A Stereoscopic Video Generation Method Using Stereoscopic Display Characterization and Motion Analysis

Donghyun Kim, *Student Member, IEEE*, Dongbo Min, *Student Member, IEEE*, and Kwanghoon Sohn, *Member, IEEE*

***Abstract*—Stereoscopic video generation methods can produce stereoscopic content from conventional video filmed with monoscopic cameras. In this paper, we propose a stereoscopic video generation method using motion analysis which converts motion into disparity values and considers multi-user conditions and the characteristics of the display device. The field of view and the maximum and minimum disparity values were calculated in the stereoscopic display characterization stage and were then applied to various types of 3D displays. After motion estimation, we used three cues to decide the scale factor of motion-to-disparity conversion. These cues were the magnitude of motion, camera movements and scene complexity. A subjective evaluation showed that the proposed method generated more satisfactory video sequence.**

***Index Terms*—Broadcasting, image analysis, multimedia systems, three-dimensional vision.**

I. Introduction

MANY researchers have developed 3D video technology which offers stereoscopic perception of the human visual system. This technology has been used in various applications including information communication, broadcasting, medicine, education, the military, computer games, animation, CAD and so on. The 3D display devices that have been developed allow satisfactory 3D perception and maximum eye comfort. However, 3D imaging technology has not been successful in commercial applications due to several problems. One of these problems has been a lack of 3D content.

There are many ways to generate 3D content. Information can be captured with a stereoscopic camera, and 2D content can be manually converted into 3D graphics. However, these methods are expensive, time-consuming and laborious. In this paper, we propose an automatic stereoscopic conversion algorithm based on a computer vision technique.

Automatic stereoscopic conversion (2D/3D conversion) can provide various types of 3D content because it can produce this kind of content from conventional 2D videos. Fig. 1 shows the overall concept of the stereoscopic conversion system.

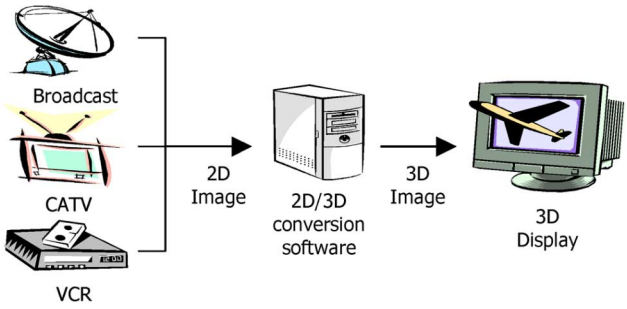

Fig. 1. Overview of the stereoscopic conversion system.

2D videos obtained from conventional broadcasting, CATV and DVDs can be converted into stereoscopic image sequences by using stereoscopic conversion technique which allows people to enjoy 3D images when using 3D display devices.

Several stereoscopic convergence algorithms have been proposed. 2D videos also contain depth perception cues such as superposition, linear perspectives, aerial perspectives, texture gradients, shadows, and motion parallax variables known as monocular cues. Stereoscopic convergence can be defined as the process of finding these monocular cues and converting them into stereoscopic cues (disparity).

Modified time difference (MTD) method detects the movements of objects and determines the delay direction and time by using the characteristics of the movements. Then, stereoscopic (left and right) images can be selected according to the time difference in the 2D image sequences [1]. The MTD method is suitable for converting images which contain simple horizontal-moving objects. However, the MTD method does not work for images that contain objects with complicated motion or those that contain no motion.

Computed image depth (CID) method uses the relative position between multiple objects in still images. The image depth is computed by using the contrast, sharpness and chrominance values of the input images [2]. Also, depth from focus method extracts depth data with a single image using blur analysis [3]. However, this method cannot be applied to all images.

The motion-to-disparity conversion method generates stereoscopic images by converting motion to disparity. This method overcomes the limitation of MTD method that convertible motion direction is restricted to horizontal direction. In order to eliminate the effect of vertical disparity values, the norm of the motion vector can be converted into horizontal disparity values

Manuscript received March 23, 2007; revised November 26, 2007. This work was supported in part by the IT R&D program of MIC/IITA 2007-F036-01 Objective quality assessment system based on human visual system] and in part by the MIC, Korea, under the ITRC support program supervised by the IITA IITA-2005-(C1090-0502-0027).

The authors are with the Electrical and Electronic Engineering Department, Yonsei University, Seoul 120-749, Korea (e-mail: dhkim@diml.yonsei.ac.kr; forevertin@diml.yonsei.ac.kr; khsohn@yonsei.ac.kr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2007.914714

[4], [5]. Motion to disparity conversion method does not calculate relative depth of scene, but it generates a depth map which is related to attentional region of scene according to human visual system.

There are several structure estimation methods using motion parallax. Structures are estimated in case that camera motion is restricted to translation [6]. In addition, various type of camera movement is considered without information about the scene and camera [7]. There was a study about the effect of matching algorithms which compare feature matching and block matching [8]. By using the extended Kalman filter [9], [10], camera motion such as rotation and translation can be estimated and the structure of scenes can be estimated in terms of the point-wise depth [11], [12]. However, these methods are limited to static video scenes.

