Appl. Math. Inf. Sci. 6 No. 1S pp. 195S-200S (2012)

Applied Mathematics & Information Sciences
An International Journal

Stereo Image Synthesis by View Morphing with Stereo Consistency

Seon-Min Rhee[1,2], Yeo-Jin Yoon[1], Il-Kyu Shin[1], Yong-Guk Kim[1], Yoo-Joo Choi[3], and Soo-Mi Choi[1*]

[1]Department of Computer Engineering, Sejong University, Seoul, Korea
[2]University of Southern California, LA, California, USA
[3]Korean-German Institute of Technology, Seoul, Korea
Email Address: smchoi@sejong.ac.kr

Received: Received Jun 18, 2010; Revised Jan 18, 2011; Accepted September 12, 2011
Published online: 1 January 2012

Abstract: We show how to synthesize stereo images by adjusting camera separation to produce customized stereopsis and by changing the viewpoint to accommodate a head-tracked 3D display. We propose an improved method for stereo image synthesis from a pair of wide-baseline stereo images and corresponding disparity maps. To give proper depth in hole regions caused by disocclusion, we introduce a hole filling method which maintains consistency between a pair of stereo images. The results from a user study demonstrate that our hole filling method gives a correct depth effect than existing hole filling.

Keywords: Stereo Images, View Morphing, Disparity Maps

1 Introduction

The increasing availability of 3D displays, such as 3D monitors, 3D TVs and 3D cinema, is raising consumer demand for stereo content [1-3]. Most 3D displays use the difference in viewpoint between the viewer's eyes to produce a stereoscopic depth effect. Stereo content for these displays is produced using two cameras, and the distance between them is an important parameter in stereopsis [4]. As the camera separation increases, the depth effect increases; but a too great separation can produce dizziness or headache. When the camera separation is equal to the distance between their own eyes, a viewer feels comfortable and perceives depth accurately.

Since the separation of stereo cameras is fixed, stereo images should be modified for individual viewers. In addition, if a viewer is using a head-tracked 3D display which tracks the position of the viewer's head, the viewpoint of the stereo images needs to be changed to match the viewer's location. We call this process stereo image synthesis. It can be achieved by view morphing [5], which is widely used to synthesize an image at an arbitrary view point from two original images with dense point correspondences. In our previous work, Rhee et al. [6] proposed a method to generate an intermediate stereoscopic view using a wide-baseline stereo camera. The method was able to synthesize a pair of stereo images which has an arbitrary viewpoint and a binocular parallax. However, the pixels in holes caused by disocclusion were not recovered correctly.

* **Corresponding Author: Soo-Mi Choi, smchoi@sejong.ac.kr**

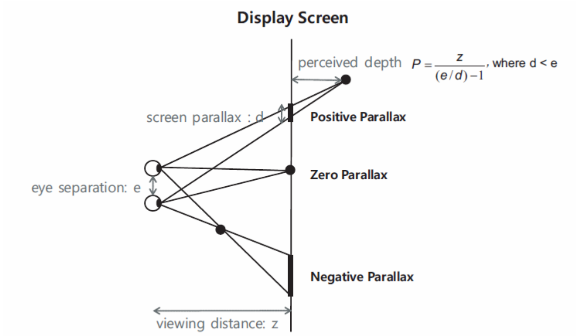


Fig.1 Principle of depth perception using screen parallax

We also shortly discussed the potential of stereo image synthesis for an auto-stereoscopic 3D display [7].

Using two images which have a binocular disparity, we can generate a screen parallax that gives us the depth perception relative to a reference plane (i.e. a display screen), as illustrated in Fig.1. In the case of positive parallax in which a point for the right eye is located to the right of the left eye's location, the point is perceived to be behind the display screen. In the case of negative parallax in which a point for the right eye is located to the left of the left eye's location, our brain interprets the point as being in front of the display screen, as if floating in space. Lastly, in the case of zero parallax a point is perceived at the surface of the display screen.

