

Effect of fundamental depth resolution and cardboard effect to perceived depth resolution on multi-view display

Jae-Hyun Jung,¹ Jiwoon Yeom,¹ Jisoo Hong,¹ Keehoon Hong,¹ Sung-Wook Min,^{2,*} and Byoungho Lee,¹

¹School of Electrical Engineering, Seoul National University, Gwanak-Gu Gwanakro 599, Seoul 151-744, Korea

²Department of Information Display, Kyung Hee University, Dongdaemoon-Gu, Seoul 130-701, Korea

*mins@khu.ac.kr

Abstract: In three-dimensional television (3D TV) broadcasting, we find the effect of fundamental depth resolution and the cardboard effect to the perceived depth resolution on multi-view display is important. The observer distance and the specification of multi-view display quantize the expressible depth range, which affect the perception of depth resolution of the observer. In addition, the multi-view 3D TV needs the view synthesis process using depth image-based rendering which induces the cardboard effect from the relation among the stereo pickup, the multi-view synthesis and the multi-view display. In this paper, we analyze the fundamental depth resolution and the cardboard effect from the synthesis process in the multi-view 3D TV broadcasting. After the analysis, the numerical comparison and subjective tests with 20 participants are performed to find the effect of fundamental depth resolution and the cardboard effect to the perceived depth resolution.

©2011 Optical Society of America

OCIS codes: (110.2990) Image formation theory; (100.6890) Three-dimensional image processing.

References and links

1. A. Kubota, A. Smolic, M. Magnor, M. Tanimoto, T. Chen, and C. Zhang, "Multiview imaging and 3DTV," *IEEE Signal Process. Mag.* **24**(6), 10–21 (2007).
2. M. Okutomi and T. Kanade, "A multiple-baseline stereo," *IEEE Trans. Pattern Anal. Mach. Intell.* **15**(4), 353–363 (1993).
3. B. Lee, J.-H. Park, and S.-W. Min, *Digital Holography and Three-Dimensional Display*, T.-C. Poon, ed. (Springer US, 2006), Chap. 12.
4. Y. Kim, K. Hong, and B. Lee, "Recent researches based on integral imaging display method," *3D Research* **1**(1), 17–27 (2010).
5. D. Minoli, *3DTV Content Capture, Encoding and Transmission: Building the Transport Infrastructure for Commercial Services* (John Wiley and Sons, 2010), Chap. 3.
6. J.-H. Jung, J. Hong, G. Park, K. Hong, S.-W. Min, and B. Lee, "Evaluation of perceived depth resolution in multi-view three-dimensional display using depth image-based rendering," in *Proceedings of IEEE Conference on 3DTV Conference 2011* (Antalya, Turkey, 2011), pp. 1–4.
7. C. van Berkel and J. A. Clarke, "Characterisation and optimisation of 3D-LCD module design," *Proc. SPIE* **3012**, 179–186 (1997).
8. C. van Berkel, "Image preparation for 3D-LCD," *Proc. SPIE* **3639**, 84–91 (1999).
9. Y.-G. Lee and J. B. Ra, "New image multiplexing scheme for compensating lens mismatch and viewing zone shifts in three-dimensional lenticular displays," *Opt. Eng.* **48**(4), 044001 (2009).
10. H. Kim, J. Hahn, and H.-J. Choi, "Numerical investigation on the viewing angle of a lenticular three-dimensional display with a triplet lens array," *Appl. Opt.* **50**(11), 1534–1540 (2011).
11. J.-C. Liou and F.-H. Chen, "Design and fabrication of optical system for time-multiplex autostereoscopic display," *Opt. Express* **19**(12), 11007–11017 (2011).
12. C. Fehn, "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV," *Proc. SPIE* **5291**, 93–104 (2004).
13. K.-J. Oh, A. Vetro, and Y.-S. Ho, "Depth coding using a boundary reconstruction filter for 3-D video systems," *IEEE Trans. Circ. Syst. Video Tech.* **21**(3), 350–359 (2011).

14. Y. Zhao, C. Zhu, Z. Chen, D. Tian, and L. Yu, "Boundary artifact reduction in view synthesis of 3D video: from perspective of texture-depth alignment," *IEEE Trans. Broadcast* **57**(2), 510–522 (2011).
15. J.-H. Jung, K. Hong, G. Park, I. Chung, J.-H. Park, and B. Lee, "Reconstruction of three-dimensional occluded object using optical flow and triangular mesh reconstruction in integral imaging," *Opt. Express* **18**(25), 26373–26387 (2010).
16. M. Tanimoto, T. Fujii, and K. Suzuki, "View synthesis algorithm in view synthesis reference software 2.0 (VSR2.0)," ISO/IEC JTC1/SC29/WG11 Doc. M16090, Feb. 2009.
17. A. Woods, T. Docherty, and R. Koch, "Image distortions in stereoscopic video systems," *Proc. SPIE* **1915**, 36–48 (1993).
18. T. Koike, A. Yuuki, S. Uehara, K. Taira, G. Hamagishi, K. Izumi, T. Nomura, K. Mashitani, A. Miyazawa, T. Horikoshi, and H. Ujike, "Measurement of multi-view and integral photography displays based on sampling in ray space," in *Proceedings of IDW '08 Technical Digest* (Niigata Convention Center, Japan, 2008), pp. 1115–1118.
19. H. Yamanoue, M. Okui, and I. Yuyama, "A Study on the relationship between shooting conditions and cardboard effect of stereoscopic images," *IEEE Trans. Circuits Syst. Video Technol.* **10**(3), 411–416 (2000).
20. H. Yamanoue, M. Okui, and F. Okano, "Geometrical analysis of puppet-theater and cardboard effects in stereoscopic HDTV images," *IEEE Trans. Circuits Syst. Video Technol.* **16**(6), 744–752 (2006).
21. J. Cutting and P. Vishton, *Perception of Space and Motion*, W. Epstein, ed. (Academic Press, 1995), Chap. 3.
22. Philips (in Coop with 3D4YOU), "Response to New Call for 3DV Test Material: Beergarden," ISO/IEC JTC1/SC29/WG11 Doc. M16421, Apr. 2009.

1. Introduction

In recent years, three-dimensional television (3D TV) broadcasting environment has been constructed with the development of 3D display and digital broadcasting technology by many research groups, broadcasters and equipment manufacturers [1–5]. The system architecture of recently commercialized 3D TV broadcasting is composed of capturing the stereo images of 3D object or the single image with the depth map, transmitting the 3D contents with compression algorithm and displaying them in the commercialized 3D TV set [2–5].

