Effects of Depth Map Quantization for Computer-Generated Multiview Images using Depth Image-Based Rendering

Minyoung Kim[1], Yongjoo Cho[2], Hyon-Gon Choo[3], Jinwoong Kim[3] and Kyoung Shin Park[4]
[1] Department of Computer Science, Sangmyung University, Seoul, Korea
[e-mail: pupleshine@gmail.com]
[2] Division of Digital Media Technology, Sangmyung University, Seoul, Korea
[e-mail: ycho@smu.ac.kr]
[3] Electronics and Telecommunications Research Institute, Taejon, Korea
[e-mail: {hyongonchoo, jwkim}@etri.re.kr]
[4] Department of Multimedia Engineering, Dankook University, Cheonan, Korea
[e-mail: kpark@dankook.ac.kr]
*Corresponding author: Kyoung Shin Park

Received May 25, 2011; revised July 27, 2011; revised August 25, 2011; accepted September 20, 2011; published November 29, 2011

Abstract

This paper presents the effects of depth map quantization for multiview intermediate image generation using depth image-based rendering (DIBR). DIBR synthesizes multiple virtual views of a 3D scene from a 2D image and its associated depth map. However, it needs precise depth information in order to generate reliable and accurate intermediate view images for use in multiview 3D display systems. Previous work has extensively studied the pre-processing of the depth map, but little is known about depth map quantization. In this paper, we conduct an experiment to estimate the depth map quantization that affords acceptable image quality to generate DIBR-based multiview intermediate images. The experiment uses computer-generated 3D scenes, in which the multiview images captured directly from the scene are compared to the multiview intermediate images constructed by DIBR with a number of quantized depth maps. The results showed that there was no significant effect on depth map quantization from 16-bit to 7-bit (and more specifically 96-scale) on DIBR. Hence, a depth map above 7-bit is needed to maintain sufficient image quality for a DIBR-based multiview 3D system.

Keywords: Depth image-based rendering, depth map quantization, multiview intermediate image generation, 3D displays

This work was supported by the IT R&D program of the Ministry of Education, Science and Technology in Korea, [Development of Technologies for Depth Map Acquisition and Its Application to 3DTV Broadcast, ETRI 2011].

DOI: 10.3837/tiis.2011.11.017

1. Introduction

There has been a growing interest in multiview 3D display systems due to the increase in the performance and capabilities of 3D content generation and rendering. A multiview 3D system displays multiple independent views of a scene simultaneously, with various viewing angles. The system offers multiple users the experience of a 3D visual sensation without the need for special glasses. This system generally involves capturing, processing, transmitting, and displaying multiview 3D scenes. The easiest way to support such a system is to use an array of multiple cameras to simultaneously capture a scene, which provides a smooth change between viewpoints and transmits them to the receiver side. However, the major problems with this process are the high-cost of capturing multiview video and the technical difficulty of the simultaneously transmitting multiview images, which significantly increases the amount of raw data compared to normal 2D video or stereo-pair video. Moreover, the video captured by this process may not be compatible with all multiview 3D displays, which differ from each other in terms of the number of views and screen size.

For efficiency reasons, depth image-based rendering (DIBR) was introduced as one of the main technologies in multiview 3D displays [1][2]. When DIBR is used, 2D color video and depth information (also called depth map) are used instead of multiview images to synthesize device-independent "virtual" views (with different view angles and screen sizes) of a scene (also known as intermediate views) [3][4]. The Philips WOW auto-stereoscopic 3D display is an example of the use of DIBR to reconstruct 9-view intermediate images for a 3D visual experience [5]. However, high-quality DIBR-based multiview intermediate image generation is a challenging task, since it requires precise depth information [6].

Depth maps are generated using a depth camera (such as ZCam by Israel 3DV Systems Inc. or Axi-vision by Japan NHK). The camera directly measures the distance of objects in a scene from the time required to bounce an infrared or ultrasound signal back to a sensor [7][8]. Another method is to estimate the disparity between stereoscopic pair images [9]. However, the depth map obtained from a depth camera is error-prone, and the disparity estimation often results in a low-quality depth map. Hence, many approaches have been proposed to improve the quality of depth maps in order to yield better multiview intermediate images. For example, a median filter or asymmetric Gaussian smoothing filter is applied when pre-processing the depth map to produce sharp and high-quality DIBR-based stereoscopic images [10][11]. Disparity compensation is also used to estimate dense depth maps [12]. More work has been done on up-sampling depth images [13][14] and depth-edge enhancement in order to correct and sharpen the object boundaries by extracting information from available high-resolution color images [15][16].

As discussed earlier, DIBR often involves a considerable amount of pre-processing and post-processing operations to compensate for the deficiencies of the depth maps. While pre-processing and post-processing algorithms have been studied extensively in the past years, few quality assessments have been performed on depth map quantization in DIBR. In recent studies, it has been found that the quantization of pixel depth introduces noise in DIBR [17][18]. In addition, a subjective evaluation shows that 20-scale depth map quantization levels are sufficient for synthesizing the stereoscopic images when a scene consists of simple objects, such as a sphere or cone [19]. At present, a multiview 3D display system typically uses an 8-bit depth map due to the limitations of current depth cameras [1][6][20]. Hence, we performed a comprehensive evaluation to find out how various depth map quantization levels

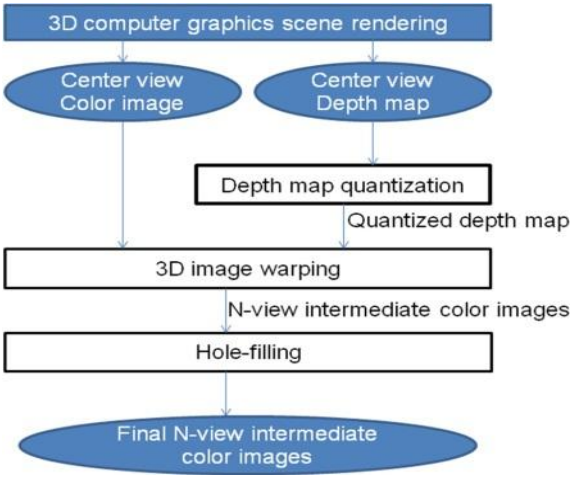

Fig. 1. Block diagram of DIBR-based multiview intermediate image generation for depth map quantization.

affect the overall quality of DIBR.

In this study, a computer-generated scene with an additional 16-bit depth map was used to evaluate the effect of a more accurate depth map. Moreover, a depth map with low quantization levels was evaluated to estimate the minimum required accuracy of depth information in DIBR. The experiment was conducted using different depth map quantization levels for constructing multiview intermediate images using DIBR. Each pixel depth was quantized to the discrete number of the pixel depth value, from a 16-bit (65536 scales) to a 1-bit (2 scales) depth. Four 3D computer-generated scenes were selected to represent video with a variety of depth and foreground object complexity and used two multiple-camera array alignments (parallel versus convergent). The image quality was evaluated using the peak signal-to-noise ratio (PSNR) of the original 5-view camera images taken directly from the scene and the 5-view intermediate images by DIBR.