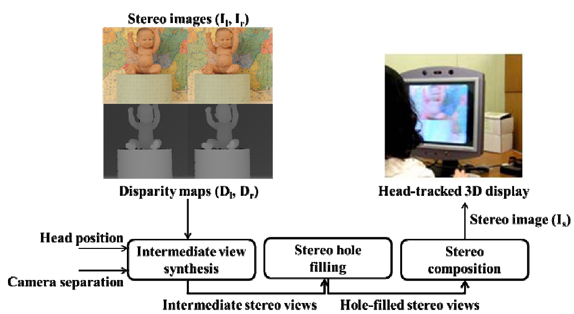


Fig.2 Overall process of the proposed stereo image synthesis

In this study, we propose an improved method for stereo image synthesis from a pair of original wide-baseline stereo images and corresponding two disparity maps of the target scene. Fig.2 shows an overview of our approach to stereo image synthesis for a head-tracked 3D display. A user's head position is estimated in real-time using a stereo camera attached on the monitor and it is used to generate intermediate stereo views. Each pixel in holes, caused by partial or complete disocclusion, is recovered by considering stereo consistency.

Previously a similar approach [8] was proposed in a sense that it synthesizes a virtual view using a pair of stereo images and their disparity maps. The difference from our method is that it generates a single image rather than stereo images. Even though two virtual views could be separately synthesized by the method, it cannot provide 3D depth effect in hole regions. The main contribution of our work is that it synthesizes a pair of stereo images with accurate depth effect in hole regions, allowing a user to control depth effect and the viewpoint.

2 Generating intermediate stereo views and hole filling

We now describe how we synthesize stereo images by view morphing to produce an arbitrary camera separation; and how we determine the disparity and the pixel value of each pixel in hole parts with stereo consistency to give a proper depth effect.

2.1 Intermediate stereo view synthesis using disparity maps

We create stereo images with a new camera separation by first determining an intermediate virtual camera separation. Fig.3 illustrates how a virtual central camera position $\mathbf{C}_i$ is obtained from the images $\mathbf{I}_0$ and $\mathbf{I}_1$, acquired from the cameras positioned at $\mathbf{C}_0$ and $\mathbf{C}_1$ respectively.

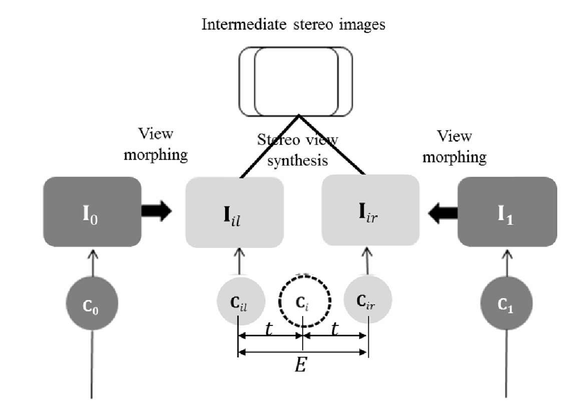


Fig.3 Virtual camera position for creating intermediate stereo images

We can then create the new camera positions $\mathbf{C}_{il}$ and $\mathbf{C}_{ir}$, each of which is offset by t from $\mathbf{C}_i$. We locate $\mathbf{C}_0$ at the origin of a world coordinate system. Thus the camera positions for the stereo images are $\mathbf{C}_0$= (0, 0, 0) and $\mathbf{C}_1$= ($\mathbf{C}_x$, 0, 0) assuming that two cameras are rectified [13]. And the virtual camera position $\mathbf{C}_i$ for the intermediate stereo image can be determined by linear interpolation between $\mathbf{C}_0$ and $\mathbf{C}_1$ as follows:

$$\mathbf{C}_i = (1-s_i)\mathbf{C}_0 + s_i\mathbf{C}_1 = (s_i\mathbf{C}_x, 0, 0) \text{ for } 0 \le s_i \le 1. \quad (1)$$

