

# Multi-directional Hole Filling Method for Virtual View Synthesis

Ji-Hun Mun<sup>1</sup> · Yo-Sung Ho<sup>1</sup>

Received: 30 March 2015 / Revised: 14 October 2015 / Accepted: 19 October 2015 / Published online: 29 October 2015  
© Springer Science+Business Media New York 2015

**Abstract** In 3D video applications, view synthesis is widely used for generating a virtual viewpoint images. However, the quality of synthesized image is dissimilar to captured color image. Furthermore, noise occurs at object boundary areas in synthesized images. In this paper, we propose a new hole filling method to improve the synthesized image quality. In order to synthesize an image without the noise, we remove the boundary noise by expanding it as a hole region. Based on the expanded hole region, we use proposed hole scanning and directional hole filling method in simultaneous time. Additionally, we compute a cost value among the candidate pixels to find out proper hole filling pixel. We conducted experiments on various test sequences for evaluation of the proposed method. The proposed method produces better objective results than other conventional hole filling methods.

**Keywords** 3D-warping · View synthesis · Hole filling · Free viewpoint

## 1 Introduction

In general, the 3D video system supports immersion and reality to viewers. Multi-view camera systems are broadly used for producing 3D video content. There have been many re-

search activities on camera geometrics and parameter alignment techniques.

Due to the technical development, the demand for 3D video systems has incredibly increased. 3D video technology is widely used in various fields, e.g., films, games, medical imaging, and TV broadcasting systems. The technology allows viewers to enjoy more realistic stereoscopic images in a number of viewpoints. 3D video technology will affect next generation video content business. This business prospect is already proved by a number of research data. Since 2002, ATTEST (advanced three dimensional television system technologies) of Europe has conducted extensive researches on 3D processing techniques [1]. Similarly, 3D image processing and holography techniques have been broadly studied [2]. Furthermore, Japan also has developed free viewpoint TV systems using ray-space techniques [3].

In addition, hybrid multi-view video systems are widely used to generate 3D video content. In general, hybrid multi-view camera systems consist of multiple color cameras and depth cameras, e.g., Kinect and ToF (Time-of-Flight) cameras [4]. A clock generator is used for synchronization of multiple color and depth cameras. Thus, multiple color camera and depth camera can capture the scene in simultaneous time.

Figure 1 describes the overall procedure to generate 3D video content using a hybrid multi-view camera system. Basically, color and depth cameras are calibrated to extract the camera parameters. Then, we perform color correction to diminish color inconsistency among multiple color cameras. Subsequently, multi-view image rectification is conducted for compensating geometric disparities of such cameras [5].

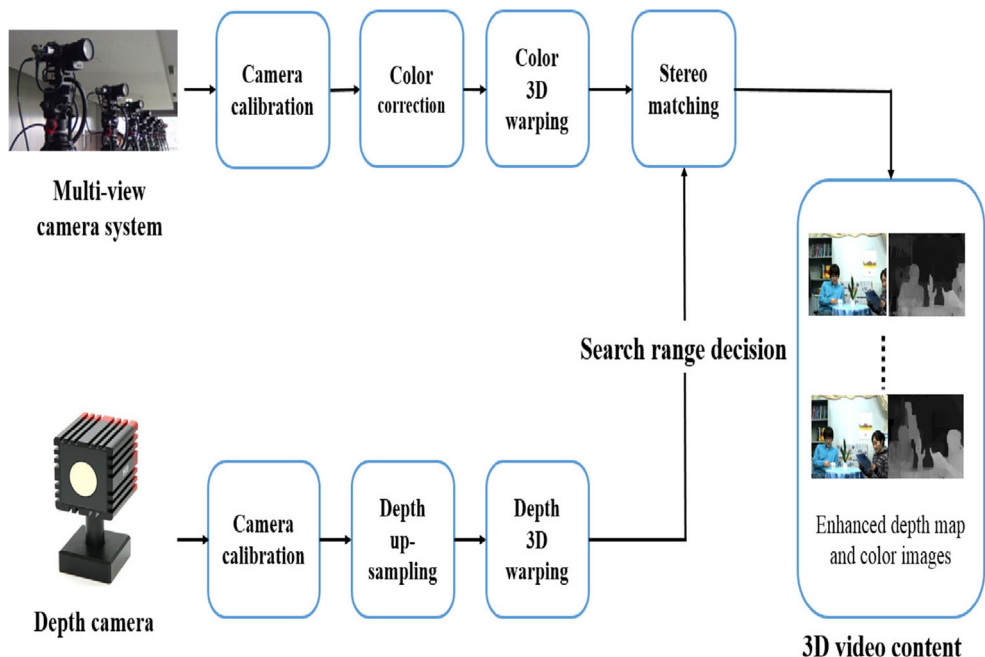
Captured depth images need up-sampling and 3D warping procedure, due to inconsistent resolution to color cameras and camera position differences, respectively [6]. Sequentially, the warped depth information reduce the disparity search range for stereo matching. While such a constrained disparity range