This paper presents the experimental method conducted to find the effective quantization level of the depth map for generating reliable, accurate multiview intermediate images for a multiview 3D display system. It then discusses the experiment results and concludes by presenting future research possibilities.

2. Experimental Setup

This section describes the experimental setup for the evaluation of the depth map quantization in depth image-based rendering (DIBR). The primary objective of this experiment was to estimate the depth map quantization level that affords acceptable image quality for multiview intermediate image generation. Hence, the initial depth map for a 3D scene was converted to the corresponding quantized depth map in the pre-processing step of DIBR. In the experiment, we evaluated the 5-view intermediate images generated by DIBR for each of the four 3D scenes and two camera configurations, with the depth map quantized to a discrete number of pixel depth values, from a 16-bit (65536 scales) to a 1-bit depth (2 scales).

Fig. 1 shows the block diagram of our DIBR-based multiview intermediate image generation system for a 3D computer graphics scene. The system takes the center-view color

Fig. 2. Four 3D computer graphics scenes (the color image at the center view and the gray-scale image representing its associated depth map): **(a)** Farm, **(b)** Zebra, **(c)** Car, **(d)** Treasure.

image and its associated depth map, which has the same image resolution as the color image. Depth map quantization is performed after the depth map is acquired. Here, each pixel-depth value is rounded to one of the number of levels needed to produce a quantized depth map. The system then processes the 3D image warping, which involves the back-projection of 2D points (per pixel from the center-view color image) into a 3D world, depending on their depth value from the quantized depth map. Then the 3D points are re-projected onto the 2D image plane of the multiview virtual cameras to create the multiview intermediate color images. Finally, the hole-filling algorithm (based on Zhang's method [10]) is applied to produce the final multiview intermediate color images. This system also provides the multiview images captured directly from multiple camera viewpoints and from the users' navigation of the 3D scene.

Fig. 2 shows four 3D computer graphics scenes (the center-view color image and its associated depth map) selected for the experiment: Farm, Zebra, Car, and Treasure. Greater luminance in the associated gray-scale depth image represents a closer distance (smaller depth value) to the viewer. The Farm scene consists of a landscape with a vast expanse of field and sky. This scene is quite monotonous when compared to the other scenes. The Zebra scene shows a striped zebra in front of a house on rocky ground. The Car scene consists of a curvy vehicle and a repeated texture pattern on the wall and the floor. The Treasure scene consists of numerous colorful objects, such as a crown, a ring, a sword, a big purse, and other jewels on a table. The Car and Treasure scenes have more foreground objects located closer to the viewer than do the Farm and Zebra scenes. These 3D computer graphics scenes were rendered at an image resolution of 800×800 pixels.

Fig. 3 shows the two configurations of the virtual camera alignment used in the experiment (i.e., a 5-view camera array aligned in parallel versus convergent configuration). Note that the center of the 5-view camera array is the third-view. All data for the two camera configurations were obtained separately so as to compare their differences. The array of five cameras in the parallel configuration looks at the depth of infinity, since the axes of the multiple cameras are arranged in parallel. In contrast, the array of five cameras in the convergent configuration focuses on the same point of convergence, in the same way a human would focus on a specific object within a scene. We evaluated these two configurations because the different camera alignments result in different 3D representations [21]. We used a 5-view camera array in the

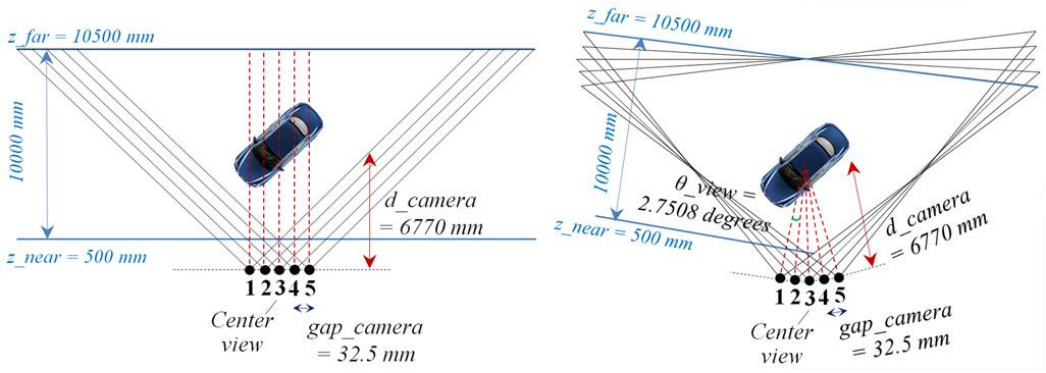

Fig. 3. Two virtual camera alignment configurations of 5-view camera arrays: **(a)** Parallel, **(b)** Convergent.

experiment to limit the size of the holes on DIBR caused by the increased number of camera arrays.

As shown in **Fig. 3**, the gap between cameras, gap_camera, was set to 32.5 mm, on the basis of Eq. 1, given that the average adult inter-ocular distance is approximately 65 mm. The number of virtual cameras between the left and right eyes, delta, is set to three to enable comfortable 3D viewing, as this ensure that there is no crosstalk effect. In addition, the convergence angle of the two virtual cameras, θ_view, was set to 2.7508°, based on Eq. 2, with the viewing distance from a virtual camera to the origin of the 3D scene, d_camera, set to 6770 mm. The z_near and z_far clipping plane of the virtual camera, view frustum, were set to 500 mm and 10500 mm, respectively, in order to cover all four 3D scenes used in this experiment.

$$gap\_camera = \frac{65}{(delta - 1)} \tag{1}$$

$$\theta\_view = 2 \ \ \text{asin}\left(\frac{gap\_camera}{2 \ \ d\_camera}\right) \tag{2}$$

Fig. 4 shows the depth map (800×800 pixels) of the Zebra scene with sixteen different quantization levels. **Fig. 4 (a)** shows the gray-scale image of the initial depth map generated by computer graphics; this is identical to a 16-bit depth (65536 scales). **Fig. 4 (b)-(p)** show the gray-scale images of the depth map quantized to a 15-bit depth (32768 scales), 14-bit depth (16384 scales), 13-bit depth (8192 scales), 12-bit depth (4092 scales), 11-bit depth (2048 scales), 10-bit depth (1024 scales), 9-bit depth (512 scales), 8-bit depth (256 scales), 7-bit depth (128 scales), 6-bit depth (64 scales), 5-bit depth (32 scales), 4-bit depth (16 scales), 3-bit depth (8 scales), 2-bit depth (4 scales), and a 1-bit depth (2 scales), respectively. The quantization of per-pixel depth is performed using Eq. 3, where 16 represents the maximum number of depth bits that can be generated by computer graphics, n is the number of depth bits for quantization, and z specifies the respective gray-level value.