Let s_{il} and s_{ir} be interpolation coefficients for $\mathbf{C}_{il}$ and $\mathbf{C}_{ir}$ respectively. If s_t is defined as the half value of the difference between s_{il} and s_{ir}, then the new camera positions $\mathbf{C}_{il}$ and $\mathbf{C}_{ir}$ can be defined by

$$\begin{aligned} \mathbf{C}_{il} &= \mathbf{C}_i - t\frac{\mathbf{C}_1 - \mathbf{C}_0}{\|\mathbf{C}_1 - \mathbf{C}_0\|} = ((s_i - s_t)\mathbf{C}_x, 0, 0), \\ \mathbf{C}_{ir} &= \mathbf{C}_i + t\frac{\mathbf{C}_1 - \mathbf{C}_0}{\|\mathbf{C}_1 - \mathbf{C}_0\|} = ((s_i + s_t)\mathbf{C}_x, 0, 0). \end{aligned} \quad (2)$$

Given the virtual camera separation $E(=2t)$, s_t can be obtained by:

$$\begin{aligned} E &= \|\mathbf{C}_{ir} - \mathbf{C}_{il}\| = \sqrt{(2s_t\mathbf{C}_x)^2}, \\ s_t &= \frac{E}{2\sqrt{\mathbf{C}_x^2}} = \frac{E}{2\|\mathbf{C}_1 - \mathbf{C}_0\|} = \frac{t}{\|\mathbf{C}_1 - \mathbf{C}_0\|}. \end{aligned} \quad (3)$$

The interpolation coefficients s_{il} and s_{ir} can be defined as follows:

$$s_{il} = s_i - s_t, \; s_{ir} = s_i + s_t \quad (4)$$

Equation (4) permits the virtual camera separation to be adjusted for synthesis. This equation also allows to generate the images $\mathbf{I}_{il}$ and $\mathbf{I}_{ir}$ from the virtual camera positions $\mathbf{C}_{il}$ and $\mathbf{C}_{ir}$ respectively.

Let a scene point with Euclidean coordinate (x, y, z) be expressed by $\mathbf{P}$ and a Euclidean image point (x, y) by $\mathbf{p}$. The points $\mathbf{p}_0$ and $\mathbf{p}_1$ are projections of a scene point $\mathbf{P}$ within the images $\mathbf{I}_0$ and $\mathbf{I}_1$ respectively. The projected coordinate of $\mathbf{P}$ within $\mathbf{I}_{il}$, the point $\mathbf{p}_{il}$, can be calculated by linear interpolation of $\mathbf{p}_0$ and $\mathbf{p}_1$. In our study, we assume that a stereo image is taken by two parallel cameras. In a similar way, the point $\mathbf{p}_{ir}$ can be obtained by Equation (5), s_{il} and s_{ir} are limited to the range 0 to 1.

$$\mathbf{p}_{il} = (1-s_{il})\mathbf{p}_0 + s_{il}\mathbf{p}_1, \; \mathbf{p}_{ir} = (1-s_{ir})\mathbf{p}_0 + s_{ir}\mathbf{p}_1. \quad (5)$$

We assume that a pair of dense disparity maps between $\mathbf{I}_0(x, y)$ and $\mathbf{I}_1(x, y)$ are given. To create the images $\mathbf{I}_{il}$ and $\mathbf{I}_{ir}$, the pixel value $\mathbf{I}_0(x, y)$ or $\mathbf{I}_1(x, y)$ can be mapped onto $\mathbf{I}_{il}(x', y')$ and $\mathbf{I}_{ir}(x', y')$ using forward mapping [9] as follows:

$$\begin{aligned} \mathbf{I}_0(x, y) &\to \begin{cases} \mathbf{I}_{il}((1-s_{il})x + s_{il}(x - \mathbf{D}_0(x, y)), y), \\ \mathbf{I}_{ir}((1-s_{ir})x + s_{ir}(x - \mathbf{D}_0(x, y)), y), \end{cases} \\ \mathbf{I}_1(x, y) &\to \begin{cases} \mathbf{I}_{il}((1-s_{il})(x + \mathbf{D}_1(x, y)) + s_{il}x, y), \\ \mathbf{I}_{ir}((1-s_{ir})(x + \mathbf{D}_1(x, y)) + s_{ir}x, y). \end{cases} \end{aligned} \quad (6)$$

