Virtual Control of Optical Axis of the 3DTV Camera for Reducing Visual Fatigue in Stereoscopic 3DTV

Jong-Il Park, Gi Mun Um, Chunghyun Ahn, and Chieteuk Ahn

In stereoscopic television, there is a trade-off between visual comfort and 3-dimensional (3D) impact with respect to the baseline-stretch of a 3DTV camera. It is necessary to adjust the baseline-stretch at an appropriate distance depending on the contents of a scene if we want to obtain a subjectively optimal quality of an image. However, it is very hard to obtain a small baseline-stretch using commercially available cameras of broadcasting quality where the sizes of the lens and CCD module are large. In order to overcome this limitation, we attempt to freely control the baseline-stretch of a stereoscopic camera by synthesizing the virtual views at the desired location of interval between two cameras. This proposed technique is based on the stereo matching and view synthesis techniques. We first obtain a dense disparity map using a hierarchical stereo matching with the edge-adaptive multiple shifted windows. Then, we synthesize the virtual views using the disparity map. Simulation results with various stereoscopic images demonstrate the effectiveness of the proposed technique.

Keywords: Stereoscopic TV, stereo matching, visual fatigue, view synthesis.

I. Introduction

Giving a sense of 3-dimensional (3D) vision plays a key role in realizing a highly realistic display system. Recent interest in 3D HDTV reflects the human desire for a more realistic display [1].

One of the key issues in realizing 3DTV is to reduce visual fatigue while maintaining sufficient 3D reality. A binocular 3DTV camera, which is mostly used in a conventional 3DTV system, consists of two cameras. A well-known problem in a conventional 3DTV system is that visual comfort degrades while the 3D impact enhances when increasing the baseline-stretch (defined as the distance between the two cameras) of the 3DTV camera. Recently, it was reported in many subjective tests using several outdoor scenes [2] that one can obtain an optimal baseline-stretch at about the average distance between human pupils (about 65 mm).

However, the lens diameter of a commercial TV camera for broadcasting purpose is usually longer than 65 mm. Therefore, the baseline-stretch of a 3DTV camera is unavoidably wider than the desirable distance.

The stereoscopic images taken by a 3DTV camera with a wide baseline-stretch not only give an uncomfortable feeling but also tend to show some subjectively undesirable effects such as a "puppet theater effect" when they are displayed in a stereoscopic display [3]. Thus, new virtual-view images are necessary, as if we took them using a 3DTV camera with a small baseline-stretch. By doing this, we can virtually control the optical axis of a 3DTV camera.

In this paper, we propose a novel optical axis control technique that controls the position of optical axis of a 3DTV camera by synthesizing the virtual views between two TV cameras. Figure 1 illustrates the concept of the proposed system.

Manuscript received Apr. 22, 2003; revised Oct. 05, 2004.

This work was supported in part by the research fund of Hanyang University.

Jong-Il Park (phone: +82 2 2290 0368, email: jipark@hanyang.ac.kr) is with the Division of Electrical and Computer Engineering, Hanyang University, Seoul, Korea.

Gi Mun Um (email: gmum@etri.re.kr), Chunghyun Ahn (email: hyun@etri.re.kr), and Chieteuk Ahn (email: ahnc@etri.re.kr) are with Digital Broadcasting Research Division, ETRI, Daejeon, Korea.

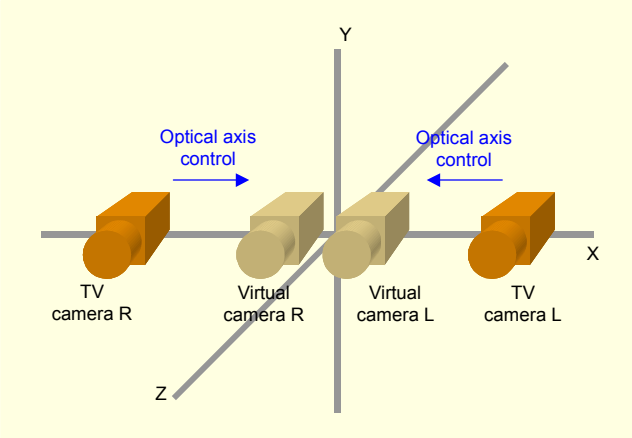

Fig. 1. Virtual control of the optical axis of a 3DTV camera.