Alternatively, a method which uses the detection of a vanishing line has been also proposed [13]. Videos can be classified into outdoor, landscape, outdoor with geometric elements and indoor categories. When vanishing points and lines are detected, the depth map can be generated according to the video type. Another method uses the sampling density of spatial temporal interpolation in human visual characteristics [14]. Some methods use the Pulfrichi effect, the time delay measured by the difference of the amount of light in both eyes [15], [16]. However, this method has not proven effective for either still images or complex images.

In this paper, we propose an automatic framework of a stereoscopic video generation system which uses the motion-to-disparity conversion method. Multi-user conditions and the characteristics of stereoscopic displays were considered for stereoscopic content generation. We also used motion analysis which calculated three cues that were used to decide the scale factor of motion-to-disparity conversion. These cues were the magnitude of motion, camera movements and scene complexity.

The rest of the paper is organized as follows. In Section II, we describe the proposed stereoscopic convergence system and discuss the algorithms of the proposed system. Experimental results and conclusions are provided in Sections III and IV, respectively.

II. Proposed Algorithm

The proposed algorithm is a general stereoscopic video generation algorithm based on the motion-to-disparity conversion method. A block diagram of the proposed stereoscopic conversion algorithm is shown in Fig. 2. It consists of four stages: stereoscopic display characterization, motion estimation, motion-to-disparity conversion and stereo generation. During the stereoscopic display characterization stage, the field of view and the maximum and minimum disparity values were determined to consider multi-user conditions and the characteristics of the display device. Motion was estimated by using a bidirectional KLT (Kanade-Lucas-Tomasi) feature tracker based on color segmentation. After motion estimation, the scale factor of motion-to-disparity conversion was determined with multiple cues. The estimated cues were the magnitude of motion, camera movements and scene complexity. Finally, stereo views were generated by using a computed depth map and the original video.

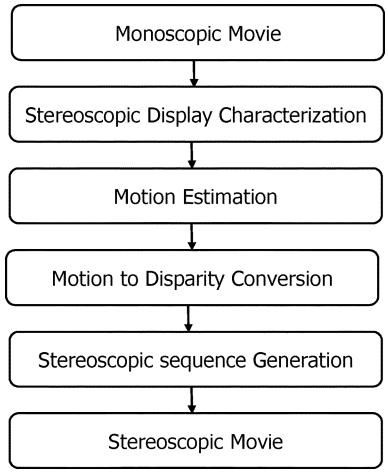

Fig. 2. Block diagram of the overall system.

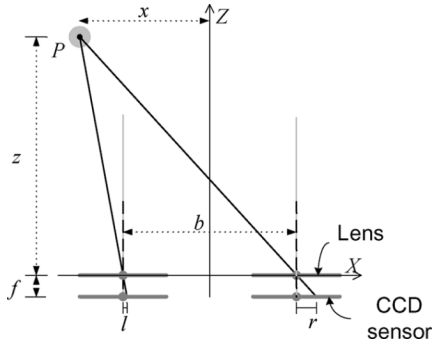

Fig. 3. Stereoscopic geometry.

A. Stereoscopic Display Characterization

In general, conversion systems must consider the circumstances in which content is displayed. These kinds of circumstances may include not only location but also the size of the audience, the illumination and sound conditions, and so on. Content generators must also decide whether a specific conversion system is a real-time or non-real time system. For example, if the location is a theater built for hundreds of people, a glass stereoscopic display device and a non-real time conversion method should be chosen. For stereoscopic PDAs (Personal Digital Assistants), a non-glass display such as a lenticular or parallax barrier should be selected, and the conversion method should be able to work in real-time. Once the types of display devices, conversion methods and number of audience are determined, we have to determine field of view, minimum and maximum disparity values. Fig. 3 shows stereoscopic geometry and Fig. 4 shows the factors that form the field of view in display devices. The field of view is generally determined by the minimum and maximum viewing distances and viewing angles. Field of view is important because of narrow viewing angle of stereoscopic display.

Human visual systems perceive depth as shown by (1).

$$Disparity(cm) = \frac{fb}{z} \tag{1}$$

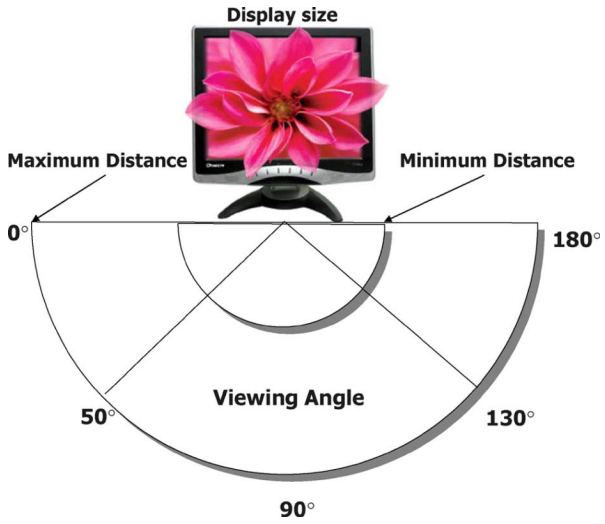

Fig. 4. Field of view of stereoscopic displays.

where f is the focal length of the camera, b is the baseline between the cameras, z is the distance between the cameras and the object, as shown in Fig. 3.