Here $\mathbf{D}_0$ is the given disparity map for the image $\mathbf{I}_0$ indicating the disparity to the image $\mathbf{I}_1$, and $\mathbf{D}_1$ is the disparity map for $\mathbf{I}_1$ to $\mathbf{I}_0$.

2.2 Hole filling with stereo consistency

The forward mapping process described in the previous section can create holes where there is no data for part of the images $\mathbf{I}_{il}$ and $\mathbf{I}_{ir}$. As shown in Fig.4, disocclusion leads to holes because the originally hidden information is missing. Two types of holes appear on the destination images. Partial holes are not visible from the reference camera but visible from the secondary camera. Complete holes are not visible from any of two cameras. Two methods are commonly used to fill holes: the first is to copy pixel values from around the hole into the missing pixels [5], [10]; the second is the inpainting technique [11]. Previous work has focused on filling partial holes, and complete holes are filled in by an interpolation technique within one image without considering the other image of a stereo image pair. However, these complete holes cannot be disregarded in complex scenes and should be handled by considering stereo consistency.

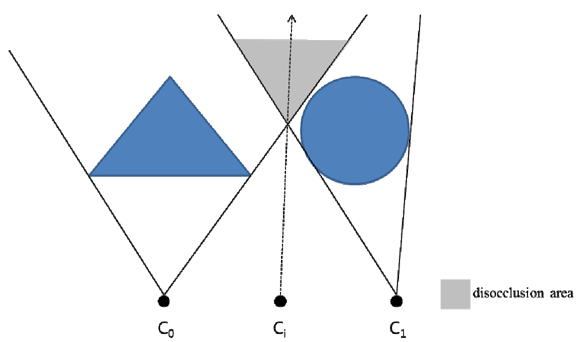


Fig.4 The disocclusion area produced by a virtual camera movement

We propose a new method for filling complete holes. The pixels within the holes are filled using an existing method for one of the stereo image pair; and then the corresponding pixels in the other image are mapped from the first image in such a way that there is stereo consistency between a pair of stereo images. In our method, a hole is assumed as a continuous planar area with a left or right neighbor. We compare two disparities d_l and d_r for the left and right neighbor pixels of the horizontal scan line of the hole. If one disparity is smaller than the other, we can assume that the area is farther than the other. Therefore, we take smaller disparity d_h as the disparity of the hole. If we assume that complete holes in the left image $\mathbf{I}_{il}$ are filled by a function $f(x, y)$, which can be any hole filling method such as interpolation of the neighborhood pixels, then complete holes in the right image $\mathbf{I}_{ir}$ can be filled with the same value in the hole apart from d_h in the left image. After that, the region remained as a hole in $\mathbf{I}_{ir}$ can be filled with $f(x, y)$ as follows:

$$\mathbf{I}_{ir}(x, y) = \begin{cases} \mathbf{I}_{il}(x + d_h, y) & \text{if } \mathbf{I}_{il}(x + d_h, y) \in \text{ hole area,} \\ f(x, y) & \text{otherwise.} \end{cases} \quad (7)$$

Since we can generate $\mathbf{I}_{il}$ and $\mathbf{I}_{ir}$ both from $\mathbf{I}_0$ and $\mathbf{I}_1$ as shown in Equation (7), the partial holes can be filled by referring to each other during the intermediate stereo view synthesis. Finally, the left and right images are composited using a D4D format (vertically interlaced) to be displayed on an auto-stereoscopic 3D display. The pixels in this display are divided into odd and even columns, which are directed to the left and right eye respectively. We adapted the OpenGL stencil mask to present stereo images in alternate columns.