$$Q(z) = \frac{z}{2^{16} \big/ 2^{n}} \ with \ \ z \in [0,...,2^{16} - 1] \tag{3}$$

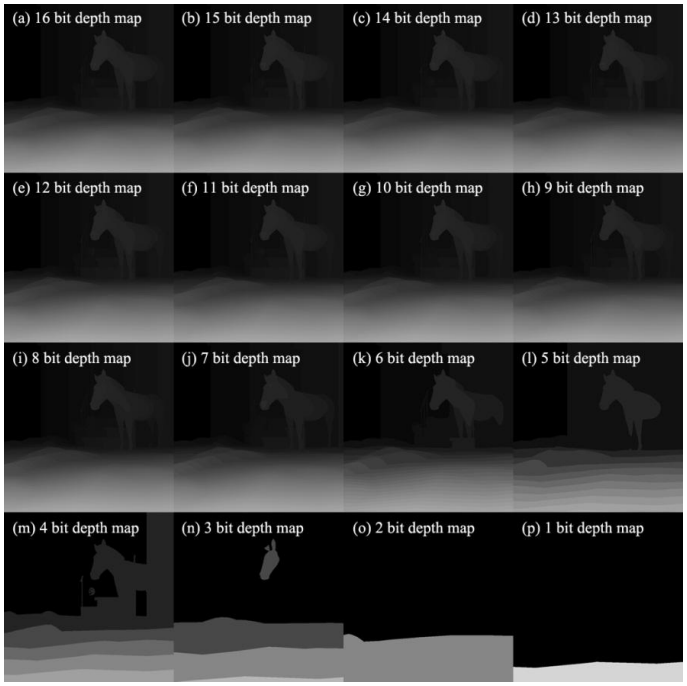

Fig. 4. The Zebra scene with sixteen depth map quantization levels

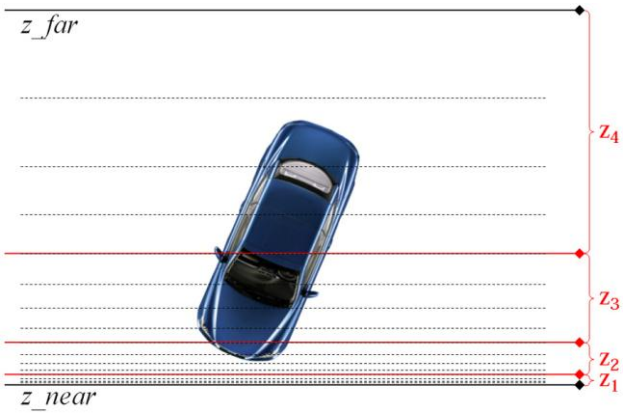

Fig. 5. Depth map quantization to 2-bit depth (4 scales)

The pixels in the depth map determine the distance from the associated color image pixel to the viewer. For example, the depth values range from 0 to 65535 in the initial 16-bit depth map, where 0 represents the position closest to the viewer and 65535 represents the position farthest from the viewer in a 3D scene. In the 8-bit quantized depth map, the initial 16-bit per-pixel depth values are grouped into 256 scales, ranging from 0 to 255. In the 1-bit quantized depth map, the depth values are either 0 or 1, where 0 represents the foreground and 1 represents the background.

As an example, the quantization mapping process from a 16-bit depth to a 2-bit depth is illustrated in **Fig. 5**. The initial 16-bit per pixel depth representation corresponds to a 65536 gray-level value, which is grouped into four equally spaced bins (shown as z1, z2, z3, and z4 in **Fig. 5**). However, the depth (i.e., the z-value in Eq. 3) is non-linear due to the perspective

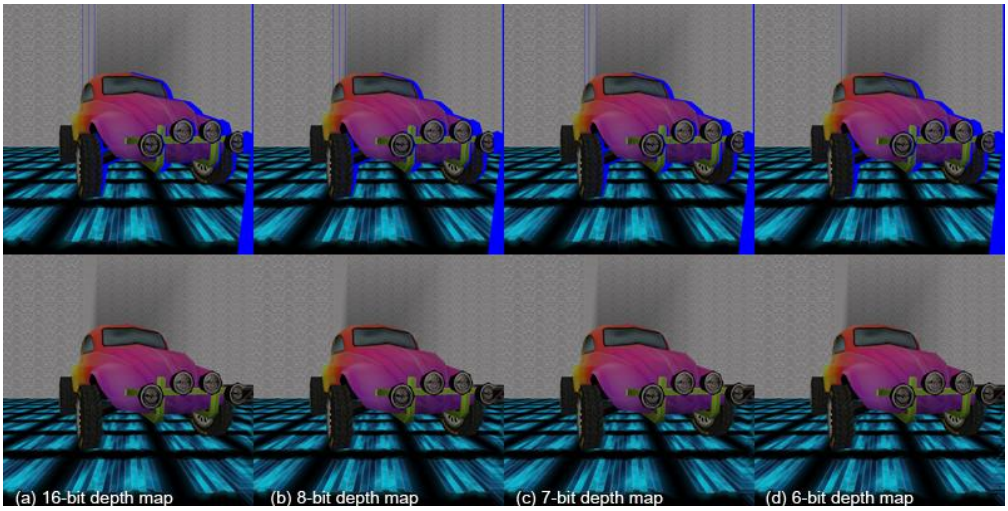

Fig. 6. The Car scene taken at the fifth-view camera viewpoint (in the parallel camera alignment configuration) before and after hole-filling: **(a)** intermediate image with 16-bit depth map, **(b)** 8-bit depth map **(c)** 7-bit depth map, **(d)** 6-bit depth map

division in computer graphics. This means that there is a high degree of precision close to the viewer and less precision farther away [22]. Hence, the quantized depth is also distributed non-linearly within a near and far plane.

Fig. 6 shows the DIBR-based intermediate images of the Car scene taken from the fifth-view camera viewpoint in the parallel camera alignment, both before hole-filling (the blue parts are the holes) and after hole-filling (the hole-filled final intermediate image) with (a) 16-bit (OpenGL uses a 16-bit depth map by default), (b) 8-bit, (c) 7-bit, and (d) 6-bit depth map quantization. As shown in the blue sections of **Fig. 6** (before hole-filling), the locations and numbers of holes differ slightly among different depth map quantization levels. In this case, the final intermediate images constructed with a 16-bit and an 8-bit depth map look almost similar, whereas the contour artifacts at the borders increase slightly when the depth map quantization is 7-bit or below.