---

✉ Ji-Hun Mun  
jhm@gist.ac.kr  
Yo-Sung Ho  
hoyo@gist.ac.kr

<sup>1</sup> Gwangju Institute of Science and Technology (GIST), 123  
Cheomdan-gwagi-ro, Buk-gu, Gwangju 61005, Republic of Korea

**Figure 1** Hybrid multi-view camera system.



enhances the accuracy stereo matching, synthesized images still contain a noise and blurring effects. Such effects compromise the viewing quality [7].

Generally, stereo and multi-view camera systems are used to generate 3D video content. Multi-view camera video content supports a reality and immersion without any assistance devices. The multi-view video content generates a many viewpoint images, therefore, we can select preferred viewpoint images.

Recently, research about the multi-viewpoint and free-viewpoint image technique have been executed by MPEG. They also develop 3D video coding techniques for broadcasting system. Since the multi-view video content have a lots of data sets, it cause a bandwidth overloading problem. In order to solve limited bandwidth and heavy data problems, effective coding technologies are developed [8]. Virtual view synthesis method is also used for solving such issues.

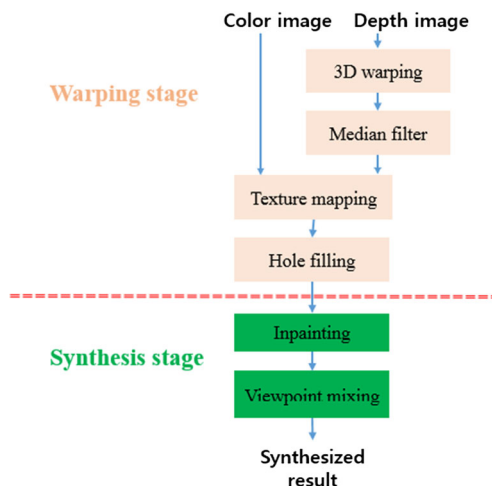


**Figure 2** Hole appearance on left and right synthesized images.

MPEG develops the VSRS (View Synthesis Reference Software) to create a generalized virtual viewpoint images [9]. Virtual viewpoint images are generated using a depth information. The depth information is not entirely correct with the real world coordination, thus synthesized image includes a many hole region as represented in Fig. 2. In this paper, we propose an effective hole filling and noise removal method to enhance the synthesized image quality.

### 2 3D Video System

Generally, the 2D video system do not requires a depth information when generating video content. However,



**Figure 3** Virtual view synthesis algorithm.



Figure 4 Noise artifacts in synthesized images.

the 3D video system uses the depth information to support a reality and immersion in 3D video content. Depth images can be acquired by depth cameras such as ToF and Kinect cameras. Moreover, they can be estimated via stereo matching using captured color images from multi-view or stereoscopic camera systems.

Two input data which color and depth images are usually used to generate 3D video content. Camera geometrical error and parameter values effect on the generated image, thus image rectification and color correction have to be performed before generating video content.

Generated video content images are differently displayed depending on the types of screen devices [10]. If display devices support a 3D video image, then we can feel the reality and immersion. However, increased number of cameras for multi-view systems cause a spatial and data overloading problem. Generally, virtual viewpoint synthesis method is used to solve that kind of problems. Virtual viewpoint images are generated by three steps: 3D warping, boundary noise removal and hole filling.

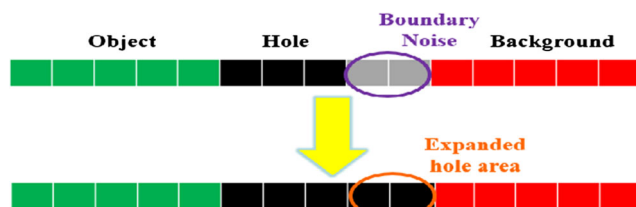


Figure 5 Hole expansion for boundary noise removing.

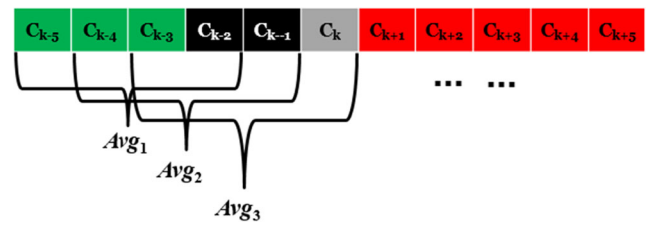


Figure 6 Boundary noise handling method.