The proposed optical axis control technique consists of two steps: a hierarchical stereo matching with an edge-adaptive shifted matching window, and a virtual view synthesis. In section II, we first describe the hierarchical stereo matching algorithm with an edge-adaptive shifted matching window scheme for obtaining a disparity map. Then, we explain a novel view synthesis algorithm using the disparity map in section III. Our simulation results for various stereoscopic images and discussions are described in section IV. Finally, conclusions and future works are presented in section V.

II. Stereo Matching with the Edge-Adaptive Shifted Matching Window

It is very important to obtain a dense and accurate disparity map from stereo images in order to synthesize high quality virtual view images. We propose a coarse-to-fine hierarchical scheme along with an edge-adaptive shifted matching window. The algorithm assumes parallel cameras. When the optical axes of the cameras are not parallel, they can be made parallel in most of the stereo setups assumed in this paper by applying a rectification algorithm [4].

Stereo matching is a process to find corresponding points between reference and target images. The disparity is defined as the position difference (usually a horizontal position difference with epipolar aligned images). Therefore, it is also called disparity estimation.

There has been great progress in the field of dense stereo matching for the last 10 years. A good review can be found in [5]. Extensive work has been exerted on the real-time extraction of dense disparity maps from multiple camera views. As a result, some camera systems producing dense disparity maps in real-time are commercially available [6]. On the other hand, a sophisticated approach has been proposed to enhance the quality of disparity estimation. The multi-view global optimization technique [7] and bi-directional Bayesian technique [8] could give better disparity maps at the expense of huge computation. However, they are not well-suited for practical applications due to their complexity. Moreover, the computational structures of such methods are not regular. Therefore, we try to develop a computationally efficient and regular method so that it can be easily implemented with hardware for real-time processing.

Disparity estimation around occlusion boundaries has a tendency to give erroneous disparity values because an occlusion area within a window affects the accuracy of disparity values for the center point of the matching window. This is a main drawback of conventional stereo matching with local support. Thus, we need to exclude occlusion areas adaptively in the process of stereo matching. To do this, we use spatially shifted matching windows

(a) Left-shifted

(b) Centered

(c) Right-shifted

Fig. 2. Three kinds of shifted matching windows.

[7], [9]. As shown in Fig. 2, there are three kinds of shifted matching window: left-shifted, centered, and right-shifted.

The disparity of each pixel is calculated as follows. We first calculate the sum of the square difference for three types of shifted matching window along the epipolar line. Then we regard the displacement that gives the smallest sum of the square difference as the disparity of the pixel.

It is interesting to notice that the region without vertical edges hardly includes the occlusion area in the usual stereo setup. Thus, there is no need to apply all of the above three kinds of windows to all pixels. Instead, we apply only the centered window if there is no edge in a window. Otherwise, we apply all three types of matching window for the stereo matching.

The advantage of using an edge-adaptive shifted window technique is the reduction of computational complexity. A simple edge detection using a Sobel operator requires negligible computation compared with multiple matching. Therefore, the computational complexity for non-edge areas can be reduced up to about 1/S of those using non-adaptive shifted windows, where S is the number of shifted matching windows. The qualities of both disparity maps obtained by the two techniques are almost the same while the computational complexity of the edge-adaptive scheme is reduced considerably, as described in section IV.

Moreover, we use the coarse-to-fine hierarchical matching scheme that successively propagates the results of a coarse layer to its neighboring layer in a resolution pyramid. In this way, we attempt to maximize the qualities of stereo matching results while maintaining the computational complexity to an acceptable level.

In local stereo matching, we assume that at least one window out of the three types of window include a common image part observable from both cameras, and can thus give correct disparity values. However, this may not be true for the occlusion areas. Thus, we detect the occlusion areas at the final step of obtaining disparity maps, as described below.

To check the validity of the disparity values, two disparity maps are obtained for both left and right views. Then, we compare the disparity of a pixel in one image with that of the corresponding pixel in the other image over its entirety. If the disparity of the pixel differs by more than one pixel, the pixel is considered as an occluded one. Otherwise, it is considered as a non-occluded pixel. In this way, we can detect the occlusion areas. Then, the disparities of the occlusion areas are directionally interpolated along the scan-line from the left for the left view and from the right for the right view under the assumption that occlusion areas belong to the background.