In the experiment, the image quality measurements were compared using the peak signal-to-noise ratio (PSNR) between the original 5-view camera images captured directly from a 3D scene and the 5-view intermediate images generated by DIBR with different depth quantization levels. The original camera-captured images represent the ground truth view, whereas the DIBR-based intermediate images represent the synthesized view. Here, the PSNR between two images I and K is defined as:

$$MSE_{color} = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left\|I_{color}(i,j) - K_{color}(i,j)\right\|^2$$

$$MSE_{total} = \frac{1}{c}\sum MSE_{color} \tag{4}$$

$$PSNR = 10\cdot\log_{10}\left(\frac{MAX_I^2}{MSE_{total}}\right) = 20\cdot\log_{10}\left(\frac{MAX_I}{MSE_{total}}\right)$$

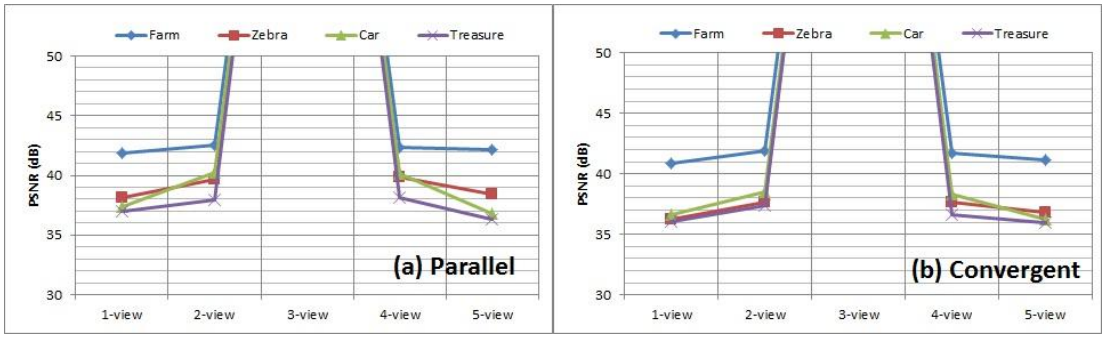

Fig. 7. Average PSNR between the original camera images taken from the scene and the DIBR intermediate image with the initial 16-bit depth map, at each camera viewpoint, and for four 3D scenes and two camera alignment configurations: **(a)** Parallel, **(b)** Convergent.

MSEcolor is the mean squared error (MSE) for each RGB channel, calculated using the color differences between two images, I and K, and the image width and height, m and n. MSEtotal is the total mean squared error for a color image, calculated as the sum of MSEcolor divided by the total number of signals, c. This represents the noise, which is calculated as the average of the three mean squared errors of the difference between matched pixels for each color signal of the two images. Then, the PSNR for a color image is calculated using MSEtotal and MAXI (the maximum brightness value on the color image). This illustrates the logarithmic ratio of the maximum possible power versus the mixed noise between two images.

To make more general comparisons, we calculated the average of the PSNR values between the original camera images and the DIBR-based intermediate images. This was done over twenty sequences, for each 3D scene, captured from different camera view positions and angles in the scene.

3. Results

First, we analyzed the PSNR values between the image taken directly from the computer-generated 3D scene and the intermediate image generated by DIBR with the initial 16-bit depth map in order to evaluate the image quality degradation caused by DIBR. We then used these results as a reference in order to compare the DIBR intermediate images generated with other quantized depth maps, and to evaluate the image quality degradation due to the quantization of pixel depth. Furthermore, we performed a more detailed analysis on the DIBR intermediate image with fine-scale quantization levels between an 8-bit and a 6-bit depth map. Finally, we compared the depth map quantization for the CG scenes against the real scenes.

3.1 Effect of a 16-bit Depth Map on DIBR

Fig. 7 shows the average PSNR value between the image taken directly from the 3D scene (using the correct-viewing depth map) and the intermediate images generated using DIBR with the initial 16-bit depth map (65536 scales) for 20 frames of each 3D scene taken from five camera viewpoints and two camera alignment configurations. This result depicts the image quality degradation caused by DIBR-based multiview intermediate image generation. Note that the DIBR intermediate image at the third view is identical to the original (center-view) color image taken directly from the scene; this is because DIBR performs 3D warping based

on the center-view color image and the depth map. Hence, the MSE between them is 0, and consequently, the PSNR becomes infinite.

Overall, the quality of the DIBR intermediate images with the initial 16-bit depth map when compared to the original camera-captured images at each camera viewpoint produced a PSNR value of above 35 dB. The average PSNR was 42.24 dB for the parallel configuration and 41.42 dB for the convergent configuration in the Farm scene, 39.03 dB (parallel) and 37.08 dB (convergent) in the Zebra scene, 38.64 dB (parallel) and 37.41 dB (convergent) in the Car scene, and 37.36 dB (parallel) and 36.49 dB (convergent) in the Treasure scene.

A three-way ANOVA test was conducted to explore the impact of scene, camera alignment, and camera viewpoint on the PSNR. All three main effects were significant ($p = 0.000$). The image quality gradually degraded when the camera viewpoint moved away from the center view and towards either of the sides (for example, the 1- or 5-view as compared to the 2- or 4-view). This occurred because the leftmost or rightmost sides of the scene were lost, so the pixel color could not be retrieved from the center-view reference image.

The image quality degradation became more visible as more objects were located closer to the viewer in the 3D scene. The PSNR of the Farm scene was approximately 3.21 dB higher than the PSNR of the Zebra scene. The PSNR of the Zebra scene was approximately 0.4 dB higher than the PSNR of the Car scene. The PSNR of the Car scene was approximately 1.28 dB higher than the PSNR of the Treasure scene. These results show that the overall quality of DIBR was more sensitive to complex scenes than to scenes with fewer objects and apparent object boundaries.

The PSNR value between the original camera image and the DIBR intermediate image was significantly affected by the camera alignment (parallel vs. convergent). The image quality was slightly better (approximately 1 dB) when the camera array was aligned in parallel than when it was toe-in aligned (i.e., convergent). This was because the convergent camera alignment causes relatively high disocclusion problems. In contrast, the parallel camera alignment only considered the left or right sides of camera movement, which made it easier to search for appropriate pixels to refer for hole-filling.

The scene and camera viewpoint interaction was significant ($p = 0.000$). The effect of the camera viewpoint was the most significant in the Car scene, which consisted of a checked pattern on the wall and a grid pattern on the floor, as the camera viewpoint moved from the center view to either of the sides (approximately 2~3 dB difference between viewpoints). In contrast, the camera viewpoint had little impact on the DIBR for the Farm scene. The scene and camera alignment interaction was also significant ($p = 0.004$). The effect of the camera alignment configuration was significant in the Zebra and Car scenes, but had little impact on the DIBR for the Farm and Treasure scenes.