Although the stereoscopic technique based 3D TV broadcasting is central to the mainstream technology, the autostereoscopic multi-view display will be developed as the next-generation 3D TV for overcoming the limitation of the number of views and the use of glasses [6–11]. For the compatibility between stereoscopic and multi-view 3D TV broadcasting, the contents format for multi-view display has to keep the stereo images and the additional depth map information. In multi-view 3D TV broadcasting, the view synthesis process is needed to generate the multi-view images from the stereo images and the depth map information. In the synthesizing process, the accuracy and quality of synthesized view image depend on the synthesizing algorithm and the depth resolution of depth map [12–16]. However, the depth map with high depth resolution needs the wide bandwidth in the transmission process, which leads to the high costs of all broadcasting systems.

The commercialized multi-view display for 3D TV broadcasting is mostly the slanted lenticular display technology, which has the limitation of expressible depth resolution fundamentally [17,18]. Therefore, the transmitted depth resolution of 3D TV will be limited by the fundamental depth resolution of slanted lenticular system. Even if the depth resolution of depth map in the transmitted contents format is higher than the fundamental depth resolution of the multi-view system, the information will be wasted and inexpressible.

In addition, the depth perception of human visual system (HVS) is decreased by the distance from observer to 3D object. The observer distance of the slanted lenticular multi-view display and the perceived depth resolution are both fixed, which affects to the depth resolution of 3D contents format. Additionally, the cardboard effect is one of the key factors to decrease the perceived depth resolution in the multi-view display [19–21].

From the fundamental depth resolution of multi-view display and the depth perception of HVS, we can assume the saturated value in perceived depth resolution exists in the multi-view 3D broadcasting. This research finds and analyzes the threshold of perceived depth resolution based on the technical factors from the specification of multi-view display and broadcasting process. The evaluation of saturation value of depth resolution in the perceived

depth resolution will provide the guideline for the manufacturer of the multi-view display and the 3D TV broadcasting systems.

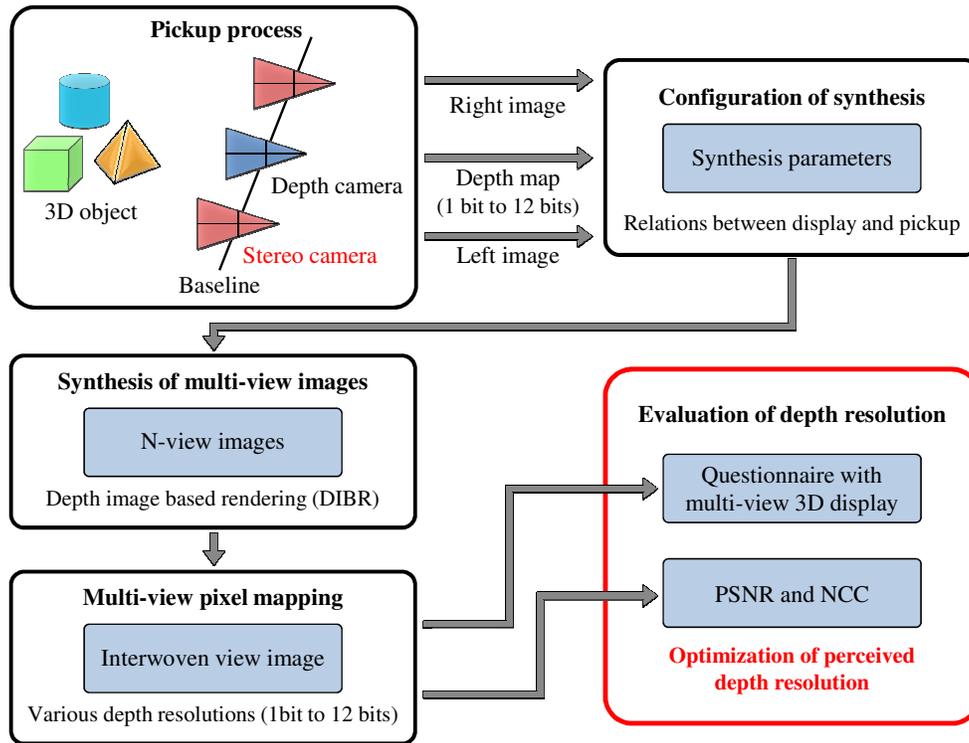


Fig. 1. Evaluation process of perceived depth resolution in multi-view display.

Figure 1 shows the detailed process for the evaluation of perceived depth resolution. First, we capture the stereo images and the depth map of 3D object with the variation of depth resolution from 1 bit to 12 bits. The 3D object and the stereo pickup specification are founded by the computational pickup scheme using OpenGL to easily change the parameters of pickup and depth information. To reduce the cardboard effect, we analyze the relation of parameters between the pickup and the synthesis. After capturing, the multi-view images are synthesized from the stereo images and the depth map with varying the depth resolution using depth image based rendering (DIBR) [12–16]. With varying the depth resolution, the synthesized multi-view images are compared with the ground truth view image in the peak signal-to-noise ratio (PSNR) and the normalized cross-correlation (NCC) to find the threshold of depth resolution numerically. In the interweaving process, the synthesized multi-view images with the different depth resolutions are mapped to the interwoven image for displaying the multi-view images on the slanted lenticular display [8]. The subjective test is performed with the reconstructed 3D images using the 9-view slanted lenticular display. The experimental results are presented and analyzed in this paper.

2. Principles of fundamental depth resolution of multi-view display and multi-view synthesis

A. Fundamental depth resolution from specification of slanted lenticular display

In HVS, the observer perceives the depth resolution sensitively nearby the reconstructed 3D object. The sensitivity of perception is inversely proportional to the distance from observer to 3D display. Therefore, the observer distance is the most important specification in the evaluation process. To evaluate the perceived depth resolution in HVS, the optimized

observer distance has to be decided by the multi-view display parameters. This paper analyzes the perceived depth resolution with the fixation of observer distance from the multi-view display specification.