III. Synthesis of Virtual View Images

With the disparity map obtained by the stereo matching described in section II, an arbitrary virtual view image can be synthesized using the following procedures: forward mapping, disparity map interpolation for uncovered areas, backward mapping and finally, spatial interpolation of the virtual view image. Each procedure is conceptually illustrated in Fig. 3. This technique is a modification of the multi-view synthesis method [10] in order to apply it to 2-view images.

We first determine the target position of virtual view images on the basis of the measured baseline-stretch of a real 3DTV camera. Once the target position of the virtual 3DTV camera is

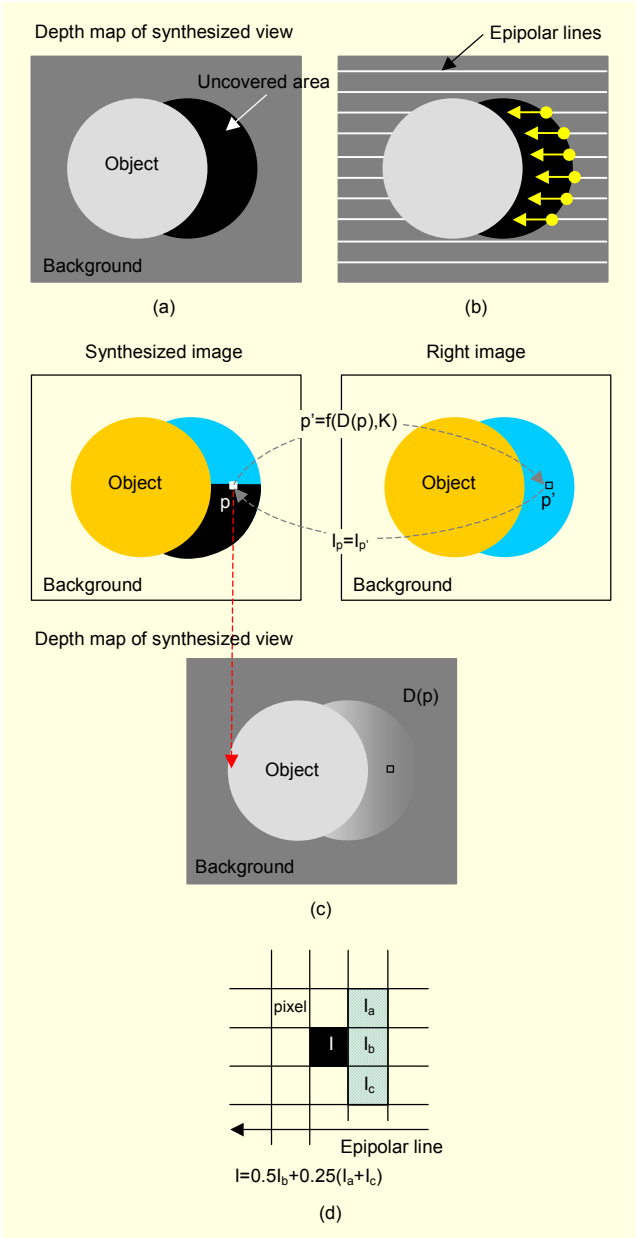

Fig. 3. An illustration of the view synthesis procedure: (a) forward mapping, (b) directional interpolation of the disparity map along the epipolar line, (c) backward mapping using the interpolated disparity, and (d) spatial interpolation.

determined, a forward mapped image can be readily obtained from original stereo images.

In forward mapping [11], uncovered areas appear due to the difference of the disparities. How to fill out these areas is of primary concern in the view synthesis. We try to obtain the colors of the areas by guessing the disparities of these areas on the basis of the reasonable assumption that uncovered areas belong not to foreground objects but to background areas. The disparities of background areas are directionally propagated to interpolate the uncovered areas along the scan-line, as shown in Fig. 3(b).

Then, the color information I_p for each pixel of the uncovered area is fetched from the other image (e.g., left image for a virtual right image and vice versa) using a guessed disparity D_p, as shown in Fig. 3(c).

In this backward mapping, we test the validity of the mapping. If the target point in the original image belongs to a detected occlusion area, we consider the mapping as valid and take the color information from there. Otherwise, we consider the backward mapping as invalid and interpolate the color of the pixel spatially using the neighboring pixels, as shown in Fig. 3(d).