3.2 Effect of Depth Map Quantization on DIBR

Fig. 8 shows the average PSNR by depth map quantization levels, ranging from 16-bit (65536 scales) to 1-bit (2 scales) for each 3D scene, with two camera alignment configurations. Overall, the average PSNR between the original camera-captured images and the DIBR intermediate images with a 7-bit depth map or above was over 35 dB, and the mean difference was less than approximately 0.5 dB for all 3D scenes and camera configurations. The quality of the DIBR was considerably degraded by low depth map quantization levels, especially below a 4-bit depth map. Scheffe post-hoc analysis revealed that, on average, at least a 7-bit depth map was required to avoid significant image quality degradation in DIBR.

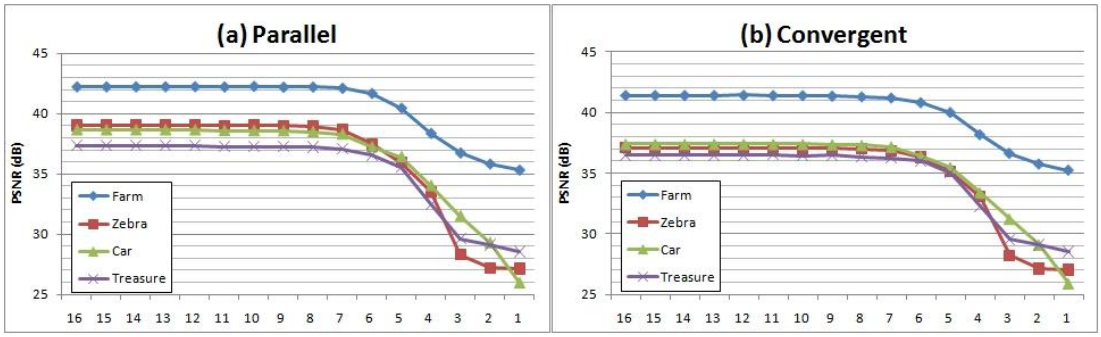

Fig. 8. Average PSNR by depth map quantization levels (16- to 1-bit depth) for four computer-generated 3D scenes and two camera alignment configurations

It appeared that the overall PSNR was significantly degraded by the scene. The closer the foreground objects were located to the viewer in a scene, the lower the PSNR was for DIBR multiview intermediate image generation. In particular, the average PSNR difference for the depth map quantization between the Farm scene and the Treasure scene was 5 dB or greater for both camera alignment configurations. In addition, a significant interaction effect of depth map quantization was found in scenes with repeated texture patterns. For example, the PSNR was highly degraded by the depth map quantization in the Car and Zebra scenes.

For each 3D scene, a two-way multivariate analysis of variance test was performed to investigate the effect of the four camera viewpoints and the sixteen depth map quantization levels on two multivariate response variables (i.e., PSNR by parallel and convergent camera configurations). The effect of the two main factors on the PSNR for all camera alignment configurations in all 3D scenes was significant. The PSNR was significantly degraded as the camera viewpoint moved from the center to the side (regardless of depth map quantization) due to more hole-filling being required. The mean difference between camera viewpoints was less than approximately 1 dB for the Farm scene, approximately 1~2 dB for the Zebra and Treasure scenes, and approximately 2~3 dB for the Car scene.

The PSNR was also significantly degraded by the depth map quantization levels. In the Farm scene, post-hoc comparisons of depth map quantization indicated that the PSNR from 16- to 6-bit depth map quantization did not differ for the parallel (mean difference < 0.55 dB) and convergent (mean difference < 0.60 dB) camera configurations. In the Zebra scene, the PSNR from 16- to 7-bit depth map quantization did not differ for parallel (mean difference < 0.37 dB) and convergent (mean difference < 0.22 dB) camera configurations. In the Car scene, the PSNR from 16- to 7-bit depth map quantization did not differ for parallel (mean difference < 0.34 dB) and convergent (mean difference < 0.23 dB) camera configurations. In the Treasure scene, the PSNR from 16- to 6-bit depth quantization did not differ for parallel (mean difference < 0.79 dB) and convergent (mean difference < 0.48 dB) camera configurations.

In particular, the PSNR was more significantly degraded by low depth map quantization levels, such as a 3-bit depth map or below. In the Car and Zebra scenes, the mean difference between a 16-bit and a 3-bit or lower depth map was greater than 10 dB, and it was approximately 6~8 dB in the Farm and Treasure scenes. The effect of depth map quantization and camera viewpoint interaction was significant in the Zebra and Car scenes for both camera configurations. That is, the PSNR difference between camera viewpoints increased (up to approximately 4 dB) as the depth map quantization level decreased. This indicated that rubber

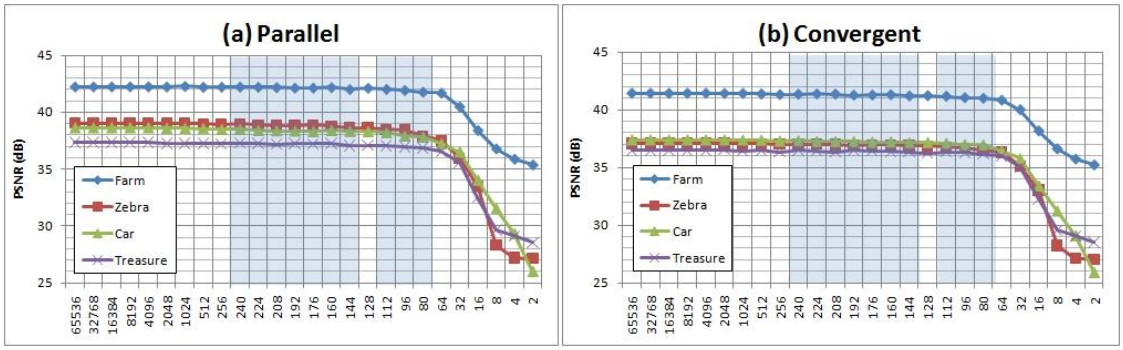

Fig. 9. Average PSNR by fine-scale depth map quantization levels for four computer-generated 3D scenes and two camera alignment configurations

Table 1. Average PSNR by fine-scale depth quantization levels between 8-bit and 5-bit for four 3D scenes and two camera alignment configurations