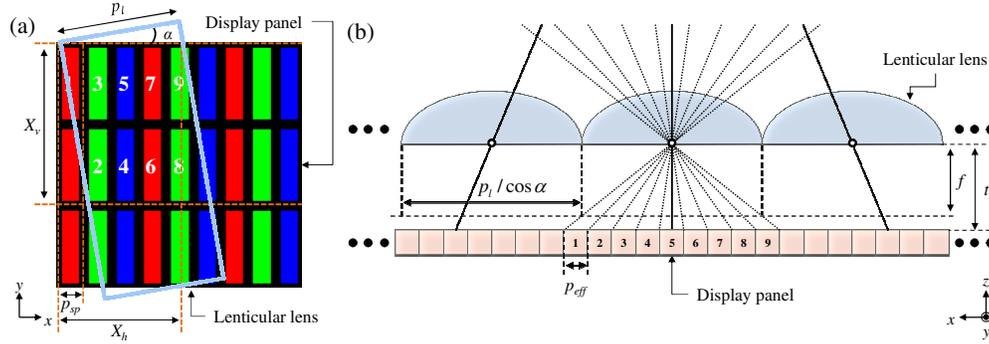


Fig. 2. Multi-view 3D display based on slanted lenticular lens: parameters of slanted lenticular system (a) in front view and (b) in upper view.

In these days, most of commercialized multi-view display systems adopt the slanted lenticular display which is proposed by Philips Research Laboratories in 1997 [7–11]. According to them, the number of views per the slanted lens in the horizontal direction X_h and the vertical direction X_v in N -view display using slanted lenticular lens array are given as

$$X_h = \frac{N}{X_v} = \frac{tp_l}{fp_{sp} \cos \alpha}, \quad (1)$$

where p_l is the lens pitch, f the focal length, p_{sp} the sub-pixel pitch, α the slant angle, and t the gap between lens and display panel. As shown in Fig. 2, the effective pixel pitch p_{eff} in the slanted lenticular display is the same as the half of sub-pixel pitch p_{sp} . The observer distance from lens array D is determined from the relation between the interocular distance d_e and the magnified pixel pitch of display. The number of views per the interocular distance k is determined to the integer values for showing the stereoscopic images at the whole multi-view positions. Therefore, the distance between each viewpoint g is defined as d_e/k .

The observer distance is derived from the ratio of g and p_{eff} , which also follows the ratio of slanted lens pitch and width of the N -view area pixels as shown in Eq. (2).

$$D = \frac{td_e}{kp_{eff}} = \frac{tp_l}{Np_{eff} \cos \alpha - p_l}. \quad (2)$$

From Eqs. (1) and (2), the lens pitch p_l is determined by the relation between interocular distance d_e and effective pixel pitch p_{eff} as follows:

$$p_l = \frac{Np_{eff}d_e \cos \alpha}{kp_{eff} + d_e}, \quad (3)$$

$$t = \frac{Nfp_{eff} \cos \alpha}{p_l}. \quad (4)$$

Equation (4) determines the gap between display panel and lenticular lens t from Eqs. (2) and (3). The observer distance D is determined from Eq. (2) with the calculated p_l and t from Eqs. (3) and (4). To evaluate the perceived depth resolution, the observer position is fixed at the observer distance from Eqs. (2) – (4).

From the specification of multi-view display, the expressible depth planes and range are limited between the near depth plane D_N and the far depth plane D_F [11]. In the general case of stereoscopic display or multi-view display, the feasible range of disparity is limited from 1% to 5% of the width of display resolution, which is more decreased and limited than D_N and D_F by the crosstalk because of the color dispersion, the lens distortion and the misalignment [7–11].

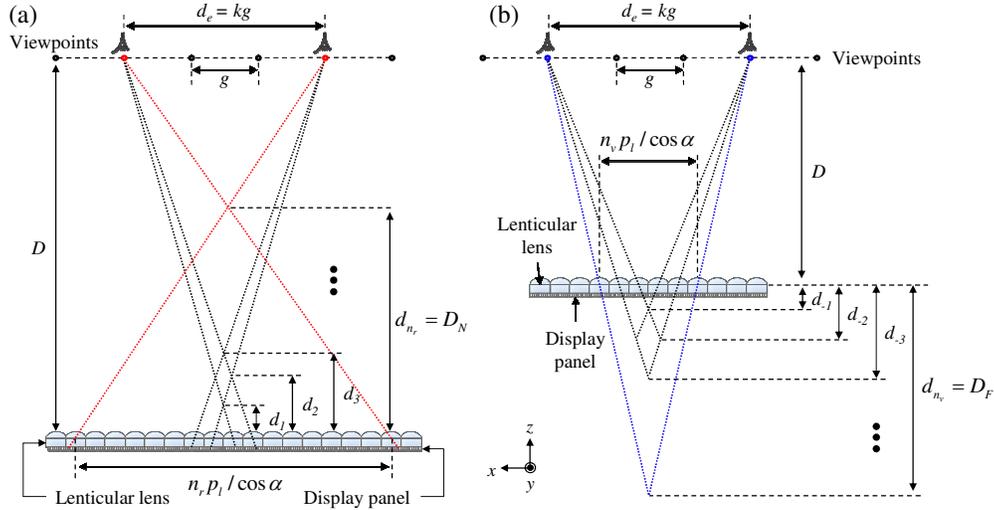


Fig. 3. Expressible depth planes in (a) real and (b) virtual mode of multi-view display ($k = 3$).

The principle of multi-view display for representing 3D object is the same as the method of stereoscopic display in the fixed viewpoint, even though the number of viewpoints is N . The expressible depth planes are quantized with the finite lens disparity as shown in Fig. 3. Equation (5) shows the depth plane determined by the n lens disparity d_n .

$$d_n = \frac{np_l D}{kg \cos \alpha + np_l}. \quad (5)$$

As shown in Fig. 3(a), the cross point of two different rays from left and right lenses is located in front of the display panel when the 3D object is reconstructed in the real mode. The lens disparity n is the positive integer value in the real mode and the nearest depth plane D_N is formed from the maximum lens disparity in the real mode n_r . Furthermore, the 3D object in the virtual mode is reconstructed at the rear of display panel with the lens disparity of the negative integer value as shown in Fig. 3(b). The smallest lens disparity n_v forms the farthest depth plane D_F and the interval of depth planes in the virtual mode is larger than the real mode. The maximum lens disparity n_r and the minimum lens disparity n_v are derived as follows:

$$n_r = \frac{D_N kg \cos \alpha}{(D - D_N) p_l}, \quad n_v = \frac{D_F kg \cos \alpha}{(D - D_F) p_l}. \quad (6)$$

From the multi-view display specification, the expressible depth level of display which is the same as the fundamental depth resolution n_d is determined by the difference of maximum lens disparity and minimum lens disparity as follows:

$$n_d = n_r - n_v = \frac{kg \cos \alpha}{p_l} \left(\frac{D_N}{D - D_N} - \frac{D_F}{D - D_F} \right). \quad (7)$$

If the capturing and transmitting processes provide the depth map with lower depth resolution than the expressible depth level of display, the multi-view display represents the 3D object with quantization and cracking in the depth direction. On the other hand, the depth resolution of reconstructed 3D object is limited by the expressible depth resolution of display even though the depth resolution of depth map is higher than the expressible depth level. However, the assumption of depth resolution limitation is considered except the characteristics of perception in HVS.