IV. Simulation Results and Discussions

The performance of the proposed technique presented in the preceding sections is verified by applying it to various stereoscopic images ranging from well-known database images (3D image databases of Univ. of Tsukuba and ITE Japan) to indoor and outdoor images taken by a set of 3DTV cameras with broadcasting quality. We implemented the proposed technique using Microsoft Visual C++.

Figure 4 shows the matching and view synthesis results for the "City" image. The size of each image is 640×480 pixels. The maximum disparity of the "City" image is less than 45 pixels. Figures 4(a) shows the original left and right images. Figure 4(b) shows the disparity map obtained by the conventional hierarchical block matching method without using shifted windows i.e., with single window. A five-layer pyramid is constructed for the hierarchical matching, and the size of the matching window in each level is set to 3×3. On the right are some parts of the synthesized virtual left view image where the baseline stretch is reduced to 40% of the original length. We see some noisy reconstruction errors in the synthesized virtual view image. Figure 4(c) shows the disparity map obtained by using the proposed edge-adaptive shifted windows, and the virtual view image using the disparity map. Comparing Figs. 4(b) and 4(c), we can see a noticeable improvement of the disparity map by the use of shifted windows. When comparing the virtual views, we can also see that the proposed matching window dramatically improves the quality of the synthesized virtual view image. Figures 4(d)

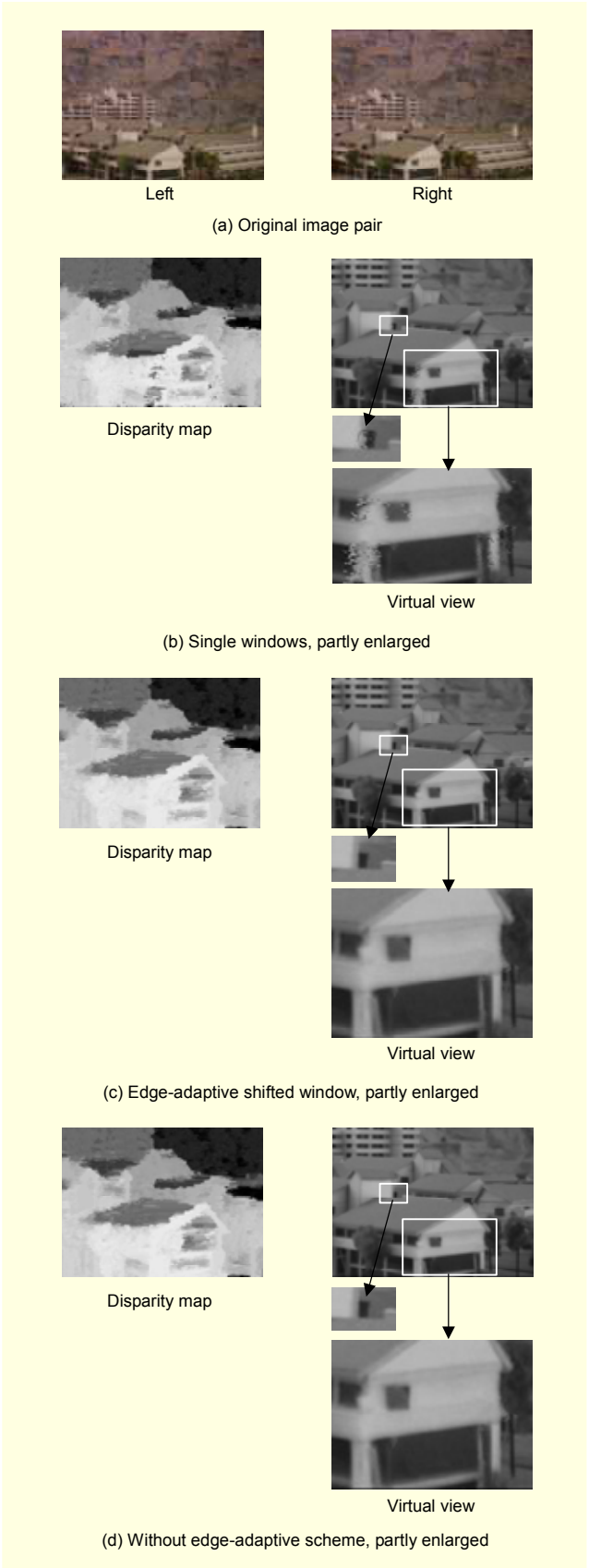

Fig. 4. Simulation results for the "City" images (Univ. of Tsukuba).

