of the inter-view consistency, compared to the previous methods.

5. CONCLUSION

In this paper, we proposed an inter-view consistent hole filling method in view extrapolation for multi-view image generation. The experimental results demonstrated that the proposed method could significantly improve the inter-view consistency for multi-view images and depth maps, compared to those of previous methods. In addition, quantitative results showed that the proposed method outperformed the previous methods. For the future work, we plan to add a temporally consistent hole filling framework to handle the video hole filling. Also, future research will develop an objective quality metric for synthesized multi-view stereo images. Moreover, we will demonstrate the effectiveness of our proposed hole filling method objectively via the synthesized multi-view stereo quality metric rather than objective image quality metric such as PSNR.

6. REFERENCES

[1] 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.

[2] L. Tran, C. Pal, and T. Nguyen, "View synthesis based on conditional random fields and graph cuts," in *Proc. IEEE Int'l Conf. Image Processing*, pp. 433-436, Sep. 2010.

[3] 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 Picture Coding Symp*., pp. 1-4, May. 2009.

[4] 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.

[5] H.-An. Hsu, C.-K. Chiang, and S.-H. Lai, "Spatio-temporally consistent view synthesis from video-plus-depth data with global optimization," *IEEE Trans. Circuits and Systems for Video Technology,* vol. 24, no. 1, pp. 74-84. Jan. 2014.

[6] I. Daribo and H. Saito, "A novel inpainting-based layered depth video for 3DTV," *IEEE Trans. on Broadcasting*, vol. 57, no. 2, pp. 533-541, Jun. 2011.

[7] C.-M. Cheng, S.-J. Lin, and S.-H. Lai, "Spatio-temporally consistent novel view synthesis algorithm from video-plus-depth sequences for autostereoscopic displays," *IEEE Trans. Broadcasting*, vol. 57, no. 2, pp. 523–532, Jun. 2011.

[8] I. Ahn and C. Kim, "Depth-based disocclusion filling for virtual view synthesis," in *Proc. IEEE Int. Conf. Multimedia Expo,* pp. 109–114, Jul. 2012.

[9] S. Lee, Y. J. Jung, H. Sohn, F. Speranza, and Y. M. Ro, "Effect of stimulus width on the perceived visual discomfort in viewing stereoscopic 3-D-TV," *IEEE Trans. Broadcasting*, vol. 69, no. 4, pp. 580-590, Dec. 2013.

[10] F. L. Kooi and A. Toet, "Visual Comfort of binocular displays," in *Displays*. New York: Elsevier, 2004.

[11] Y. Jung, H. Sohn, S. Lee, Y. M. Ro, and H. W. Park, "Quantitative measurement of binocular color fusion limit for non-spectral colors," *Optics Express*, vol. 19, no. 8, pp. 7325-7338, Apr. 2011.

[12] J. Boxtel, D. Alais, C. J. Erkelens, and R. Ee, "The Role of Temporally Coarse Form Processing during Binocular Rivalry," *PLoS ONE*, vol. 3, no. 1, Jan. 2008, DOI:10.1371/journal.pone.0001429.

[13] J.M.L. Murillo and A.A. Rodriguez, "Algorithms for Gaussian band-width selection in kernel density estimators," *in Proc. Neural Inf. Process. Syst*., 2008.

[14] Y. J. Jeong, Y. J. Jung, and D. Park, "Depth image based rendering for multi-view generation," *Journal of the Society for Information Display (JSID)*, vol. 18, no. 4, pp. 310-316, Apr. 2010.

[15] R. Chen, X. Li, "Fill in Occlusion Regions on Remotely Sensed Images Using Texture Synthesis Technique," in *Proc. IEEE Int'l Conf. Internet Computing and Information Services,* pp. 54-57, Sep. 2011.

[16] Middleburry stereo datasets, available at http://vision.middlebury.edu/stereo/data/

[17] M. Tanimoto, T. Fujii, and K. Suzuki, "View Synthesis Algorithm in View Synthesis Reference Software 2.0 (VSRS 2.0)," Lausanne, Switzerland, ISO/IEC JTC1/SC29/WG11 M16090, Feb. 2008.

[18] H. Zhou and J. Zheng, "Adaptive patch size determination for patchbased image completion," in *Proc. IEEE Int'l Conf. Image Processing*, pp. 421–424, Sep. 2010.

[19] 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.