It is possible to compute the disparity pixel-to-centimeter ratio by considering the size and resolution of the display as follows:

$$Disparity(pix) = \frac{Resolution(pix)}{DisplaySize(cm)} \times Disparity(cm) \tag{2}$$

where $Disparity(pix)$ is the distance between the corresponding points l and r in image coordinate as shown in Fig. 3.

A suitable disparity value range which can enable a stereopsis fusion can be determined by the display device and human depth perception characteristics. During the stereoscopic display characterization stage, we determined the field of view and the maximum and minimum disparity values according to the location of the user, the display size and so on. The maximum and minimum disparity values were determined by adjusting various disparity values and verifying the success of stereoscopic fusion, measured in terms of each viewing angle and viewing distance. A suitable limitation of disparity was determined by a large number of participant groups because individual discrepancies existed in the stereopsis fusion process. Stereoscopic display characterization is performed one time when we choose stereoscopic display device and viewing area. Maximum and minimum disparity values and field of view have constant values for given display device and viewing area. If we change the stereoscopic display or viewing area, we have to perform stereoscopic display characterization again.

B. Motion Estimation

Stereoscopic video generation using motion-to-disparity conversion assigns depth to moving objects. Therefore, acquiring dense and accurate motion maps is an important process.

In our experiments motion maps were calculated by using color segmentation and the KLT feature tracker. A color segment-based method was used for robust estimation in textureless regions and at the boundaries of objects, and it was also used in the stereoscopic matching algorithm [17], [18]. We assumed that the level of motion was uniform in each segment. Using this assumption, we were able to track a few features in the segments and generate an accurate and dense motion map by using the KLT feature tracker.

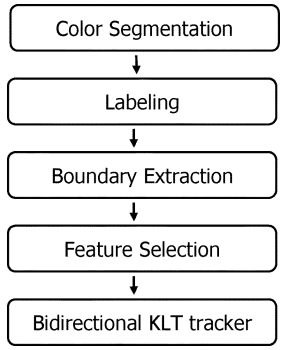

Fig. 5. Block diagram of motion estimation.

The motion estimation process consisted of five stages: color segmentation, labeling, boundary extraction, feature selection and feature tracking. The color segmentation stage separated the images with similar color areas. The labeling stage numbered the separated areas so that they could be distinguished from other areas. After color segmentation and labeling, the feature points were extracted from the boundary of the color segments and bidirectional motion estimation was performed. Fig. 5 shows a block diagram of the motion estimation process.

The mean shift algorithm (MSA) was utilized for color segmentation [19]. In general, the MSA estimates the density gradient of feature spaces and does not require multiple parameters. These are important characteristics for robust color segmentation. The MSA was used to calculate the mean of the high density areas in feature spaces. Conventional segmentation algorithms require the size and shape parameters of kernels and the number of neighboring pixels, but the MSA utilizes a sphere-shaped-window kernel to minimize these parameters. After color segmentation, labeling was performed which numbered connected pixels with the same value. In order to select the feature points after the labeling process, the size and contour information of each segment were calculated. The size of the segmented areas was calculated by the number of pixels with the same label. The segmented areas were organized by size and contour information in order to select many features from large segments.

Motion estimation was performed with the KLT feature tracker, which was composed of the feature tracking and extraction stages, which selected features by detecting large luminance gradients and matched them by comparing their similarity values with consecutive frames [20]. In the proposed method, feature points were selected from the contours of the color segmented area and features were tracked by the KLT feature tracker. Motion was computed by calculating the discrepancy of corresponding feature points in successive frames. Bidirectional tracking was then performed to increase the accuracy of feature tracking. Bidirectional tracking detected false matching that occurred with fast moving objects. When feature tracking failed or when features were not extracted from the

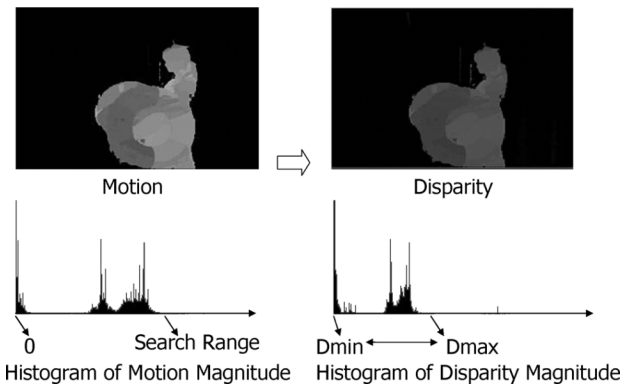

Fig. 6. Conversion of motion map to disparity map.

segments, interpolation was performed with the neighboring segment's color and distance information.