B. View synthesis parameters from specification of stereo pickup and multi-view display

In the multi-view 3D broadcasting, the relation among the stereo pickup specifications, the multi-view synthesis parameters and the display specification affect the perception of depth resolution of reconstructed 3D object. The 3D display has not only the planar resolution, but also the resolution of depth direction to display the 3D volume object. However, the observer perceives the depth resolution of reconstructed 3D object differently because of the various cues and effects in HVS. One of the depth perception effects in HVS is the cardboard effect [19–21].

The cardboard effect refers to a phenomenon where the 3D objects represented at different depth planes appear as flat layers to observers. If the spatial thicknesses of represented 3D object and acquired 3D object are different, the reconstructed 3D object is perceived to the observer as the planar images with the cardboard effect. To analyze the occurrence condition of cardboard effect, the specification of stereo pickup and the relation between multi-view synthesis and display are shown as Fig. 4. In the multi-view 3DTV broadcasting, the first step is capturing the stereo images and the depth information using shift sensor configuration as shown in Fig. 4(a). As shown in Fig. 2, the principal point of lenticular lens is shifted to form the view images at the observer distance D because N -pixel behind one lenticular lens is larger than lens pitch p_l . Therefore, the configuration of shift sensor model uses the same principle as the multi-view display, and the plane of convergence is established by a small shift h of the sensor targets as shown in Fig. 4(a) [5,12]. For generating depth map, the near depth plane D_{NS} and far depth plane D_{FS} of 3D objects are set around the convergence distance D_S .

After acquiring the stereo images and depth map, N -view images are synthesized by the recapturing of the reconstructed 3D object in virtual space using DIBR as shown in Fig. 4(b). The coordinate of stereo pickup is scaled with the magnification factor r which is defined as D/D_S . In the DIBR process, N virtual cameras capture the reconstructed 3D object which is reconstructed from the depth map and the stereo images, considering the specification of multi-view display [12–16]. The distance from reconstructed 3D object to multi-view cameras is equal to the observer distance D . The multi-view cameras from C_L to C_N and the scaled stereo camera C_L' and C_R' are located at the same baseline and aligned in the scaled interaxial distance of stereo camera rg_s .

In this situation, the interaxial distance of multi-view cameras g_c is set with consideration of the specification of multi-view display. Although the convergence distance of multi-view cameras is the same as the observer distance D , the near and far depth planes of display D_N and D_F are different from the near and far depth planes of reconstructed 3D object rD_{NS} and rD_{FS} . If rD_{NS} and rD_{FS} are smaller than D_F and D_N , the multi-view display can show the 3D object in the expressible depth range without flipping and cracking. Therefore, g_c is determined to be the same as g , and the observer can watch the 3D object without the cardboard effect.

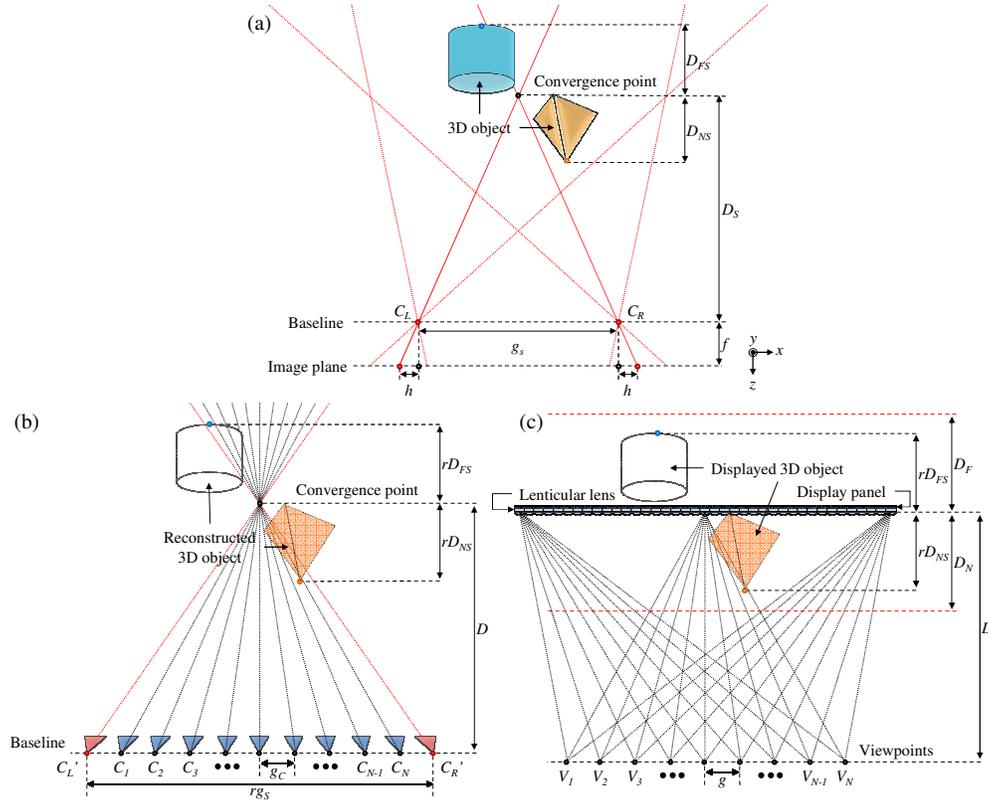


Fig. 4. Parameters of stereo to multi-view camera configuration and display: (a) stereo pickup, (b) multi-view pickup, and (c) multi-view display.