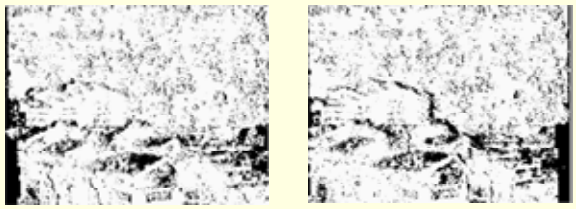

Fig. 5. Detected occlusion areas.

shows the disparity map and synthesized virtual view image, obtained by using a shifted matching window without an edge-adaptive scheme. Comparing Figs. 4(c) and 4(d), it is hard to see the difference between them. These results imply that the proposed stereo matching with the edge-adaptive shifted matching windows significantly reduces computational complexity without sacrificing the quality of the synthesized virtual views. We have observed similar tendencies in the results of applying the proposed technique to various stereo images.

It is interesting to notice that the quality of the disparity map is relatively poor in textureless areas but the quality of the synthesized images is not degraded, which convinces us of the effectiveness of the proposed technique.

Figure 5 shows the detected occlusion areas using the bidirectional check of the disparity maps for the images in Fig. 4. We see that most of the important occlusion areas are correctly detected although there are some noisy areas, especially in textureless regions.

To evaluate the performance of the proposed technique, we tried to quantitatively measure the quality of the computed disparity map. We compute the following quality measure based on known ground truth data. The percentage of badly matched pixels [5] is defined by

$$B = \frac{1}{N}\sum_{(x,y)} \left(\left| d_{est}(x,y) - d_{true}(x,y) \right| > \delta_d \right) \times 100 ,$$

where δ_d is a disparity error tolerance. For the experiments in this paper, we use $\delta_d = 1.0$.

Figure 6 shows the error between the ground truth map and the estimated disparity map using the proposed technique.

(a) Original image pair

(b) Error compared with the ground truth map: window size 3 x 3

(c) Error compared with the ground truth map: window size 5 x 5

Fig. 6. Errors of disparity map compared with the ground truth map. The badly matched pixels are shown in black.

Table 1. Percentage of badly matched pixels.

	Window size: 3×3	Window size: 5×5
Single window	29.1%	30.5%
Edge-adaptive shifted window	25.1%	18.1%
Shifted window without edge-adaptive scheme	25.1%	17.9%

Table 1 shows the error percentages of the estimated disparity maps. The size of the original image is 384×288 pixels, and the effective area compared with the ground truth map is 348×252 pixels (cut off by 18 pixels from left, right, top, and bottom). Figures 6(a) shows the original left and right images. The "face" images are provided by Univ. of Tsukuba. A five-layer pyramid is constructed for the hierarchical block matching, and the size of the matching window in each level is set to 3×3. The searching range of the disparity is set to [0, 16] pixels. Figure 6(b) shows the error of a disparity map obtained by a hierarchical block matching method using a single window on the left. The black areas represent badly matched pixels. The errors of disparity maps obtained by using the proposed edge-adaptive shifted windows and by using a shifted matching window without an edge-adaptive scheme are also shown. Figures 6(c) shows the results when the matching window size is 5×5. Comparing the results of single window and shifted window, we see noticeable improvements, especially in object boundaries. The error percentage is reduced by 4% for a 3×3 window and 12.4% for a 5×5 window. When comparing the middle and the right images in Figs. 6(b) and 6(c), it is hard to discriminate the difference between them. The computational cost of the edge-adaptive scheme is in between the single-window matching and multiple-window matching depending on the edge distribution of the image. In this case, the CPU times for a single-window, edge-adaptive 3-window, and 3-window matching are 1.89, 2.59, and 3.59 seconds, respectively, on a Pentium-IV 1.7 GHz Windows machine. These results demonstrate that it is reasonable to use the proposed edge-adaptive shifted windows for the reduction of computational complexity.