		Parallel					Convergent				
bit	scale	Farm	Zebra	Car	Treasure	Total	Farm	Zebra	Car	Treasure	Total
8	256	42.2263	38.9222	38.4951	37.2209	39.2161	41.3102	37.0123	37.3086	36.3485	37.9949
	240	42.2269	38.9650	38.5004	37.2726	39.2412	41.3434	37.0027	37.3260	36.5130	38.0463
	224	42.2182	38.8861	38.3988	37.2570	39.1900	41.3705	37.0240	37.2632	36.4398	38.0244
	208	42.1857	38.8386	38.3629	37.1842	39.1429	41.3551	36.9929	37.2629	36.3456	37.9891
	192	42.1355	38.8495	38.3468	37.2626	39.1486	41.2588	36.9278	37.2475	36.4921	37.9815
	176	42.1429	38.8626	38.2979	37.2150	39.1296	41.2694	36.9514	37.1946	36.4330	37.9621
	160	42.1802	38.7725	38.4021	37.2496	39.1511	41.2998	36.9380	37.2162	36.3788	37.9582
	144	42.0132	38.6526	38.3275	37.0524	39.0114	41.1938	36.9047	37.1491	36.3376	37.8963
7	128	42.1266	38.6478	38.2955	37.0895	39.0399	41.2159	36.8569	37.1827	36.2630	37.8796
	112	42.0282	38.4748	38.1855	37.0459	38.9336	41.1682	36.7762	37.1168	36.3496	37.8527
	96	41.9286	38.4353	37.9072	36.9568	38.8070	41.0475	36.6784	36.9964	36.2677	37.7475
	80	41.7814	37.9102	37.8120	36.8549	38.5896	40.9937	36.4785	36.9802	36.1486	37.6503
6	64	41.6867	37.5012	37.2186	36.5614	38.2420	40.8244	36.3073	36.4498	36.0066	37.3970
5	32	40.4866	35.9483	36.5164	35.5642	37.1289	40.0131	35.1349	35.4832	35.0274	37.3897

sheet artifacts caused by inappropriate hole-filling in DIBR were more perceptible, as more vivid texture patterns appeared in the scene (such as the stripped Zebra pattern, and the checked pattern and the grid pattern of the Car scene).

3.3 Effect of Fine-Scale Depth Map Quantization on DIBR

Fig. 9 shows the average PSNR by depth map quantization levels from 16-bit (65536 scales) to 1-bit (2 scales), and the fine-scale depth map quantization levels between 8-bit (256 scales) and 6-bit (64 scales) for four 3D scenes and two camera alignments. **Table 1** shows the PNSR by fine-scale depth map quantization levels ranging specifically from 256-scale (8-bit depth) to 32-scale (5-bit depth). The depth map quantization level was shown as the per-pixel bit

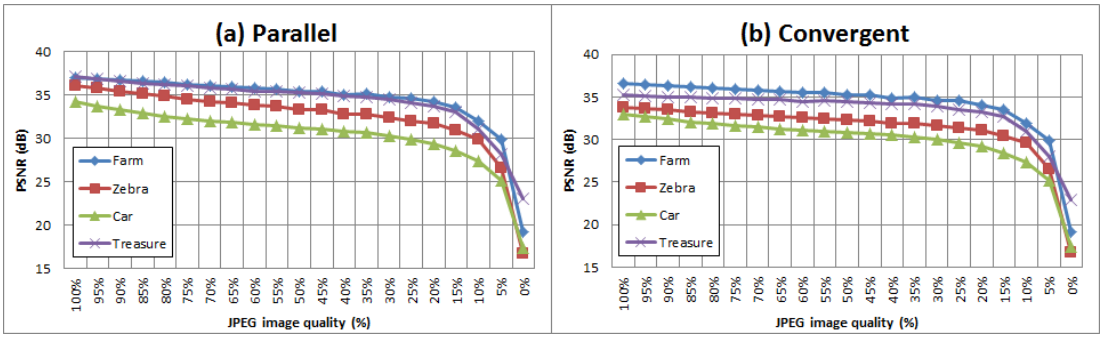

Fig. 10. Average PSNR by JPEG image quality (Q=100~0%) for four 3D scenes

depth and the depth scale (i.e., the range of depth available for each depth map quantization level).

Overall, the PSNR between the original camera-captured images and the DIBR intermediate images with 32-scale depth map or below was greatly degraded for all scenes, regardless of camera alignment configurations. Also, post-hoc comparisons of depth map quantization indicated that the PSNR was not significantly different from a 65536-scale (i.e., the initial 16-bit) to a 96-scale depth map quantization, in which the mean difference was less than approximately 0.5 dB.

In the Farm scene, the PSNR from 65536- to 64-scale depth did not differ for the parallel (mean difference < 0.55 dB) and convergent (mean difference < 0.60 dB) camera configurations. In the Zebra scene, the PSNR from 65536- to 96-scale depth did not differ for the parallel configuration (mean difference < 0.59 dB), and the PSNR from 65536- to 80-scale depth did not differ for the convergent (mean difference < 0.60 dB) camera configuration. In the Car scene, the PSNR from 65536- to 128-scale depth did not differ for the parallel configuration (mean difference < 0.34 dB), and the PSNR from 65536- to 112-scale depth did not differ for the convergent (mean difference < 0.29 dB) camera configuration. In the Treasure scene, the PSNR from 65536- to 64-scale depth did not differ for the parallel (mean difference < 0.79 dB) and convergent (mean difference < 0.48 dB) camera configurations.

Fig. 10 shows the image quality degradation measured by the PSNR with respect to JPEG compression of the first DIBR image frame of each 3D scene. JPEG compression is performed using the IJG implementation of JPEG [23]. For the full quality images (Q = 100%), the PSNR was between 34.1 dB and 37.1 dB, and between 32.9 dB and 36.5 dB for the parallel and convergent camera configurations, respectively. With respect to the effect of JPEG compression, the Car scene, with high-frequency textures, had the lowest PSNR, showing that the image quality was more affected by compression. In contrast, the Treasure scene, with no texture, had a higher PSNR, showing that the image quality was less affected by compression.

3.4 Effect of Depth Map Quantization on DIBR for Real Scenes

Fig. 11 shows the average PSNR by depth map quantization levels from 8-bit (256 scales) to 1-bit (2 scales), and the fine-scale depth map quantization levels between 8-bit (256 scales) and 6-bit (64 scales) for Breakdancer and Ballet sequences from Microsoft [24]. In this experiment, we analyzed 100 frames of Breakdancer and Ballet sequences at 2-, 3-, 4-, 5-, and 6-viewpoints (i.e., a 5-view camera array in a convergent alignment). The intrinsic and extrinsic camera parameters offered by these sequences were directly applied to our DIBR-based multiview system. When compared to the CG scenes, much larger holes appeared

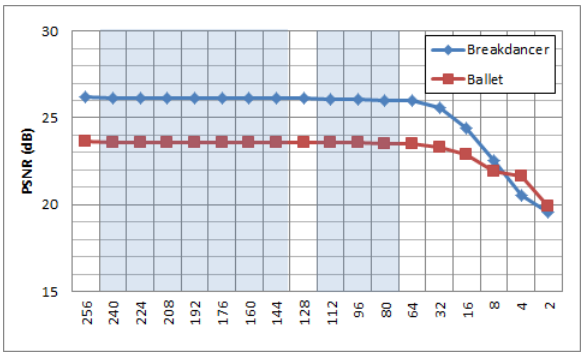

Fig. 11. Average PSNR by fine-scale depth map quantization levels for Breakdancer and Ballet scenes

on the DIBR intermediate images, and hence, the average PSNR was more degraded in these sequences. Overall, the average PSNR of the original camera-captured images and the DIBR intermediate images with an 8-bit (256 scales) depth map at five camera viewpoints was 26.23 dB for the Breakdancer sequences and 23.62 dB for the Ballet sequences.