### 3 Virtual View Synthesis

In VSRS, stereo color image and depth image are used for an input data to generate a virtual viewpoint images. Figure 3 represent the flow chart of view synthesis procedure. The view synthesis method can be separated into two stage.

In the warping stage, we perform depth image 3D warping to change the image coordination to 3D space. The depth camera coordination and virtual world coordination are not the same, hence on the warped depth image includes some hole region. VSRS uses the median filter at the warped depth image to remove the hole region.

When performing the color image 3D warping (texture mapping), the filtered depth image is used for guide information. Due to the hole region in depth image, the warped color image also contains a hole region. The hole area has a negative effect during the view synthesis, since it has to be clearly removed. For a remaining hole region, VSRS uses the inpainting method. The viewpoint mixing method selects a relevant view point image to create an accurate synthesized image.

#### 3.1 Boudnary Noise Handling

Conventional hole filling methods did not consider a boundary noise effect, which occurred near the object neighbor region. The boundary noise affects the quality of synthesized images. As indicated in Fig. 4, the noise artifacts are emerged in synthesized images.

Many kinds of boundary noise are generated during the color image 3D warping. Since the boundary noise influence the hole filling result, it has to be eliminated

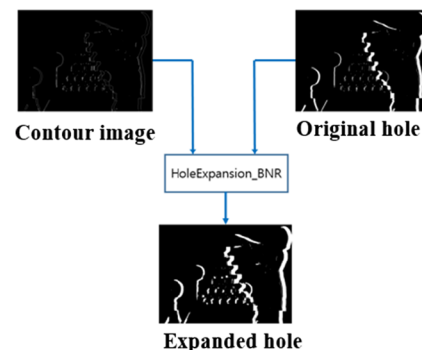
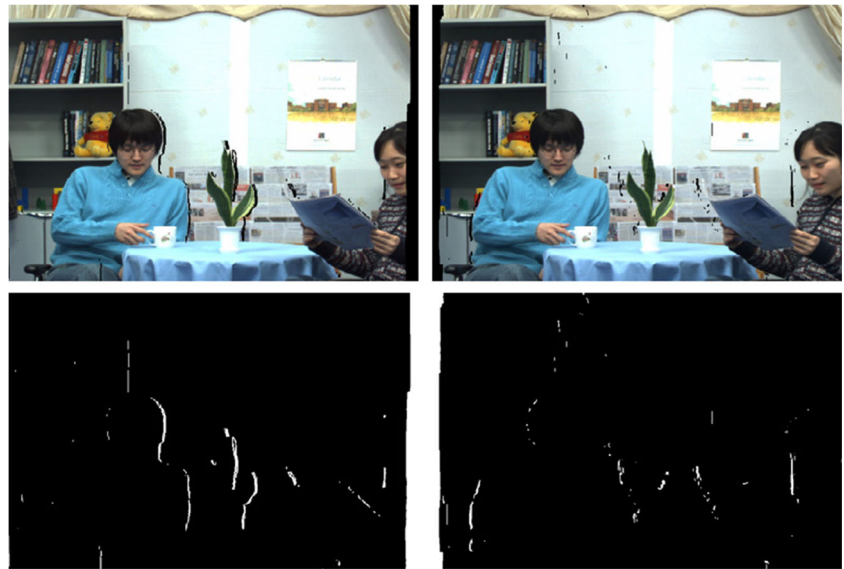


Figure 7 Input and output date for boundary noise handling.

**Figure 8** Left and right image 3D warping results.



from the warped color image before performing the hole filling.

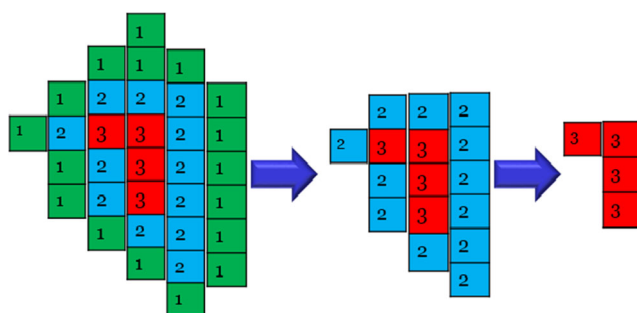
In this paper, we propose boundary noise removing method as represented in Fig. 5. Since the boundary noise occurred near the hole area, we compare the boundary noise region with noise neighbor area. By replacing the boundary noise area with a hole regions, we can efficiently remove the noise region. Figure 6 represents the proposed boundary noise removing method.