We also tested the proposed technique in the case of the indoor and outdoor stereoscopic scenes.

Figure 7 shows the results for the "Dog and Flower" stereoscopic image. The baseline stretch L of the 3DTV camera is 110 mm. The size of the image is 720×480 pixels. The maximum difference of the disparity is within 30 pixels. Figures 7(a) shows the original left and right images. The rectified and shifted images for stereo matching using an optimum rectification algorithm [4] are shown in Fig. 7(b). Figures 7(c) shows synthesized virtual view images where the

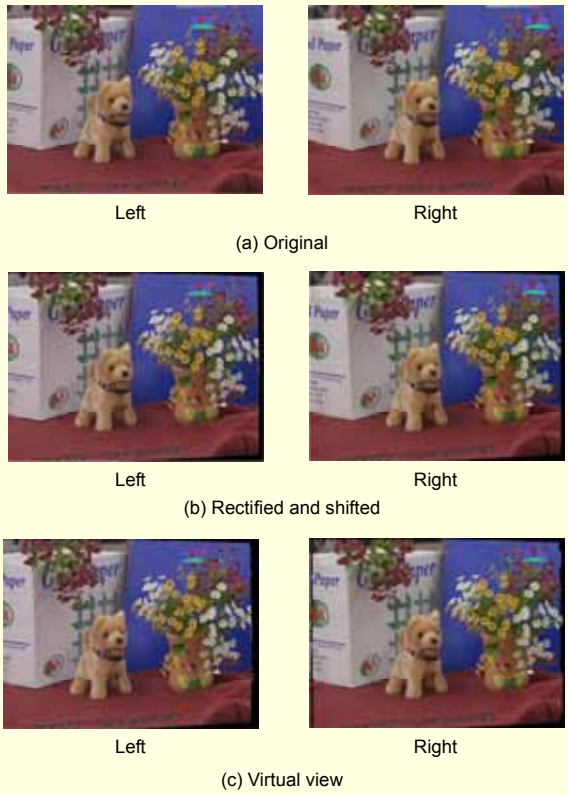

Fig. 7. Simulation results for the "Dog and Flower" indoor scene.

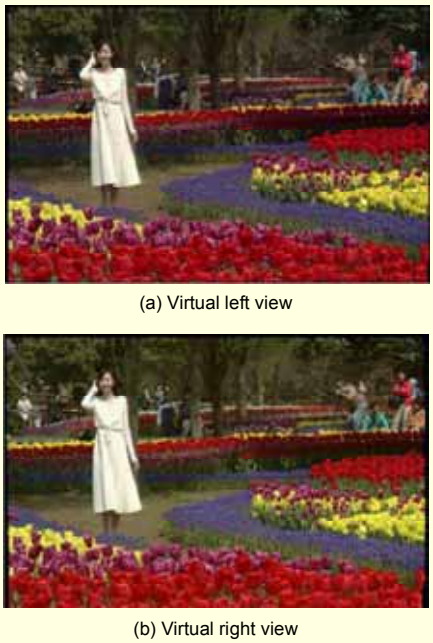

Fig. 8. Simulation results for the outdoor stereoscopic HD scene.

baseline stretch is reduced to 40% of the original one.

Figure 8 shows the results for the "Flower Garden" scene shot in stereoscopic HD (provided by ITE, Japan). The size of the image is 1920×1080 pixels. The baseline stretch L is 65 mm. The maximum difference of the disparity is within 70 pixels. Figures 8(a) and 8(b) are generated virtual stereo images where the baseline-stretch is reduced to 40% of the original one. We see the quality of the images is quite acceptable.

From the above results and other experiments not reported here, the disparity maps obtained by the proposed method using an edge-adaptive shifted matching window scheme show consistently sharper and more correct edges at object boundaries compared to those of conventional methods. The virtual views using the enhanced disparity maps naturally show a more compelling visual quality.

Even with all of the improvements, there still are many disparity errors due to the inherent limitation of 2-view area-based stereo matching. These errors inevitably produce some artifacts around occlusion areas. However, it is interesting to point out that the small artifacts around these occlusion areas are not very perceptible if the images are shown in the stereoscopic display, which justifies the effectiveness of the proposed method.