A three-way ANOVA test was conducted to explore the impact of scene, camera viewpoint, and depth map quantization on PSNR. All three main effects were significant ($p = 0.000$). Similar to the CG scenes, the image quality degraded when the camera viewpoint moved away from the center to either side (approximately a 2~3 dB difference between 2- or 6-view and 3- or 5-view) for both sequences. The PSNR of Breakdancers was approximately 2.5 dB higher than the PSNR for Ballet when the depth map quantization was over 64-scale. The mean difference was gradually reduced with lower quantization levels. In addition, post-hoc comparisons of the depth map quantization revealed that the PSNR from a 256-scale to 96-scale depth map was not significantly different for the Breakdancer sequences (mean difference < 0.18 dB). In the Ballet scene, the PSNR from a 256- to 64-scale depth quantization did not differ (mean difference < 0.10 dB).

4. Discussions and Conclusions

In this paper, we investigated how the quantization levels of a depth map may affect the image quality of multiview intermediate images generated using DIBR in order to find suitable per-pixel depth bit information for 3D display systems. In prior works, the image quality of DIBR had been improved by more elaborate depth map estimation. Hence, much work has been done on pre-processing the depth map to improve the DIBR process, such as applying a smoothing filter, disparity compensation, up-sampling the low-resolution depth, depth edge enhancement, etc [10][11][12][13][14][15][16]. In this research, we conducted an experiment to evaluate which depth map quantization levels are sufficient for DIBR multiview rendering without the need for depth map pre-processing.

This experiment first evaluated the image quality of DIBR on the initial 16-bit depth map for four computer-generated 3D scenes (Farm, Zebra, Car, and Treasure) and a 5-view camera array aligned in two camera alignments (parallel versus convergent). Then, the image quality was measured from the PSNR value between the original 5-view images taken directly from the 3D scene and the 5-view intermediate images generated by DIBR. The results showed that the PSNR was affected by scene, camera viewpoint, and camera alignment. Overall, the quality of DIBR with the initial 16-bit depth map was significantly degraded by scenes

containing foreground objects located closer to the viewer. The PSNR was also degraded by the camera viewpoints, especially in scenes with repeated texture patterns. These results are similar to those of Do et al [6], showing that the overall quality of DIBR is highly dependent on the complexity of the scene.

Given this fact, we further investigated the image quality of DIBR while varying the depth map quantization levels from a 16-bit (65536-scale) to a 1-bit (2-scale) depth for all scenes and camera alignments. The overall results showed that the PSNR was significantly degraded by the depth map quantization levels for each scene. Post-hoc comparisons revealed that the image quality of DIBR with a 6- to 7-bit depth map, or above (depending on the scene), did not significantly decrease from the image quality of DIBR with an initial 16-bit depth map. It appeared that the quality of DIBR was more affected by depth map quantization levels in scenes consisting of more repeated texture patterns. In such scenes, the additional effect of depth map quantization and camera viewpoint interaction should also be considered. The quality analysis with different texture quality and foreground object complexity CG [17] also indicated that a high depth map quantization level was necessary in order to enhance the quality of DIBR.

In our detailed analysis of fine-scale depth map quantization levels from 256-scale (8-bit depth) to 32-scale (5-bit depth) for the four computer-generated 3D scenes, the results showed that at least a 96-scale depth map was required to avoid significant image quality degradation in DIBR. The detailed analysis of fine-scale depth map quantization for real scenes (such as Breakdancer and Ballet scenes) produced results similar to those of the computer-generated 3D scenes. That is, the PSNR from 256- to 96-scale depth did not differ for the Breakdancer sequence and the PSNR from 256- to 64-scale depth did not differ for the Ballet sequence. However, in these sequences, the overall PSNR was much lower than the computer-generated 3D scenes, due to larger holes that appeared on the DIBR intermediate images.

Hence, we conclude that the depth map generated by disparity estimation, or directly acquired by a depth camera should maintain 96-scale or higher quantization levels to generate a qualified DIBR 5-view intermediate image. Also, the image quality of DIBR by depth map quantization is more sensitive to the complexity of the scene (such as the location of foreground objects and the texture patterns). In the future, we plan to conduct a systematic evaluation of how DIBR multiview image generation is affected by the depth map quantization levels for scenes of varying complexity. We will also evaluate the subjective evaluation of the image quality for DIBR multiview intermediate images of 3D scenes displayed on a multiview 3D display, such as the Philips WOW 3DTV.

References

[1] C. Fehn, "Depth-Image-Based Rendering (DIBR), Compression and Transmission for a New Approach on 3D-TV," in *Proc. of SPIE Stereoscopic Display and Virtual Reality Systems XI*, vol. 5291, pp. 93-104, Jan. 19-21, 2004. Article (CrossRef Link)

[2] C. Fehn, "A 3D-TV Approach Using Depth-Image-Based Rendering (DIBR) ," in *Proc. of Visualization, Imaging, and Image Processing*, pp.482-487, Sep. 8-10, 2003.

[3] W.R. Mark, "Post-Rendering 3D Image Warping: Visibility, Reconstruction and Performance for Depth-Image Warping," *PhD thesis*, University of North Carolina at Chapel Hill, Apr. 21, 1999.

[4] L. McMillan Jr., "An Image-Based Approach on Three-Dimensional Computer Graphics," *PhD thesis*, University of North Carolina at Chapel Hill, 1997.