C. Motion-to-Disparity Conversion

The estimated motion vectors were converted into disparity vectors. Fig. 6 shows an example of scaling the motion vectors to the disparity vectors in a histogram. In Fig. 6, histogram of motion map and disparity map are shown. Data range of histogram of motion map is determined negative motion search range to positive motion search range. Because we use the magnitude of motion value to convert motion map into disparity map, range is 0 to search range of motion estimation. We have to adjust these motion values to disparity values whose range is determined by stereoscopic display characteristics.

The maximum and minimum values of the disparity vectors were already known by the display characteristic parameters that were determined in the stereoscopic display characterization stage. It is important to classify the scale factor of motion-to-disparity conversion without reverse depth or fatigue, because converted disparity vectors represent pseudo depth perceptions.

We used three cues to determine the scale factor of motion-to-disparity conversion: magnitude of motion, camera movements and scene complexity. The maximum disparity value was assigned when each cue indicated the maximum value. In motion-to-disparity conversion method, we utilize motion information to assign depth feeling in moving objects which are supposed to be attentional region in scene. In this case, it is difficult to deal with problem of far objects with fast motion and near objects with slow motion. This is a weak point of motion to disparity conversion method. That is why we use motion analysis to avoid these situations. We control the scale factor of motion to disparity conversion, in the several cases such as slow motion, camera motion and complicated motion scene. However, if we encounter with a scene of far objects with fast motion and near objects with slow motion, we can expect the best case that each motion is quantized to same value of maximum disparity values, and reversed depth map in the worst case.

Eq. (3) shows the maximum disparity value for motion-to-disparity conversion using the proposed three cues. The maximum disparity value was calculated by multiplying the three cues with the maximum disparity value of the stereoscopic display device.

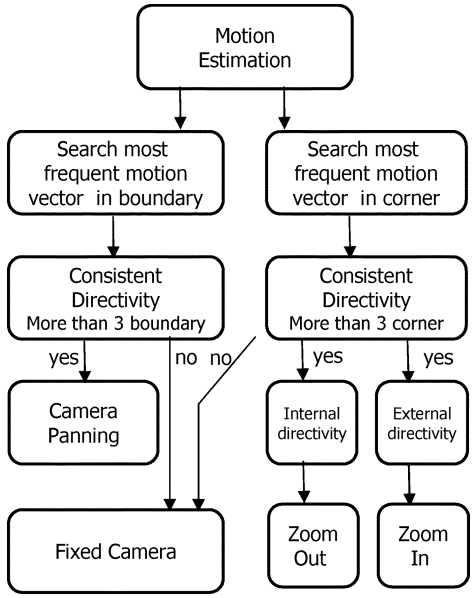

Fig. 7. Block diagram of camera movement recognition.

$$D_{max} = D_{max\ for\ display} \times Cue\,1 \times (1 - Cue\,2) \times (1 - Cue\,3) \tag{3}$$

where D_{max} represents the maximum disparity for motion-to-disparity conversion, and $D_{max\ for\ display}$ represents the maximum disparity value allowed for the characteristics of the display device.

1) Magnitude of Motion: Motion is an important factor that can be used to divide a given image into static background and dynamic foreground regions. If a moving object contains a large portion in the image and shows a different motion tendency than the surroundings, a maximum disparity value can be assigned, assuming that users are more interested in moving objects. Eq. (4) shows the magnitude of motion.

$$Cue\,1 = \alpha_1 \times \frac{M_{\max}}{Search\ range_{motion}} \tag{4}$$

where $M_{\max}$ represents the mean of the upper ten percent of the estimated motion vectors and α_1 represents the weighting factor of Cue 1.

2) Camera Movement: Stereoscopic conversion can generate disorderly results when a camera is moving because it is difficult to distinguish background and foreground regions with motion information captured with a moving camera. Users cannot experience 3D perceptions in foreground objects when using the same algorithm that is used with a fixed camera, because the motion of foreground regions is smaller than that of background regions.

Several algorithms for camera movement recognition have been proposed. The optical flow method directly analyzes patterns of optical flow by using angular distribution and the power of optical flow vectors [21]. The MPEG compressed video method directly manipulates encoded sequences in order to recognize camera motion [22].

We used a cue that recognizes camera movements in terms of fixed, panning and zooming cameras. In order to recognize panning and zooming, the motion value of the image's boundary was calculated and the most frequent value was determined. Fig. 7 shows a block diagram of camera movement recognition. This kind of recognition showed smaller computational complexity than conventional algorithms because it used only the information from the boundary regions.

Fig. 8 shows the recognition method for panning and zooming. Panning directions were classified by checking the motion tendency for three boundaries except the bottom part of the boundary in the motion map. The bottom part of the boundary is not suitable because the probability of errors caused by foreground objects is higher than that of the other boundaries. Zoom-in and zoom-out functions were classified by checking the motion tendencies in four corners of the motion map. When there was camera movement, scaling the motion vector to the disparity vector was not suitable.

When panning occurred, motion in the background and foreground regions was different. Smaller disparity values were assigned for panning because a reverse depth effect map occurred when the motion of the background region was larger than that of the foreground region. For zoom-in and zoom-out functions, the minimum disparity value was assigned in order to reduce eye fatigue. Eq. (5) shows the second cue which controlled the scale factor for the panning and zooming functions. This cue showed a lower value when the camera was moving.

$$Cue\ 2 = \alpha_2 \times \left(\frac{Block_{panning}}{Block_{boundary}} + \frac{Block_{zoom}}{Block_{corner}} \right) \quad (5)$$

where α_2 represent weighting factors for panning and zooming in Cue 2 and $Block_{boundary}$ and $Block_{corner}$ represent the arrowed areas of panning and zooming in Fig. 8, respectively.