3 Results and discussion

3.1 Results of stereo image synthesis

To test the performance of our stereo image synthesis, we used stereo images from Middlebury Institute [12], and the stereo images are shown in the form of anaglyph. Fig.5 shows the results of stereo image synthesis with a camera separation of 160mm. The adjusted camera separations are 130, 100, 70, and 40mm, respectively, and the position of the virtual stereo camera is (80, 0, 0). The Fig.6 shows the enlarged images of the rectangle regions in Fig.5.

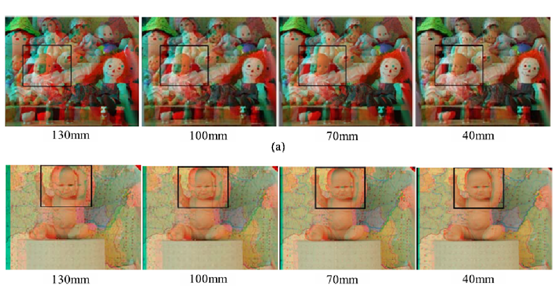


Fig.5 Stereo image synthesis with different camera separations

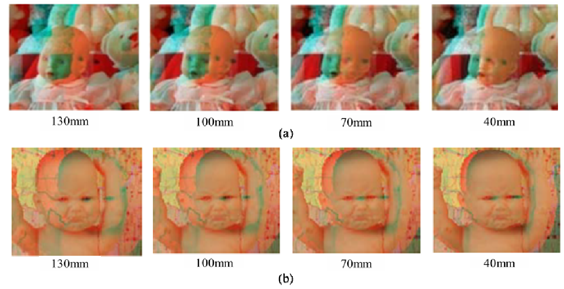


Fig.6 Enlarged images of the rectangle regions in Fig.5

Fig.7 shows four stereo images generated by positioning the camera $\mathbf{C}_i$ at (30, 0, 0), (60, 0, 0), (90, 0, 0), and (120, 0, 0), with a camera separation of 60mm. A reference line is drawn within each image to assist comparison. In Fig.7(a), we see that the nose of the doll shifts to the left, and in Fig.7(b), the box shifts to the left with the reference line.

To test the accuracy of the proposed method, we used 3D graphic models and compared the 3D rendered result from a specific intermediate viewpoint with the synthesized result by our method. In Fig.8, the synthesized stereo images are the result of our method and the right OpenGL rendered stereo images are the 3D rendered result. The synthesized stereo images are almost identical to the OpenGL rendered stereo images (ground truth) with the same camera separation. The PSNR (peak signal-to-noise ratio) is about 36dB.