In order to find out the boundary noise region in warped color image, we compute the mean value of boundary noise neighbor pixels. Since the boundary noise pixels have a different value compared to neighbor pixel values, the proposed boundary handling method correctly extract the boundary noise area.

$$Avg_p = \sum_{i=0}^3 \frac{C_k + T_i}{4}, \quad (T = -5, -4, \dots, 4, 5) \quad (1)$$

$(p = 1, 2 \dots 8)$

$$Abd_i = |Avg(P_i) - Avg(P_{i+1})|, \quad (i = 1, 2, 3 \dots 7) \quad (2)$$



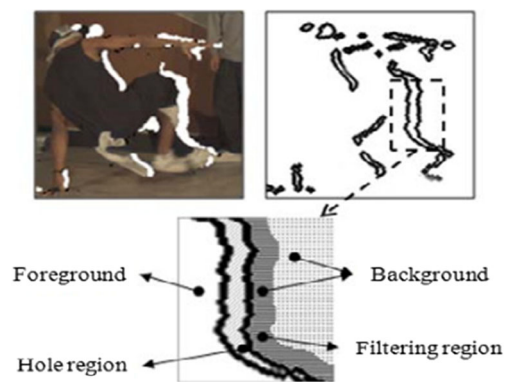
**Figure 9** Hole filling procedure.

In Eq. (1), the denominator represents the number of comparison pixels and  $C_k$  indicates the target boundary noise pixels. The computation of the average value of all the candidates is an essential procedure to change the boundary noise area to a hole region. After all candidates average values are calculated, we compute the absolute difference of each neighbor average values using Eq. (2). If the absolute difference value is bigger than the pre-determined threshold value, then  $C_k$  changes into a hole pixel.

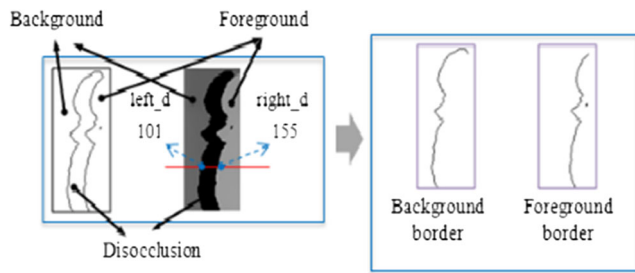
As a result of the proposed boundary noise handling method, the original hole area is expanded as represented in Fig. 7. Boundary noise contour image and original hole information are used as an input data to obtain the expanded hole image. Using the proposed boundary handling method, we can get a non-boundary noise image before performing the hole filling.

### 3.2 Conventional Hole Filling Methods

After the depth image 3D warping, VSRS uses median filter to remove the remaining hole area as represented in Fig. 3. Based



**Figure 10** Extracted hole contour lines from warped image.



**Figure 11** Comparison of depth value.

on the warped depth information, we also perform color image 3D warping. Since the real world coordination and warped depth information is not matched, the warped color images contain a hole region as represented in Fig. 8.

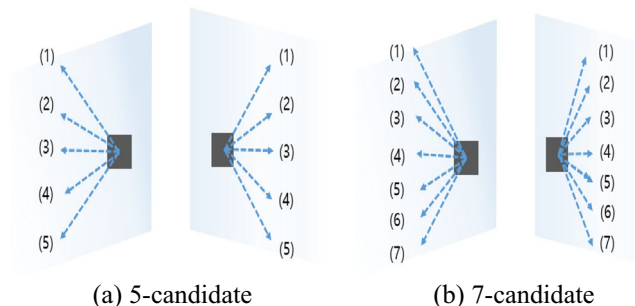
If we apply the median filter to the warped color image, then synthesized image has a noise and blurring effects due to the hole region. Consequently, the hole regions have to be clearly removed before view synthesis.

Many hole filling methods have been developed to remove the hole region in warped color image [11]. Majority of effective hole filling methods use a depth information and hole neighbor pixels to remove the hole region. In a conventional hole filling methods, the hole area  $H$  is filled with a constant color value  $C$ . The constant color value is determined by averaging the hole neighbor pixel values. That relation is represented in Eq. (3). Where  $\partial H$  indicates the hole neighbor pixels.