On the other hand, the multi-view display cannot express the exact 3D object when the depth range of reconstructed 3D object in virtual space exceeds the expressible depth range. To prevent the excess of expressible depth range in the multi-view display, the spatial thickness of reconstructed 3D object has to be magnified with the adjustment of interaxial distance of multi-view camera g_c . The interaxial distance g_c with consideration of the depth range in multi-view display is derived as shown in Eq. (8), where the rD_{NS} and rD_{FS} are larger than D_N and D_F .

$$g_c = \begin{cases} \frac{gD_N}{D - D_N} \cdot \frac{D - rD_{NS}}{rD_{NS}}, & D_{NS} \geq -D_{FS} \\ \frac{gD_F}{D - D_F} \cdot \frac{D - rD_{FS}}{rD_{FS}}, & D_{NS} < -D_{FS}. \end{cases} \quad (8)$$

From the relation between the parameters of multi-view synthesis and display, the ratio of spatial thickness between pickup and display configuration E_c is determined as follows:

$$E_c = \frac{D}{g} \cdot \frac{g_c}{D_c}. \quad (9)$$

In previous researches, the cardboard effect is vanished when the ratio of spatial thickness E_c is kept as 1 [19–21]. In addition to the ratio of spatial thickness, the light source of 3D object and background affect the cardboard effect. The use of the flat light source instead of

the point light source and the case of 3D image without background can reduce the cardboard effect. For reducing the cardboard effect in pickup and synthesizing process, the ratio of observer distance to viewpoint distance in the display and the ratio of convergence distance and camera interval in the synthesis process have to be kept by the synthesis parameters.

However, the depth camera acquires the depth information between D_{NS} and D_{FS} regardless of the expressible depth range of multi-view display from D_N to D_F because the specifications of each multi-view displays are different in the multi-view 3D broadcasting. Therefore, the parameters of synthesis are defined by the specification of multi-view display, and occurring of the cardboard effect is an inevitable phenomenon in the multi-view 3D broadcasting. From the result of cardboard effect in the display, we can assume the perceived depth resolution is decreased and the saturation of perceived depth resolution exists in the multi-view 3D broadcasting.

3. Synthesis and numerical comparison of multi-view images in PSNR and NCC with varying depth resolution

A. Stereo pickup and multi-view synthesis of 3D object with varying depth resolution

In the previous section, we discussed the fundamental depth resolution and the synthesis parameters with considering the cardboard effect, which are the factors to affect the perceived depth resolution. With consideration of the fundamental depth resolution and the cardboard effect, we perform the evaluation process of the perceived depth resolution as shown in Fig. 1. To evaluate the saturation bits of perceived depth resolution, we first acquire the stereo images and the depth information of 3D object with varying the depth resolution from 1 bit to 12 bits. The pickup process uses the computational pickup framework using OpenGL to secure the ground truth of depth map and change the depth map resolution easily.

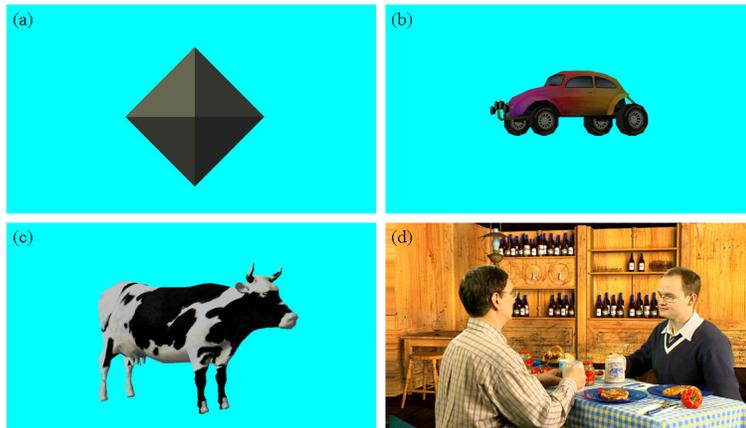


Fig. 5. Contents for evaluation process of perceived depth resolution: (a) pyramid, (b) car, (c) cow, and (d) beergarden.

To reduce the cardboard effect, we use the flat light source and the 3D object without background. The evaluation process is applied to the 3 kinds of computer graphic (CG) contents and the contents of actual beergarden objects which are captured by the 3D4YOU consortium [21] as shown in Fig. 5. The beergarden content involves the additional cues such as occlusions, perspective and shades in the background of objects. In the case of real 3D contents, the background condition, lighting condition and spatial distortion are inevitable condition in acquiring process. Therefore, we perform the evaluation process with CG contents and beergarden content to compare the condition of cardboard effect. In consideration of the recent broadcasting environment, the resolution of stereo image is set to the full HD resolution (1920 by 1080).

Before the acquiring stereo images and depth map, we set the position of 3D object from the convergence distance D_S differently for three modes such as the real, the virtual and the real-and-virtual mode. We use the three modes of contents in the evaluation process so as to find the effect of the distance from 3D object to observer because the observer distance of multi-view display is fixed. In the real mode, the whole volume of 3D object is located in front of the convergence point, and the 3D object in the virtual mode is located at the behind of convergence point. In the real-and-virtual mode, the center position of 3D object is the same as D_S to generate both the real and the virtual 3D images.

In this paper, we assume the expressible depth range of multi-view display is set to ± 150 mm and the observer distance is 1175.45 mm in front of the display panel. Table 1 shows the specification of stereo pickup and the positions of 3D object in the different modes. In CG contents, the pickup parameters D_S and g_s are set to the same as D and N_g of multi-view display to avoid the cardboard effect. On the other hand, the beergarden content which has 1897.18 mm D_{NS} and -1897.18 mm D_{FS} is captured without considering the expressible depth range of multi-view display. To generate the three modes of beergarden content, we adjust the convergence point of the beergarden content with different g_s . From the setting of near depth plane D_{NS} and far depth plane D_{FS} , the depth map is acquired with varying the depth resolution from 1 bit to 12 bits as shown in Fig. 6. In the CG contents, the depth map can be acquired with varying the depth resolution. However, the depth resolution of beergarden is fixed to 8 bits. To generate the depth map of beergarden content with 1 bit to 13 bits depth resolutions, the depth information of beergarden contents is converted from intensity of 8 bit gray level values to real distance [12]. After the conversion, the depth information is quantized with nonlinear quantization equation as follows:

$$I = (2^k - 1) \left[\frac{D_{NS} (D_{FS} - D)}{D (D_{FS} - D_{NS})} \right], \quad (10)$$

where I specifies the respective intensity of k bits depth map. Each depth map with different depth resolution is generated by the different quantization levels with bicubic interpolation method.