As a final step of our test, we performed a subjective evaluation in the laboratory. In this test, the original images (baseline stretch = L) and various synthesized images (baseline stretch = 0.8 L, 0.6 L, 0.4 L, 0.2 L, and 0) are displayed on a stereoscopic display (VRJoy shutter glass on a 21" CRT monitor). A total of five subjects were asked to score their opinions from 1 to 5 in terms of comfortableness, 3D impact, image quality, and preference among the original and processed images. Figures 9 and 10 show the average scores of the subjective evaluation for the "Dog and Flower" indoor scene and "Flower Garden" outdoor scene, respectively. The results show that comfortableness enhances while the 3D impact decreases with a decrement of the baseline-stretch, as is reported in [2]. The quality of the image degrades as the baseline stretch decreases, i.e., as the area of interpolated pixels increases. This is due mainly to the errors in disparity estimation. However, it is interesting to notice that most of the preferred images are not the original stereoscopic images but the synthesized stereoscopic images, as we see in Figs. 9 and 10. This convinces us of the usefulness of the proposed approach that virtually controls the optical axis of a stereo camera system.

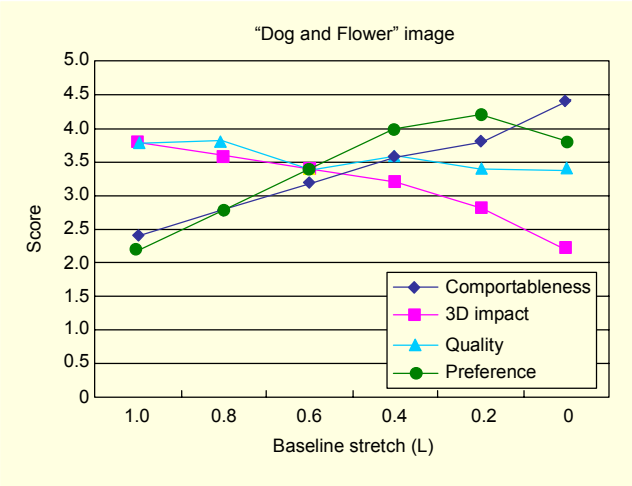

Fig. 9. The average scores of our subjective evaluation for the indoor stereoscopic scene.

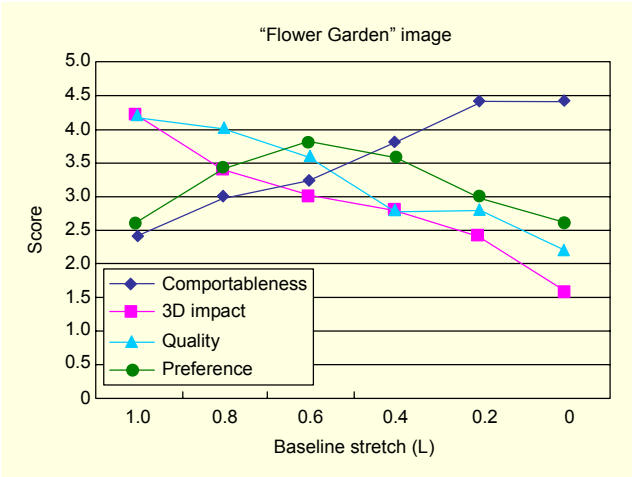

Fig. 10. The average scores of our subjective evaluation for the outdoor HD stereoscopic scene.

V. Conclusions and Future Works

We have presented a virtual optical axis control technique for a 3DTV camera by synthesizing virtual views at the desired positions of two cameras in order to reduce visual fatigue.

We first obtain a dense disparity map using a hierarchical implementation of edge-adaptive shifted stereo matching. Then, virtual views are efficiently synthesized using the disparity map. The proposed method is verified by computer simulation. The simulation results for various stereoscopic images showed visually compelling virtual views. The effectiveness of the proposed optical axis control scheme is demonstrated by a subjective evaluation.

The virtual control of the optical axis is a first step toward a versatile framework of post-processing for providing comfortable stereoscopic contents. It would be interesting as a future work to analyze various mutually affecting factors on visual comfortableness and 3D impact in stereoscopic images, such as focal length, convergence angle, resolution, and the distance to main objects using a more rigorous subjective assessment with a sufficient number of subjects evaluating a large set of stereoscopic images.