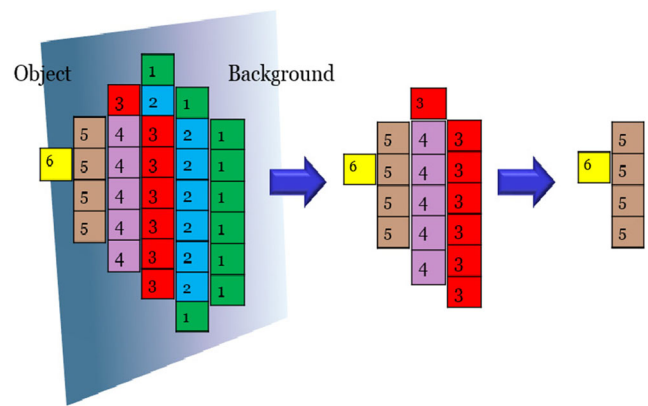
$$C = \frac{\sum_{d \forall m \in \partial H^1} [m]}{\sum_{\forall m \in \partial H^1}} \quad (3)$$

The conventional method [11] do not considers the hole filling directions and other hole neighbor pixel conditions, thus the quality of hole filling result is not extremely improved than VSRS result. Figure 9 demonstrates the conventional hole filling procedure.

In conventional hole filling method, the hole neighbor pixel values are used for hole filling pixel value. As shown in Fig. 9, the hole area is gradually removed using neighbor pixel values from the outside to inside. Since the conventional



**Figure 12** Candidates after left and right image warping.



**Figure 13** Hole filling with background consideration.

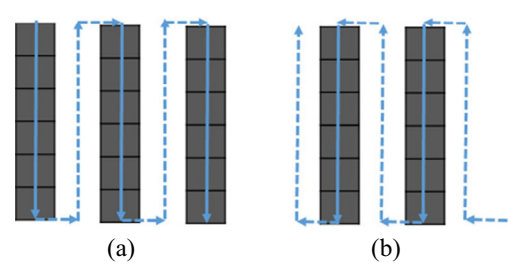
method do not efficiently fill the hole region, result image still includes a blurring and noise effects.

An advanced hole filling method was proposed using a texture background pixel value to eliminate the noise artifacts [12]. Hole regions usually occur in the background region, hence the hole regions are efficiently removed using the background pixel values. In order to find the hole region, the hole contour line is extracted from warped color image as represented in Fig. 10.

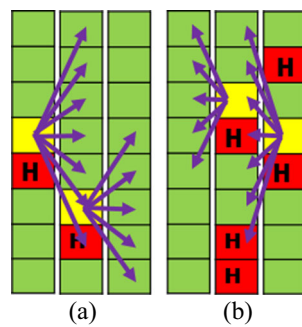
Based on the extracted hole region, the background region is determined using a depth information. As shown in Fig. 11, by comparing the depth values, the background and foreground region are distinguished.

If the left depth value is ‘101’ and the right depth value is ‘155’, then we can determine that the left region is background area. This method can be easily applied to standard hole filling procedure such as VSRS. Consequently, this algorithm improved the quality of synthesized image compared to the other conventional hole filling methods.

Even though the conventional method considers the background region pixel for hole filling, the result image quality is not significantly improved compared to original hole filling method. The warped depth image information is not perfectly correspond to real world coordination, thus the background considering method has a problem in terms of effective hole filling.



**Figure 14** Adaptive hole scanning method. **a** Hole left side is background **b** Hole right side is background



**Figure 15** Comparison of hole and substitutive pixels. **a** Left image warping hole filling candidates **b** Right image warping hole filling candidates

### 3.3 Proposed Hole Filling Method

Previously used hole filling methods create an insufficient hole filling results. In order to improve the quality of hole filling result image, we propose a new hole filling method considering a directionality of hole neighbor pixels. Usually the hole regions appear near the object. If the hole region is filled with a background pixel value, then result becomes

more reasonable. Since the hole areas do not occur within the object, background labeling is crucial in hole filling procedure.

The proposed hole filling method considers the directionality of hole neighbor pixels. Since the hole region emerges near the object area, we fill the hole region with a specific background pixel value. Left and right image warping results contain a different hole region depending on the viewpoint as shown in Fig. 8. The hole regions occur at the opposite side of input viewpoint in warped image. Furthermore, the occlusion region generates a hole region in warped color images.

Based on the depth information, we can easily determine the background area which correlates to the hole region. Additionally, we divide the hole appearance types as represented in Fig. 12. Depending on the location of background region, each side of hole candidate is adaptively applied to hole filling. Usually, the 7-candidate hole filling method can select more proper hole filling pixel value than the 5-candidate method.

Hole filling ordering is determined by location of background area. As demonstrated in Fig. 13, hole region have to be filled from the background region to object region.



**Figure 16** View synthesis result using different hole filling method.

Even though the proposed method considers the directionality of background pixels, the empty pixel values (hole) can be used as a hole filling candidate pixels. We propose an adaptive hole scanning method to solve this problem.