Table 1. Stereo pickup specification of contents for evaluation of perceived depth resolution

| Mode | Position of object (mm) | | D_S (mm) | | g_s (mm) | |
|----------------|-------------------------|-------------------|------------|------------|------------|------------|
| | CG | Beergarden | CG | Beergarden | CG | Beergarden |
| Real | 0 ~150 | 0 ~3794.35 | 1175.45 | 4742.94 | 325 | 318.9627 |
| Virtual | -150 ~0 | -3794.35 ~0 | 1175.45 | 948.59 | 325 | 63.79 |
| Real & virtual | -150 ~150 | -1897.18 ~1897.18 | 1175.45 | 2845.77 | 325 | 191.38 |

After the pickup process, the synthesis parameters for multi-view image are determined by the pickup and the multi-view display specifications. The view images are synthesized by DIBR. The DIBR methods are researched and proposed by the many groups in computer vision and image processing field [12–16]. For the reasonable evaluation of the perceived depth resolution, we try to reduce the influence of DIBR process, therefore the view synthesis reference software (VSRS) of MPEG is used as the reference method [16].

In the CG contents, the synthesis parameters are set to the same as parameters of the multi-view display because D_{NS} and D_{FS} are equal to D_F and D_N . However, the depth range of beergarden content exceeds to the expressible depth range of the multi-view display of experimental setup. Therefore, the cardboard effect occurs and the spatial thickness is reduced as shown in Table. 2.

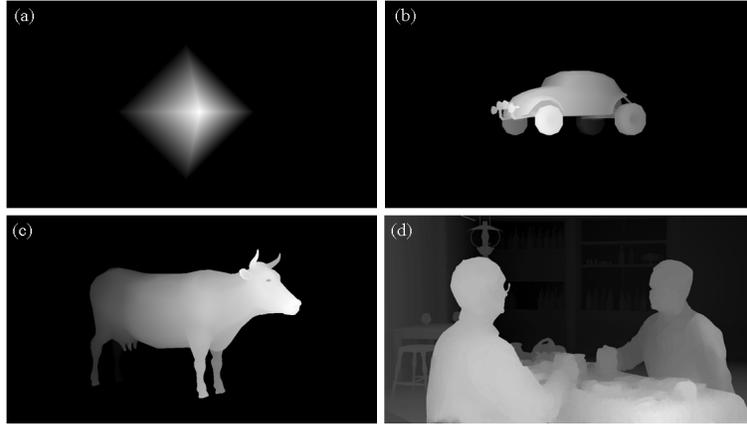


Fig. 6. Acquired depth maps of 4 contents with varying depth resolution from 1 bit to 12 bits: (a) pyramid (Media 1), (b) car (Media 2), (c) cow (Media 3), and (d) beergarden (Media 4).

Table 2. Multi-view pickup specification of beergarden contents

| Mode | R | E_c | g_c (mm) |
|----------------|--------|--------|------------|
| Real | 0.2478 | 0.0366 | 1.1893 |
| Virtual | 1.2392 | 0.1415 | 4.5975 |
| Real & virtual | 0.4131 | 0.0731 | 2.3762 |

The synthesized N -view images with different depth resolutions and modes are interwoven to each interwoven images for the multi-view display. Therefore, the interwoven images with different depth resolutions are displayed to the evaluation of perceived depth resolution with the subjective test.

B. Numerical comparison of synthesized view images in PSNR and NCC with varying depth resolution

Before the subjective test, the numerical comparison of synthesized view image is performed by PSNR and NCC. The ground truth image of pickup process with CG in 5th view image position is captured by the CG pickup framework, which is set to the reference image of PSNR and NCC calculation. We calculate the PSNR and NCC value between the synthesized 5th view image with varying depth resolutions and ground truth image in different modes as shown in Fig. 7.

As shown in Fig. 7, PSNR and NCC values with higher depth resolution are higher than lower depth resolution case although some fluctuations of PSNR and NCC values occur from the hole-filling of DIBR process. The PSNR and NCC values have the saturation of depth resolution between 5 bits and 7 bits in the CG contents. In the beergarden content, the depth information is captured by the time-of-flight camera with 8 bits depth resolution, and the maximum depth resolution of content is increased to 13 bits by image processing technique. Nevertheless, the PSNR and NCC values are saturated around 5 to 7 bits as shown in Fig. 7(d). Therefore, the numerical comparison shows the existence of saturation of depth resolution in the synthesis process, and the saturation values are 7 bits or under. From the numerical comparison, the depth resolution in the synthesis process can be reduced to the saturated depth resolution which is less than 8 bits in the conventional system.

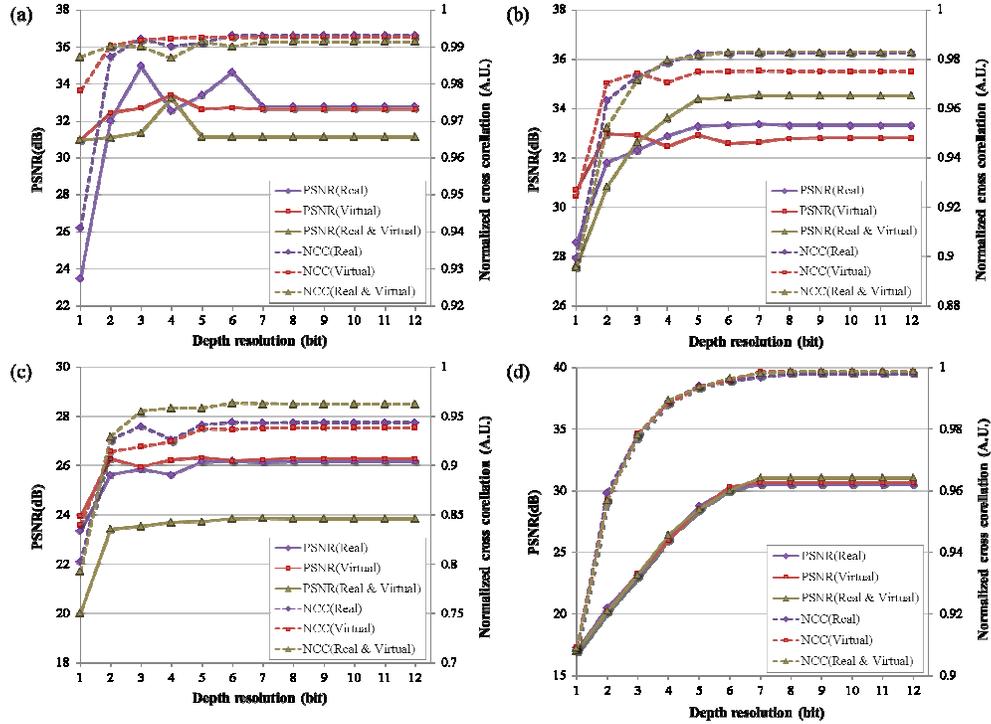


Fig. 7. Numerical comparison of synthesized view image and ground truth image in PSNR and NCC with varying depth resolution: (a) pyramid, (b) car, (c) cow, (d) beergarden.