3) Scene Complexity: The cue for scene complexity analyzed images and computed the complexity of the scenes. This cue assumed that it was hard to assign large amounts of disparity to images with complex motion patterns. For real-time implementation, the images were divided into macro blocks. The number of blocks with large differences was counted, as shown in the following equation.

$$Cue\ 3 = \alpha_3 \times \frac{Block_{complex}}{Total\ \#\ of\ Blocks} \quad (6)$$

where $Block_{complex}$ represents the number of blocks where the difference between the current block and the previous block was larger than the threshold.

4) Combining the Three Cues: The scaling factors computed by the multiple cues were combined to adjust the D_{max} value obtained by (3). After scaling the motion vector, a histogram of scaled motion was analyzed and equalization was performed. This is because 3D perceptions were maximized when disparity distribution was regularized.

Besides the proposed scaling factors, additional cues can be used according to various conditions. Also, real-time or non-real time cues can be selected according to stereoscopic conversion application.

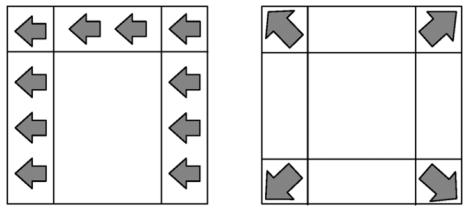

Fig. 8. Camera movement recognition using motion information.

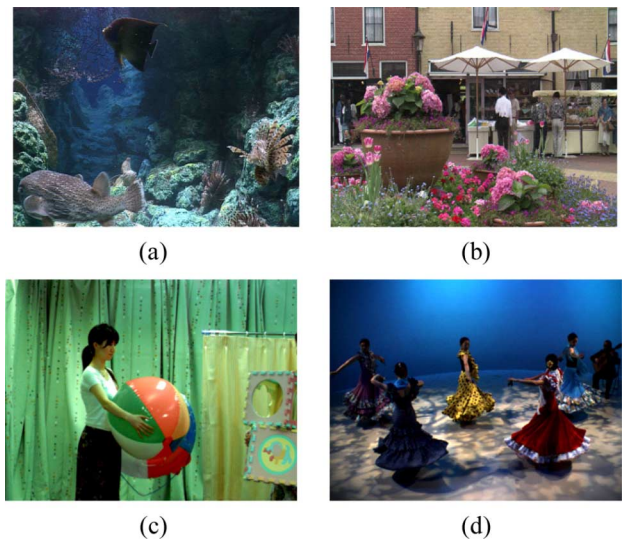

Fig. 9. Test image sequence sets. (a) Aquarium; (b) Flower Pot; (c) Akko&Kayo; (d) Flamenco.

D. Stereo Generation Obtained From Depth Maps

In the previous section, the motion vector was converted into a disparity vector using three cues. To generate a stereoscopic image pair based on disparity vectors, we used the algorithm proposed in [23]. This algorithm provided a solution for the occlusion problem in depth image based rendering using the depth smoothing method. They generated stereoscopic images with original views and corresponding depth images. However, we generated both left and right images from the reference image and the depth image which enabled stable and seamless results with the same disparity value. This approach can be extended to multi-view video generation when appropriate disparity values for multi-view displays are available.

III. Experimental Results

In order to evaluate the proposed algorithm, several sequences were used. We used two 1920 × 1080 stereoscopic image sequences called 'Aquarium' and 'Flower Pot', and two 640 × 480 multi-view image sequences 'Akko & Kayo' and 'Flamenco', as shown in Fig. 9. 'Aquarium', 'Akko & Kayo', and 'Flamenco' were captured with a fixed camera, and 'Flower Pot' was composed of scenes captured with a fixed camera as well as a panning camera.

We used a Pentium PC with a 17-inch polarized stereoscopic display device. This glass display offered a resolution of 1280

Fig. 10. Test stereoscopic image in side-by-side format for disparity and depth fusion.

TABLE I
EXAMINATION OF DEPTH FUSION ACCORDING TO VARYING DISPARITY VALUES

Test Sets	Disparity (cm)	Depth Fusion
1	2.64	×
2	2.11	×
3	1.58	×
4	1.06	o
5	0.53	o
6	0.00	o
7	-0.53	o
8	-1.06	×
9	-1.58	×
10	-2.11	×
11	-2.64	×

$\times$ 1024 pixels in 2D mode and 1280 $\times$ 512 pixels in stereoscopic mode [24]. The display characteristics were calculated through a centered striped fish in a synthesized stereoscopic image sequence, as shown in Fig. 10. The stereoscopic image was synthesized according to varying disparity values, and then we found the maximum and minimum disparity values and the field of view for the display device. These display characteristics were determined by repeated experiments.