Hole scanning directions are differently applied depending on the left and right image warping results as represented in Fig. 14. Using this adaptive hole scanning method, the empty pixel values are not used for candidate pixels.

In Fig. 14a, it represents the situation which a background is located in hole left side. Similarly, hole scanning departures from right to left side in case of Fig. 14b. Since the right side information is insufficient when performing the left image warping, the right side of warped image contains a hole region. Likewise, the right image warping result also has a hole region on the left side.

The adaptive hole scanning method fills the hole region using different directions, thus we can generate an improved hole filling result when performing the hole filling with multi-directional hole filling candidate method. Throughout the proposed hole scanning method, we can use the non-empty pixels as a hole filling candidate pixel values.

In order to determine useful pixel value among the hole filling candidate pixels, we compare the hole pixel with

candidate pixels. However, the hole pixels do not contain any value, so we have to use other pixel value instead of empty pixel value. Since the hole neighbor pixel values contain a similar pixel value with a hole pixel, we use a hole upper side pixel value as a substitutive pixel. Using a substitutive pixel value, we can accurately select a hole filling pixel value. Consequently, we compare the substitutive pixel value with candidate pixel values as represented in Fig. 15.

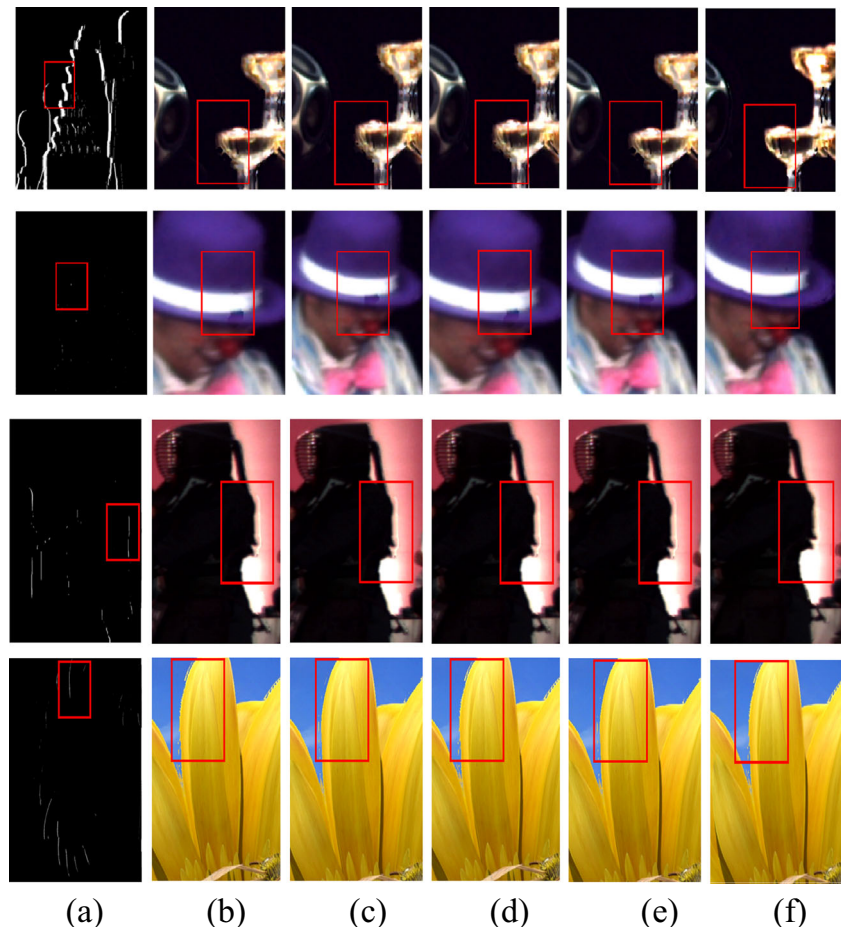
When performing the hole filling for warped left image as shown in Fig. 15a, we use right to left hole scanning direction. Similarly, in case of Fig. 15b, left to right hole scanning direction is used. Among the hole filling candidate pixels, we select a proper pixel values by comparing the substitutive pixel with candidate pixels. Equation (4) demonstrates the cost computation method for hole pixel value.

$$Cost = \frac{\sum_{x,y \in Neighbor} |NP(x,y) - CP(x,y)|}{NeighborNumber} \tag{4}$$

$$Direction = \text{argMin}\{Cost\} \tag{5}$$

In Eq. (4), *NeighborNumber* indicates the number of hole filling candidate pixels, thus the denominator value can be

**Figure 17** Enlarged test result for comparison. **a** Hole area **b** Original **c** Vazquez et al. **d** Cheon et al. **e** 5-candidate **f** 7-candidate



changed depending on the 5 and 7 candidate methods. Where  $NP$  represents the hole filling candidate pixels and  $CP$  indicates substitutive hole pixel value. Until the final hole is removed, we iteratively compute the cost value.