4. Subjective test for limitation of perceived depth resolution in multi-view display

To find the effect of fundamental depth resolution in multi-view display and the cardboard effect to limitation of perceived depth resolution, we performed the subjective test with participants using the multi-view display. The process before the subjective test is the same as the evaluation process of the numerical comparison. In subjective test, we show the interwoven image with the different depth resolutions to the observer using the 9-view slanted lenticular monitor. The 9-view slanted lenticular monitor is composed of the display panel with high resolution and the slanted lenticular lens as shown in Table. 3. The observer distance is determined from the specification of display panel and lens using Eq. (2).



Fig. 8. Experimental setup of subjective test for perceived depth resolution.

Table 3. Specification of the experimental setup

| Setup | Specification | Characteristic |
|--------------------|--|----------------|
| Display panel | Resolution | 3840 by 2400 |
| | Size | 22 inch |
| | Pixel pitch (p_p) | 0.1245 mm |
| Lenticular lens | Lens pitch (p_l) | 0.1845 mm |
| | Focal length (f) | 0.75 mm |
| | Slant angle (α) | 9.4623 degree |
| Multi-view display | Observer distance (D) | 1175 mm |
| | Number of views (N) | 9 |
| | Interval of viewpoints (g) | 32.5 mm |
| | The number of views per interocular distance (k) | 2 |
| | Near depth plane (D_N) | 150 mm |
| | Far depth plane (D_F) | -150 mm |
| | Resolution of 3D image | 1280 by 800 |

Figure 8 shows the experimental setup of subjective test. To evaluate the saturation of perceived depth resolution, we perform the experiments that the observers watch the reconstructed 3D object using the 9-view slanted lenticular display with the 4 different contents with varying the depth resolution from 1 bit to 12 bits which have three modes with different object positions. The expressible depth range of multi-view display panel is from 150 mm to -150 mm and the depth range of content cannot exceed this limitation to avoid the flipping image. Figure 9 shows the perspectives of 4 contents for the evaluation process at different viewpoints.

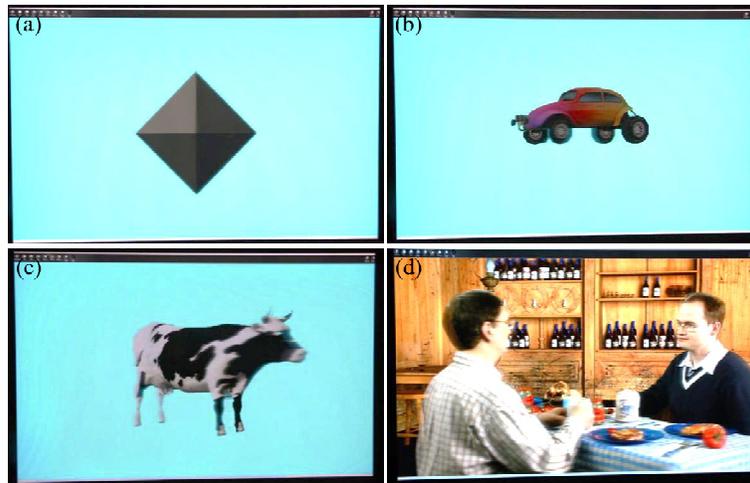


Fig. 9. Represented 3D objects in 4 contents using 9-view slanted lenticular monitor: (a) pyramid (Media 5), (b) car (Media 6), (c) cow (Media 7), and (d) beergarden (Media 8).

Before the evaluation process, the observer should adjust the offset of interwoven image because the alignment process is affected to the perception of depth in the multi-view display. The observer sits 1175 mm in front of multi-view display and adjusts the offset using the view image controller. The observer finds the acceptable viewpoint of different contents. After the alignment process, the 9-view monitor reconstructs the 3D contents with varying the depth resolution from 1 bit to 12 bits. If the observer can feel a 3D effect and find the difference between 3D images with low depth resolution and high depth resolution, the depth resolution of reconstructed 3D object is increased. After the increment of depth resolution, if the observer cannot find the difference between reconstructed 3D objects from high and low depth resolution, the saturated depth resolution in this situation is determined.

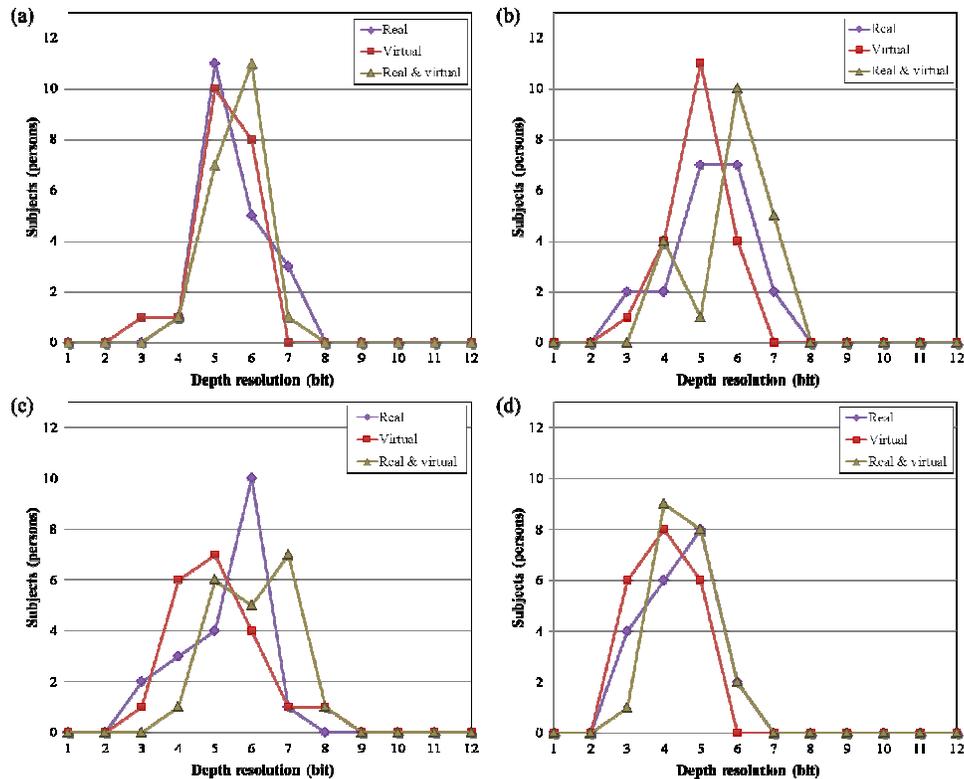


Fig. 10. Experimental result of subjective test with varying depth resolution: (a) pyramid, (b) car, (c) cow, and (d) beergarden.