Table I shows an example of whether depth fusion succeeded or not, according to each disparity value when the viewing angle was 130°. We marked "O" when the participants of experiments are possible to make stereopsis fusion, in other words they do not see the split right and left images. The unit of disparity is the pixel, which was converted into centimeter units by using (2) as follows.

$$Disparity(cm) = 0.032(cm/pix) \times disparity(pix) \quad (7)$$

We performed the experiment in Table I for every ten degrees in the field of view of the stereoscopic display. Table II shows how the field of view was set between 50° and 90° and 90° and 130° in symmetry. In Table II, we found that the maximum and minimum disparity values for display were 1.06 centimeters to -0.53 centimeters, according to Table I.

Figs. 11 to 14 shows the results of motion estimation for the four test sequences. In this figure, we found that the shape of the objects were well represented enough to assign depth perceptions to the moving objects. Note that in 'Flower Pot', there was

TABLE II
MINIMUM AND MAXIMUM DISPARITY VALUES ACCORDING TO VARYING ANGLES

Angle(°)	Maximum Disparity (cm)	Minimum Disparity (cm)
160	×	×
150	×	×
140	×	×
130	1.06	-0.53
120	1.06	-1.06
110	2.64	-2.64
100	2.64	-2.64
90	2.64	-2.64

reverse depth of the background and foreground regions. This verifies that the camera movement recognition process is essential. In general, errors may occur when the original images are roughly segmented or when there are variations in the amount of illumination.

Fig. 15 shows the results of the three cues for the four video sequences. Figs. 15(a)–(c) are the results of the three cues, respectively. Large disparity values were assigned when there were larger motion sizes, fewer camera movements, and simpler scene complexity values. Fig. 15(a) shows that 'aquarium' has the biggest magnitude of motion and Fig. 15(b) shows that the camera movements are detected in 'Flower Pot'. Weighting factors of three cues α_1, α_2, α_3 are empirically set to 1, 0.7, 0.7. We choose the largest weighting factors for magnitude of motion based on the assumption that motion-to-disparity conversion method assigns large disparity for fast moving object. We choose smaller values for other cues; camera movement and scene complexity. As previously mentioned, besides these proposed scaling factors, additional cues can be used and corresponding weighting factors should be carefully chosen.

The performance of generated stereoscopic video was evaluated in a subjective manner by comparing a conventional stereoscopic video with a stereoscopic video that was generated from one view. Subjective evaluation was performed by surveying participants after watching the stereoscopic video. Participants were composed of 30 people with normal visual acuity and stereo-acuity. They watched a randomly-ordered stereoscopic video twice and assigned the video a grade from 1 to 10, according to three evaluation items: sense of presence, protrusion, and fatigue.

We then provided three types of video sequences: videos that were acquired by a stereoscopic camera, videos that were generated by the proposed algorithm and videos that were generated by the conventional stereoscopic conversion algorithm using Dynamic Depth Cueing [25].

Fig. 16 shows the results of interlaced video of generated stereoscopic video which represent the disparity of stereoscopic video. Fig. 17 shows the results of subjective evaluation. Figs. 17(a)–(c) represent the mean of the values acquired by experiments for the four evaluation sequences. Fig. 17(d) represents the weighted sum of the three evaluation terms. A higher weighting factor was assigned to evaluation terms with lower variance values, which were considered to be reliable terms.