Subsequently, the average of differences between candidate and substitutive hole pixel is selected as a final cost value. The smallest cost value is used for a hole filling pixel value as indicted in Eq. (5). Consequently, the final hole filling pixel is determined among the hole filling candidate pixels which has a smallest cost value.

Using the proposed hole filling method, we can improve the synthesized image quality. Especially adaptive hole scanning method and multi-directional hole filling method are used to find a proper hole filling pixel value.

## 4 Experimental Results

To verify the multi-directional hole filling performance, we conduct the tests with 6 MPEG sequences; *Kendo*, *Bee*, *Pantomime*, *Champagne*, *Balloons* and *Soccer* (<http://www.jujii.nuee.nagoya-u.ac.jp/multiview-data>). Conventional hole filling methods for view synthesis [11, 12] and proposed methods are used to compare the synthesized image quality.

Multi-directional hole filling method shows a better objective results than other the conventional hole filling methods. Experiment results are represented in Fig. 16. In Fig. 17, the proposed methods and conventional methods result are indicated, especially object boundary areas.

Since the performance of view synthesis method is usually measured using a peak signal-to-noise ratio (PSNR), we present the PSNR comparison results in Table 1. The VSRS 3.5 version, conventional methods and proposed methods are used for performance comparison.

The proposed 7-candidate method performs better than other tools in most cases. However in *Bee* sequence test, the 5-candidate method is more effective than the 7-candidate method. Since the *Bee* sequence includes similar texture areas near the hole region compared to other test sequences, the similar pixel values are used when computing the cost value.

In Table 2, the average PSNR results of entire hole filling methods are indicated. As shown in Table 2, the proposed 7-candidate method has a better PSNR result than other hole filling methods. The conventional methods [11, 12] use the background pixel value and depth information for hole filling, thus they can efficiently fill the hole region compared to the VSRS method.

## 5 Conclusion

In this paper, we proposed an efficient hole filling method for virtual view synthesis. Boundary noise pixels near hole region

**Table 1** Objective evaluation of hole filling methods in terms of PSNR.

Sequence	Method	PSNR (dB)	Rating
Kendo	VSRS 3.5	35.3734	5th
	Vazquez et al.	35.7553	2nd
	Cheon et al.	35.3745	4th
	5-Candidate	35.5946	3rd
	7-Candidate	36.3960	1st
Bee	VSRS 3.5	29.9514	5th
	Vazquez et al.	29.9557	4th
	Cheon et al.	29.9610	3rd
	5-Candidate	30.4727	1st
	7-Candidate	30.2324	2nd
Pantomime	VSRS 3.5	38.7064	5th
	Vazquez et al.	39.0226	2nd
	Cheon et al.	38.8438	4th
	5-Candidate	37.9722	3rd
	7-Candidate	39.4908	1st
Champagne tower	VSRS 3.5	34.2504	5th
	Vazquez et al.	34.9841	4th
	Cheon et al.	35.1094	3rd
	5-Candidate	35.3167	2nd
	7-Candidate	35.8466	1st
Soccer	VSRS 3.5	34.2504	5th
	Vazquez et al.	35.0094	3rd
	Cheon et al.	35.1076	4th
	5-Candidate	35.4471	2nd
	7-Candidate	35.7437	1st

are expanded as a hole region. Based on the expanded hole region, we use the multi-directional hole filling method to correctly determine the hole filling pixel value. In addition, the hole scanning method is used to efficiently compare the hole filling pixel value within the non-empty pixels. Subsequently, we compute the cost value for each candidate pixels to find a minimum cost pixel value. Consequently, the hole filling pixel is properly selected from the background region. From view synthesis experiments, we confirmed that the proposed method showed a 1.03 dB average PSNR improvement over other conventional methods.