References

[1] N. Hur, G. Lee, W. You, J. Lee, and C. Ahn, "An HDTV-Compatible 3DTV Broadcasting System," *ETRI J.*, vol. 26, no. 2, Apr. 2004, pp. 71-82.

[2] L. Stelmach, W. Tam, F. Speranza, A. Vincent, and G. Um, "Visual Fatigue Reduction Techniques for Binocular 3D System," *Final Report of CRC/ETRI Collaborative Research*, June 2002.

[3] H. Yamanoue and I. Yuyama, "The Relation Between Size Distortion and Pick-Up Conditions For Stereoscopic Images," *Journal of ITE*, vol. 48, no.10, Oct. 1994 (in Japanese), pp. 116-124.

[4] J. Gluckman and S. Nayar, "Rectifying Transformations That Minimize Resampling Effects," *Proc. of IEEE CVPR 2001*, vol. I, Hawaii, Dec. 2001, pp. 111-117.

[5] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," *Int'l J. Computer Vision,* vol. 47, issue 1-3, 2002, pp. 7-42.

[6] http://www.ptgrey.com.

[7] S. Kang, S. Szeliski, and J. Chai, "Handling Occlusions in Dense Multi-View Stereo," *Proc. of IEEE CVPR 2001*, vol. I, Hawaii, Dec. 2001, pp. 103-110.

[8] P. Lim, A. Das, and M. Chong, "Estimation of Occlusion and Dense Motion Fields in a Bi-directional Bayesian Framework," *IEEE Trans. PAMI*, vol. 24, no. 5, May 2002, pp. 712-718.

[9] A. Bobick and S. Intille, "Large Occlusion Stereo," *Int'l J. Computer Vision*, vol. 33, no. 3, Sept. 1999, pp. 181-200.

[10] J. Park and S. Inoue, "Arbitrary View Generation Using Multiple Cameras," *Proc. IEEE ICIP'97*, vol. I, Santa Barbara, USA, Oct. 1997, pp. 149-153.

[11] G. Wolberg, *Digital Image Warping*, IEEE Computer Society Press, 1990.

Jong-Il Park is an Associate Professor in Division of Electrical and Computer Engineering, Hanyang University, Seoul, Korea. He received the BS, MS, and PhD degrees all in electronics engineering from Seoul National University, Seoul, Korea, in 1987, 1989, and 1995. He was a research student of University of Tokyo and a visiting researcher in NHK Science and Technology Research Laboratories, Japan, from 1992 to 1994. After working for Korean Broadcasting Institute in 1995, he joined ATR Media Integration and Communications Research Laboratories, Japan, in 1996 where he was involved in various projects on video analysis and processing, 3D video processing, and virtual reality. Since 1999, he has been with Hanyang University. His research interest includes computer graphics, computer vision, video processing, and virtual reality.

Gi Mun Um received the BS, MS, and PhD degrees in electronic engineering from Sogang University, Seoul, Korea, in 1991, 1993, and 1998. He joined Electronics and Telecommunications Research Institute (ETRI) in 1998, and he is currently with the Digital Broadcasting Research Division. He has worked as a visiting research scientist at Communications Research Centre Canada (CRC) from 2001 to 2002. His main research interests are computer vision, image-based 3D modeling/rendering (IBMR), 3DTV broadcasting system, and visual fatigue reduction technique in stereoscopic video systems.

Chunghyun Ahn obtained his PhD in remote sensing & GIS from the Chiba University of Japan. He joined ETRI in 1996, and he is currently with the Digital Broadcasting Research Division. He has interests in the fields of integration of remote sensing and GIS, 3DTV broadcasting systems, and the integration of broadcast- and location-based services in DMB systems.

Chieteuk Ahn is Vice-President and a Principal Member of the Technical Staff in the Digital Broadcasting Research Division of ETRI, Korea. He received the BS and MS from Seoul National University, Korea, in Feb. 1980 and 1982, and the PhD from University of Florida, USA in Aug. 1991. Since he joined ETRI in 1982 he has been involved in developing digital switching systems, MPEG standardization and broadcasting technology. His recent work has focused on MPEG technology as well as in developing interactive multimedia technology. He has served as an HOD of MPEG-Korea and SC29-Korea since 1996. His main interests are in the areas of multimedia signal processing, broadcasting and communications.