Figure 10 shows the result of subjective test in the different contents and modes with 20 participants. The participants are staff and students of our research groups (17 men and 3 women) with a mean age of 28.95 years (range from 23 to 37 years). All participants have the experience of the 3D display and do not have strabismus. They can feel the 3D effect from the multi-view display and perceive the reconstructed 3D objects with different depth resolution and modes. As shown in Fig. 10, the threshold of perceived depth resolution is not increased with the depth resolution of depth maps, which is saturated around 5 to 7 bits in the CG contents. In the case of pyramid content, most of the participants choose the same depth resolution as the threshold of the perceived depth resolution. On the other hand, the experimental result of cow content is a little bit spread out. The reason of different variance of experimental result is the characteristics of contents. The pyramid content has smooth, continuous and simple structure whereas the cow content has many curve and complex structures such as legs and horns. Although the result depends on the characteristic of contents and the result of the case of complex contents is spread out, the threshold of perceived depth resolution in the virtual mode marks the lowest value and the real-and-virtual mode needs the highest depth resolution.

In contrast, the result of beergarden contents marks the lower depth resolution than the CG cases because of the cardboard effect from the synthesis process. When the cardboard effect occurs, the observer perceives the reconstructed 3D object as the floating planes. Therefore, the observer is insensitive to the increments of depth resolution due to the cardboard effect.

The result of subjective test shows that the threshold of perceived depth resolution exists in the environments of multi-view broadcasting system. We calculate the fundamental depth

resolution of multi-view display to compare with the fundamental depth resolution of multi-view display and the subjective test results. The fundamental depth resolutions are 5.6724, 5.2854 and 6.4919 bits in the real, the virtual and the real-and-virtual modes from Eq. (7).

To show the tendency of the perceived depth resolution in the multi-view display with the different depth resolution, the result of subjective test in the different modes with average values is represented as shown in Fig. 11. From the tendency, the distribution of the threshold of perceived depth resolution in each mode is very similar to the fundamental depth resolution of multi-view display. However, all depth resolution values from the subjective tests mark the lower value than the fundamental depth resolution. Especially, the perceived depth resolution of beergarden content marks the lowest value because of the cardboard effect. Therefore, the observer perceives the depth resolution lower than the fundamental depth resolution of multi-view display and the depth perception in the multi-view display is more insensitive when the cardboard effect appears.

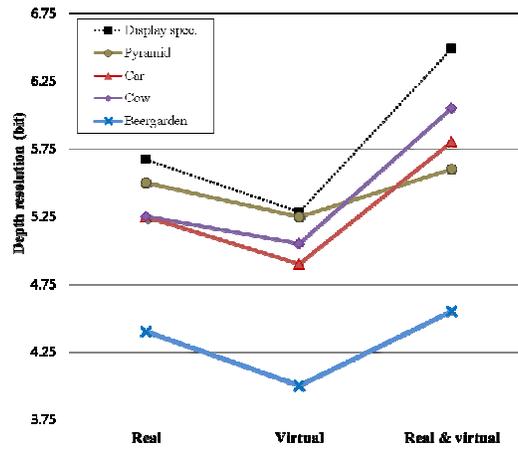


Fig. 11. Experimental result of subjective test with average values in different modes.

The cognitive factors, the technical factors and the psychological condition affect the perception of depth resolution in multi-view display. We analyze and find the threshold of perceived depth resolution based on the technical factors from the specification of multi-view display and broadcasting process. From the numerical and subjective experimental results, the technical factors to saturate the perceived depth resolution in the multi-view broadcasting are the multi-view synthesis process, the fundamental depth resolution of multi-view display and the cardboard effect.

The first technical factor occurs in the multi-view synthesis process from the stereo images and the depth information using DIBR. From the numerical experiments with PSNR and NCC values, the depth resolution of 3D contents for multi-view display is saturated around 5 to 7 bits, which comes from the DIBR process. Even if the depth resolution of 3D contents is over the saturated depth resolution from the numerical experiments, the DIBR process cannot improve the quality of synthesized view-image. The second technical factor is the fundamental depth resolution of multi-view display which is determined by the principle of multi-view display and the depth quantization problem. Therefore, the perceived depth resolution follows the lowest depth resolution from the first or second factor.

The last factor is the cardboard effect between pickup and display processes. Generally, the depth range of pickup specification exceeds the expressible depth range of multi-view display, which results in the cardboard effect from the difference of the ratio of spatial thickness. Although the depth resolution of 3D contents satisfies the numerical limitation and the fundamental depth resolution of multi-view display, the observer is becoming desensitized to the depth resolution because of the cardboard effect.

5. Conclusion

We find the effect of the fundamental depth resolution of multi-view display and the cardboard effect from the synthesis process to the depth perception on multi-view display. To find the threshold of perceived depth resolution, we analyze the fundamental depth resolution and the factors for cardboard effect and perform the evaluation process. According to the subjective tests with 20 participants and the numerical comparison with PSNR and NCC, we find the threshold of depth resolution in the view synthesis process and the limitation of perceived depth resolution in multi-view display. The perceived depth resolution is lower than the fundamental depth resolution and shows very similar distribution with the fundamental depth resolution. In addition, the cardboard effect decreases the perceived depth resolution in the multi-view display. The technical factors for the limitation of perceived depth resolution in the multi-view display are analyzed and described.

Acknowledgment

This work was supported by the IT R&D program of MKE/KEIT [KI10035337, development of interactive wide viewing zone SMV optics of 3D display].