[5] K.-J. Oh, S. Yea, Y.-S. Ho, "Hole-Filling Method Using Depth Based In-Painting for View

Synthesis in Free Viewpoint Television (FTV) and 3D Video," in *Proc. 27th Conference on Picture Coding Symposium*, pp. 233-236, May 6-8, 2009.
[6] L. Do, S. Zinger, Peter H.N. de With, "Quality Improving Techniques for Free-Viewpoint DIBR," in *Proc. of SPIE Stereoscopic Displays and Applications XXI*, vol. 7524, Jan. 18-20, 2010. Article (CrossRef Link)
[7] R. Gvili, A. Kaplan, E. Ofek, G. Yahav, "Depth keying," in *Proc. of SPIE Electronic Imaging*, vol. 5006, pp. 564–574, Jan. 20-24, 2003. Article (CrossRef Link)
[8] M. Kawakita, K. Iizuka, H. Nakamura, I. Mizuno, T. Kurita, T. Aida, Y. Yamanouchi, H. Mitsumine, T. Fukaya, H. Kikuchi, F. Sato, "High-definition real-time depth-mapping TV camera: HDTV Axi-Vision Camera," *Optics Express*, vol. 12, no. 12, pp. 2781-2794, 2004. Article (CrossRef Link)
[9] P. Fua, "A Parallel Stereo Algorithm that Produces Dense Depth Maps and Preserves Image Features," *Machine Vision and Applications*, vol. 6, no. 1, pp. 35-49, 1993. Article (CrossRef Link)
[10] L. Zhang, W.J. Tam, "Stereoscopic Image Generation Based on Depth Images for 3D TV," *IEEE Transactions on Broadcasting*, vol. 51, no. 2, pp. 191-199, 2005. Article (CrossRef Link)
[11] J. Lee, C. Kim, "Stereoscopic Image Generation with Optimal Disparity using Depth Map Preprocessing and Depth Information Analysis," *Journal of Broadcast Engineering*, vol. 14, no. 2, pp. 164-177, 2009.
[12] M. Magnor, P. Eisert, B. Girod, "Multi-View Image Coding with Depth Maps and 3-D Geometry for Prediction," in *Proc. of SPIE: Visual Communications and Image Processing*, pp.263-271, Jan. 20-26, 2001. Article (CrossRef Link)
[13] Z. Ni, D. Tian, S. Bhagavathy, J. Llach, B.S. Manjunath, "Improving the quality of depth image based rendering for 3D video systems," in *Proc. IEEE The 16th International Conference on Image Processing*, pp. 513-516, Nov. 7-10, 2009. Article (CrossRef Link)
[14] Q. Yang, R. Yang, J. Davis, D. Nistér, "Spatial-Depth Super Resolution for Range Images," in *Proc. of IEEE Computer Vision and Pattern Recognition*, pp.1-8, June 18-23, 2007. Article (CrossRef Link)
[15] Q.H. Nguyen, M.N. Do, S.J. Patel, "Depth Image-Based Rendering with Low Resolution Depth," in *Proc. of IEEE International Conference on Image Processing*, pp. 553-556, Nov. 7-10, 2009. Article (CrossRef Link)
[16] W.J. Tam, F. Speranza, L. Zhang, R. Renaud, J. Chan, C. Vazquez, "Depth Image Based Rendering for Multiview Stereoscopic Displays: Role of Information at Object Boundaries," in *Proc. of SPIE*, vol. 6016, pp. 75-85, 2005. Article (CrossRef Link)
[17] L. Do, S. Zinger, P.H.N. de With, "Objective Quality Analysis for Free-Viewpoint DIBR," in *Proc. of IEEE International Conference on Image Processing*, pp. 2629-2632, Sep. 26-29, 2010. Article (CrossRef Link)
[18] G. Leon, H. Kalva, B. Furht, "3D Video Quality Evaluation with Depth Quality Variations," in *Proc. of IEEE 3DTV Conference*, pp. 301-304, May 4-6, 2009. Article (CrossRef Link)
[19] I. Ideses, L. Yaroslavsky, I. Amit, B. Fishbain, "Depth Map Quantization - How much is sufficient?," in *Proc. of IEEE 3DTV Conference*, pp.1-4, May 7-9, 2007. Article (CrossRef Link)
[20] S. Zinger, L. Do, P.H.N. de With, "Free-viewpoint depth image based rendering," *Journal of Visual Communication and Image Representation*, vol. 21, no. 5-6, pp. 533-541, 2010. Article (CrossRef Link)
[21] A. Woods, T. Docherty, R. Koch, "Image Distortions in Stereoscopic Video Systems," in *Proc. SPIE Stereoscopic Displays and Applications IV*, vol. 1915, pp. 36-48, Feb. 1-2, 1993. Article (CrossRef Link)
[22] OpenGL.org, http://www.opengl.org/resources/faq/technical/depthbuffer.htm Why is there more precision at the front of the depth buffer?
[23] Independent JPEG Group, "JPEG library", http://www.ijg.org/
[24] C.L. Zitnick, S.B. Kang, M. Uyttendaele, S. Winder, R. Szeliski, "High-quality video view interpolation using a layered representation," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 600-608, Aug. 2004. Article (CrossRef Link)

Minyoung Kim is currently working on a Ph. D. degree at the Department of Computer Science, Graduate School, Sangmyung Univerity, Seoul, Korea. She got her Bachelor's degree at the Division of Digital Media, College of Software and her Master's degree at the Department of Computer Science, Graduate School, Sangmyung University, Seoul, Korea. Her expertise is in distributed 3D display, tiled systems and virtual reality.

Yongjoo Cho is currently an associate professor at the Division of Digital Media, College of Software, Sangmyung Univerity in Korea. He earned his Bachelor of Science in Computer Science at University of Illinois at Urbana-Champaign and Masters' of Science and Ph. D. degrees at University of Illinois at Chicago. He currently is the director of the Interactive Computing and Entertainment Laboratory at Sangmyung University. His research interest includes Virtual Reality, 3D Display, Tiled Display System, Computer Supported Cooperative Work, and Virtual Learning Environments.

Hyon-Gon Choo received his B.S. and M.S. degree in electronic engineering in 1998 and 2000 respectively, and his Ph.D degree in electronic communication engineering in 2005 from Hanyang University, Seoul Korea. He has been a Senior Member of Engineering Staff with the Broadcasting Media Research Group at Electronics and Telecommunications Research Institute (ETRI), Daejeon, Korea. His research interests include content-based audio/video analysis, multimedia protection and 3D broadcasting technologies.

Jinwoong Kim is currently director of Broadcasting &Telecommunications Convergence Media Department, ETRI in Korea. He earned his Bachelor and Master degrees in Electronics Engineering at Seoul National University of Korea, and Ph. D. degree at Texas A&M University at College Station, TX. His research interest includes audio and video signal processing and compression technologies for digital broadcasting system, especially 3DTV and UHDTV.

Kyoung Shin Park is an assistant professor at the Department of Multimedia Engineering, Dankook Univerity in Korea. She acquired her Ph. D. in Computer Science at the University of Illinois at Chicago (UIC) in 2003, with an emphasis in human computer interaction and graphics. She had also worked as a research professor at Digital Media Laboratory at Information & Communications University until she joined at Dankook University and was a visiting scholar at MIT Media Laboratory in 2004. Her research interest includes human computer interaction, virtual reality, augmented reality, collaborative environments, tiled display systems, multimedia systems and 3D imaging.