**Table 2** Average PSNR results.

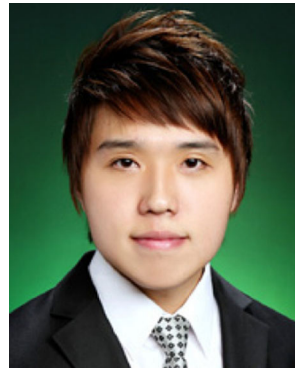
Method	Avg. PSNR (dB)	Rating
VSRS 3.5	34.5065	5th
Vazquez et al.	34.8474	3rd
Cheon et al.	34.8792	4th
5-candidate	35.2206	2nd
7-candidate	35.5419	1st



**Acknowledgments** This research was supported by the ‘Cross-Ministry Giga KOREA Project’ of the Ministry of Science, ICT and Future Planning, Republic of Korea (ROK). [GK15C0100, Development of Interactive and Realistic Massive Giga-Content Technology]

## References

1. Redert, A., Op de Beeck, M., Fehn, C., IJsselsteijn, W., Pollefeys, M., Gool, L., Ofek, E., Sexton, I., & Surman, P. (2002). ATTEST: advanced three-dimensional television system technologies. 3D data processing visualization and transmission. doi: [10.1109/TDPVT.2002.1024077](https://doi.org/10.1109/TDPVT.2002.1024077).
2. Fehn, C., Kauff, P., Op de Beeck, M., Ernst, F., IJsselsteijn, W., Pollefeys, M., Van Gool, L., Ofek, E., & Sexton, I. (2002). An evolutionary and optimised approach on 3D-TV. International Broadcast Conference, pp. 357–365.
3. Tanimoto, M. (2006). FTV (Free Viewpoint Television) for 3D scene reproduction and creation. Proc. of IEEE Computer Vision and Pattern Recognition Workshop (CVPRW), p. 172.
4. Wilburn, B., Joshi, N., Vaish, V., Talvala, V., Antunez, E., Barth, A., Adams, A., Horowitz, M., & Levoy, M. (2005). High performance imaging using large camera arrays. *ACM Transactions on Graphics*, 24(3), 765–776.
5. Cheng, H., An, P., Li, H., & Zhang, Z. (2011). Stereo image rectification algorithm for multi-view 3D display. International conference on 3D Imaging (IC3D), pp. 1–5.
6. Plath, N., Knorr, S., Goldmann, L., & Sikora, T. (2013). Adaptive image warping for hole prevention in 3D view synthesis. *IEEE Transactions on Image Processing*, 22(9), 3420–3432.
7. Macchiavello, B., Dorea, C., Hung, E.M., & Cheung, G. (2013). Saliency-cognizant robust view synthesis in free viewpoint video streaming. IEEE International Conference on Image Processing (ICIP), pp. 1603–1607.
8. Wiegand, T., Sullivan, G. J., Bjontegaard, G., & Luthra, A. (2003). Overview of the H.264/AVC video coding standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(7), 560–576.
9. MPEG. (2009). Introduction to 3D video. ISO/IEC JTC1/SC29/WG11 document N 10357.
10. Yuan, H., Calic, J., & Kondo, A. (2014). Facilitating interaction with stereoscopic 3D display devices. 3DTV-Conference: The True Vision –Capture, Transmission and display of 3D Video (3DTV-CON), pp. 1–4.
11. Vazquez, C., Tam, W. J., & Spemaza, F. (2006). Stereoscopic imaging: filling disoccluded areas in depth image-based rendering. Proc. SPIE, vol. 6392. doi: [10.1117/12.685047](https://doi.org/10.1117/12.685047).
12. Lee, C., & Ho, Y.S. (2009). View synthesis using depth map for 3D video. Asia-Pacific Signal and International Processing Association (APSIPA), pp. 1–8.



**Ji-Hun Mun** B.S. : degree in Electronic engineering from Chonbuk National University, Korea, in 2013. M.S. : degree in Information and Communication Engineering from Gwangju Institute of Science and Technology (GIST), Korea, in 2015. Ph.D. Student in the School of Information and Communication at GIST, Korea. His research interests are digital image, video coding, H.265/HEVC, Real-time stereo matching, and signal processing.



**Yo-Sung Ho** B.S. : degree in electronic engineering from Seoul National University, Seoul, Korea, in 1981. M.S. : degree in electronic engineering from Seoul National University, Seoul, Korea, in 1983. Ph.D. : degree in electrical and computer engineering from the University of California, Santa Barbara, in 1990. Electronics and Telecommunications Research Institute (ETRI), Daejeon, Korea, in 1983. Philips Laboratories, Briarcliff Manor, NY, From 1990 to 1993. Where he development of the advanced digital high-definition television system. In 1993, he rejoined the Technical Staff of ETRI and was involved in development of the Korea direct broadcast satellite digital television and high-definition television systems Since 1995, he has been with the Gwangju Institute of Science and Technology, Gwangju, Korea, where he is currently a Professor in the Department of Information and Communications. His research interest include digital image and video coding, image analysis and image restoration, advanced coding techniques, digital video and audio broadcasting, 3D television, and realistic broadcasting.