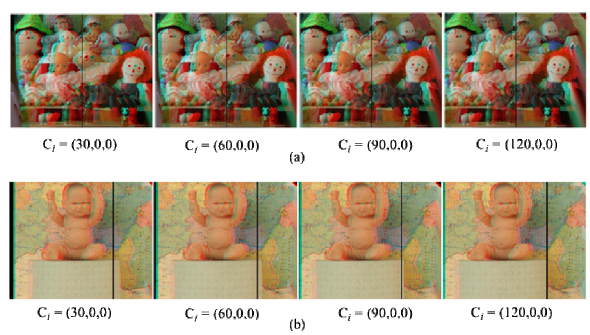


Fig.7 Stereo image synthesis with different camera positions

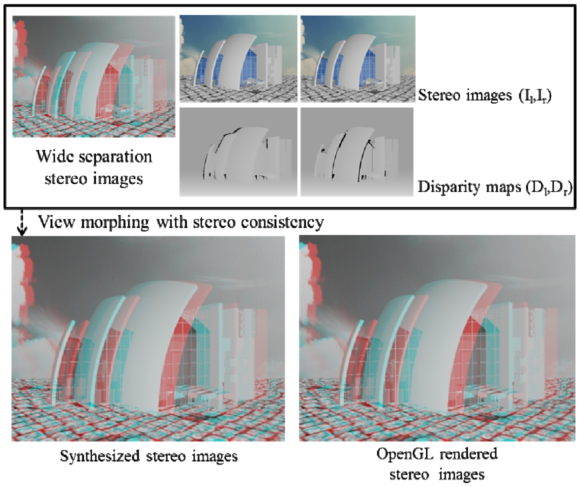


Fig.8 Comparison of synthesized stereo images and OpenGL rendered stereo images

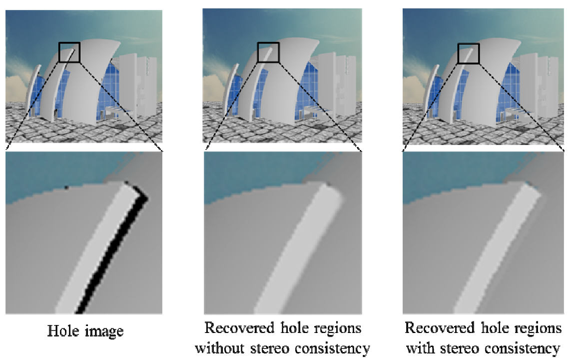


Fig.9 Comparison of stereo image synthesis with and without stereo consistency

In Fig.9, we compared the results of stereo image synthesis with and without stereo consistency, focusing on hole regions. We can see the recovered hole regions without stereo consistency are blurred.

3.2 User study

A total of 38 subjects, all in their 20s, participated in this user test. Among them, 13 had used a stereo display before. Each subject was shown the two kinds of stereo images and asked about their experience of the 3D depth effect. Camera separations of 10mm and 70mm were used. In assessing the 3D depth effect, 33 subjects out of 38 reported a difference in perceived depth between the two types of image and 32 out of these 33 said that the depth effect was greater for the 70mm camera separation than for the 10mm. Five subjects said that some difference in perceived depth was present but it was not significant. In the stereopsis test, 27 subjects among 38 observed a significant difference, and all subjects agreed that they perceived improved stereopsis with the 70mm camera separation than with 10mm. However, 11 subjects said that this difference was not significant.

To assess the methods of hole filling, the subjects were asked to compare the 3D depth of stereo images prepared by methods with and without stereo consistency. A majority of 33 subjects reported that they got a better depth effect from the stereo images synthesized by our method with stereo consistency, and the hole regions recovered without stereo consistency were smudged. Five of the subjects could not tell any differences. These results suggest that our hole filling gives a more consistent and correct depth effect than hole filling without stereo consistency.

4. Conclusion

We proposed a new method of synthesizing stereo images. When the camera separation is greater than the distance between the viewer's eyes, stereo images can be synthesized by changing either the camera separation or the viewpoint. This allows the viewer to experience an exact 3D depth effect, making viewing more comfortable. To generate more accurate stereo views, we proposed a new hole filling method using disparity information of a stereo pair. The main contribution of this paper is that it provides stereo consistency in hole regions of synthesized stereo images as well as controllable depth effect and the viewpoint. In this work, we assumed that a hole region is a continuous planar surface. We plan to provide more alternatives such as a curved surface or a linear interpolation of the neighbors as a future work.

Acknowledgements

This research was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (Nos. 2010-0003685 and 2010-0019373) and in part by the Korean Ministry of Knowledge Economy (NIPA-2011-C1820-1101-0004). The first author was partially supported by the Korea Research Foundation (No. 2007-357-D00235).

References

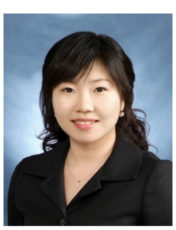
[1] Z L. Zhang and W.J. Tam, Stereoscopic image generation based on depth images for 3D TV, *IEEE Transactions on Broadcasting*. 51 (2005) 191.

[2] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic and M. Gross, Nonlinear Disparity Mapping for Stereoscopic 3D, *Proceedings of ACM Transactions on Graphics*. 29 (2010).