$$Y = w_1 Y_1 + w_2 Y_2 + w_3 Y_3 \quad (8)$$

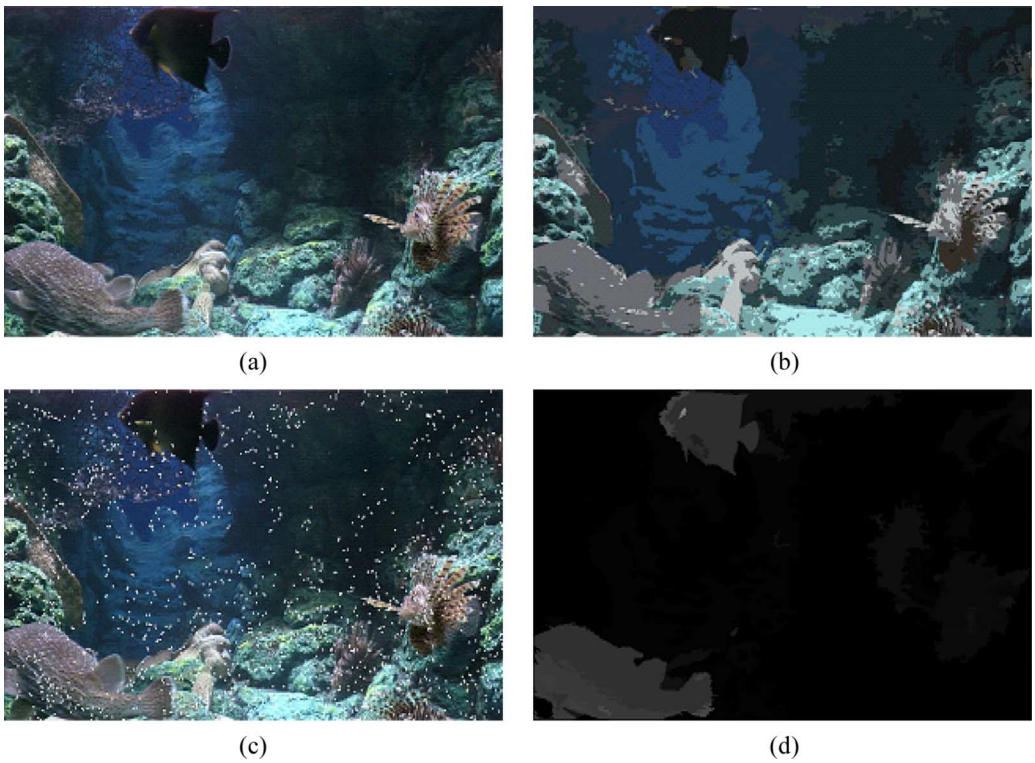

Fig. 11. Results of motion estimation (Aquarium). (a) Original image; (b) color segmentation; (c) feature selection; (d) motion estimation.

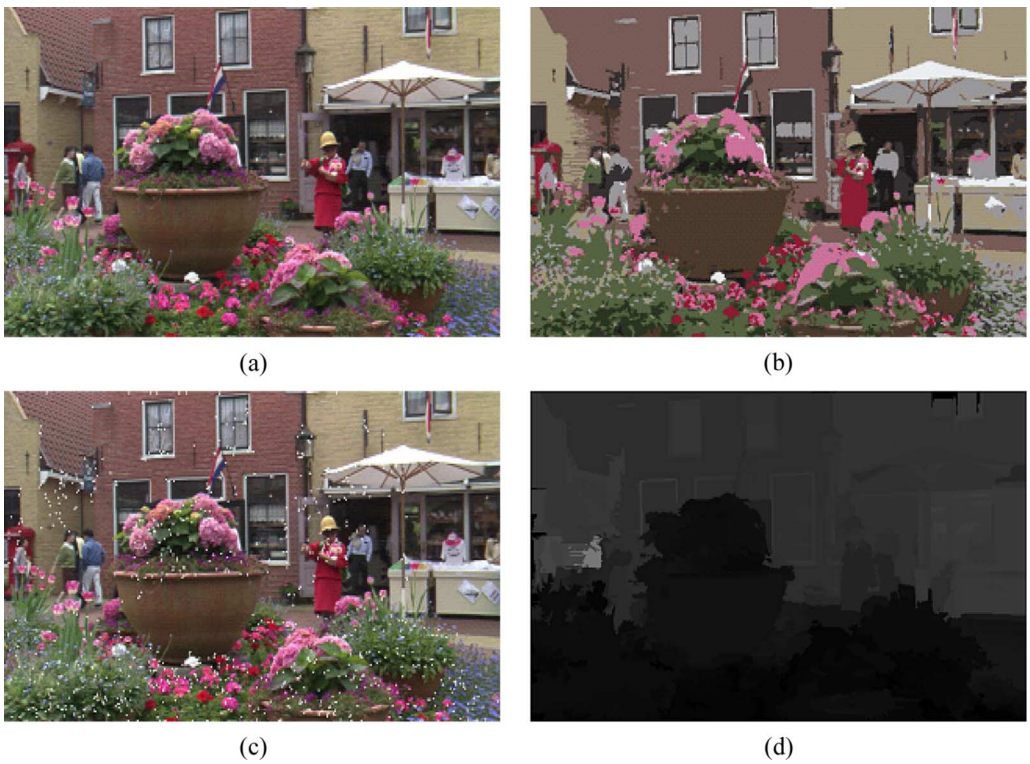

Fig. 12. Results of motion estimation (Flower Pot). (a) Original image (b) color segmentation; (c) feature selection; (d) motion estimation.

where Y_i represents the mean of evaluation terms from the participants, σ_i represents the variance of the evaluation terms from the participants. The weighted value was calculated by

$$w_i = \frac{1/\sigma_i}{1/\sigma_1 + 1/\sigma_2 + 1/\sigma_3} \tag{9}$$

The stereoscopic video captured with the stereoscopic camera obtained the highest score, and the proposed algorithm was superior to the conventional algorithm in terms of presence and protrusion. However, the stereoscopic camera did not obtain the highest score in terms of fatigue because