[3] M. Zwicker, A. Vetro, S. Yea, W. Matusic, H. Pfister and F. Durand, Resampling, Antialiasing, and Compression in Multiview 3-D Displays, *IEEE Signal Processing*

Magazine. 24 (2007) 88.
[4] M. Lambooij, W. Ijsselsteijn, M. Fortuin and I. Heynderickx, Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review, *Journal of Imaging Science and Technology*. 53 (2009).
[5] S.M. Seitz and C.R. Dyer, View Morphing, *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*. (1996) 21.
[6] S.M. Rhee and U. Neumann, Stereoscopic View Synthesis by View Morphing, *Proceedings of the 4th International Symposium on Visual Computing*. (2008) 924.
[7] S.M. Rhee, J.M. Choi and S.M. Choi, Accurate Stereo View Synthesis for an Auto-stereoscopic 3D Display, *International Conference on Consumer Electronics (ICCE)*. (2010) 375.
[8] 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, *Proceedings of the Picture Coding Symposium (PCS 2009)*. (2009) 1.
[9] D. Scharstein and R. Szeliski, A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms, *International Journal of Computer Vision*. 47 (2002) 7.
[10] A. Criminisi, A. Blake, C. Rother, J. Shotton and P.H.S. Torr, Efficient Dense Stereo with Occlusions for New View-Synthesis by Four-State Dynamic Programming, *International Journal of Computer Vision*. 71 (2007) 89.
[11] M. Bertalmio, G. Sapiro, V. Caselles and C. Ballester, Image Inpainting, *Proceedings of the 27th Annual Conference on Computer graphics and Interactive techniques*. (2000) 417.
[12] D. Scharstein and R. Szeliski, High-Accuracy Stereo Depth Maps Using Structured Light, *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR2003)*. 1 (2003) 195.
[13] A. Fusiello, E. Trucco and A. Verri, A Compact Algorithm for Rectification of Stereo Pairs, *Machine Vision and Applications*. 12 (2000) 16.

Seon-Min Rhee received her Ph.D. degree in computer science and engineering from Ewha Womans University at Seoul, Korea in 2007. Currently she is a research staff member in Samsung Advanced Institute of Technology. Her research interests include Mixed Reality, Computer Graphics, and 3D Vision. Contact her at seonminrhee@hotmail.com.



Yeo-Jin Yoon received her Master degree in computer engineering from Sejong University at Seoul, Korea in 2008. Currently she is a Ph.D. student at Sejong University. Her research interests include Computer Graphics and Virtual Reality. Contact her at yjyoon@gce.sejong.ac.kr.

Il-Kyu Shin received his Undergraduate degree in multimedia engineering from Hoseo College at Seoul, Korea in 2009. Currently he is a Master course student at Sejong University. His research interests include Computer Graphics, Virtual Reality, and Stereo Rendering. Contact him at wb_a@naver.com.

Yong-Guk Kim received his Ph.D. degree from Cambridge University in 1998. Currently he is an Associate Professor in Department of Computer Engineering in Sejong University, Seoul, Korea. His research interests include Human-Computer Interaction for Mobile devices, Computer Vision, and Visual Neuroscience. Contact him at ykim@sejong.ac.kr.

Yoo-Joo Choi received her Ph.D. degree in computer science from Ewha Womans University at Seoul, Korea in 2005. Currently she is an Assistant Professor in Korean-German Institute of Technology. Her research interests include Image Processing, Human-Computer Interaction and Smart Surveillance System. Contact her at yjchoi@kgit.ac.kr.

Soo-Mi Choi (corresponding author) received her Ph.D. degree in computer science and engineering from Ewha Womans University at Seoul, Korea in 2001. Currently she is an Associate Professor in Department of Computer Engineering in Sejong University, Seoul, Korea. Her research interests include Computer Graphics, Human-Computer Interaction, Virtual Reality, and Medical Imaging. Contact her at smchoi@sejong.ac.kr.