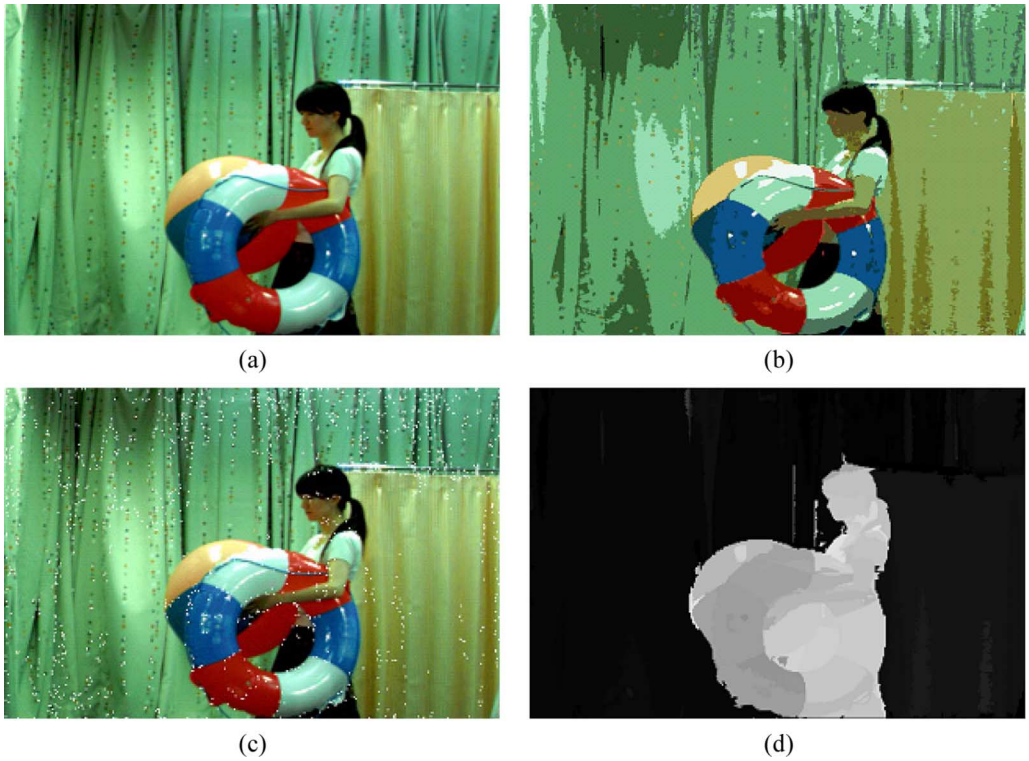

Fig. 13. Results of motion estimation (Akko&Kayo). (a) Original image; (b) color segmentation; (c) feature selection; (d) motion estimation.

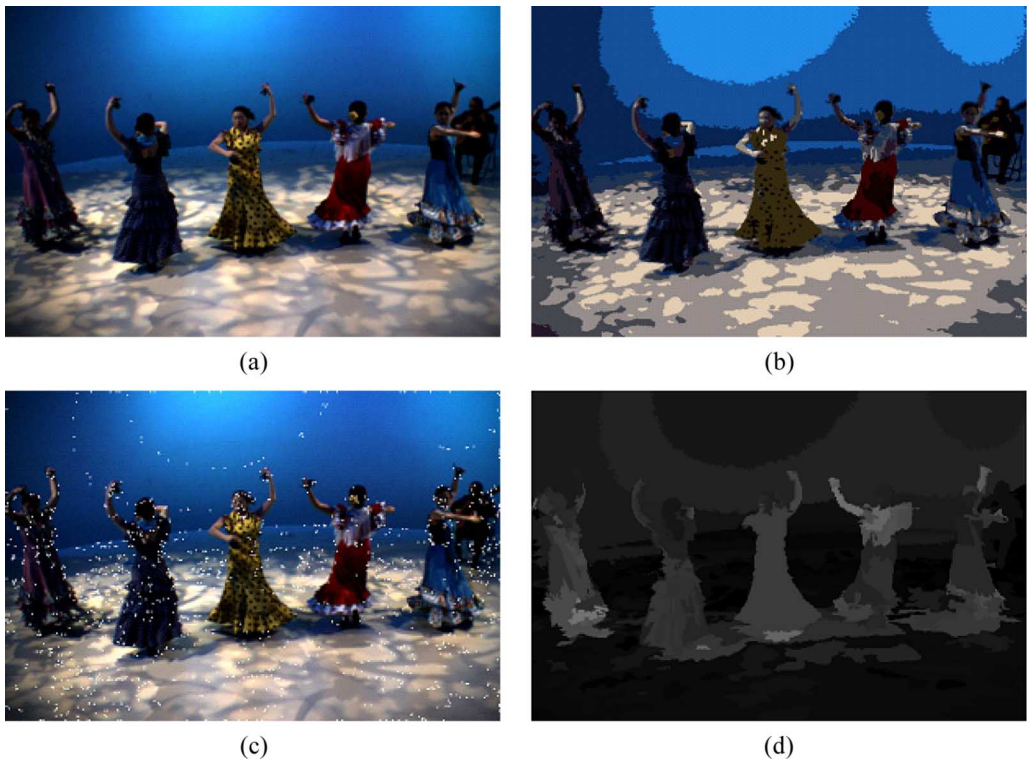

Fig. 14. Results of motion estimation (Flamenco). (a) Original image; (b) color segmentation; (c) feature selection; (d) motion estimation.

the camera arrangement was not similar to the human visual system. General expectation to have most comfortable results in stereoscopic camera was conflicted by several distortion factors which were not similar to human visual system. Various kinds of stereoscopic distortions make visual fatigue. Andrew Woods discussed the types of image distortion in stereoscopic video systems [26]. When filming stereoscopic video, we must consider the capturing parameters of stereoscopic video filming which includes camera baseline, focal length and convergent angle should be considered to adapt human visual system.

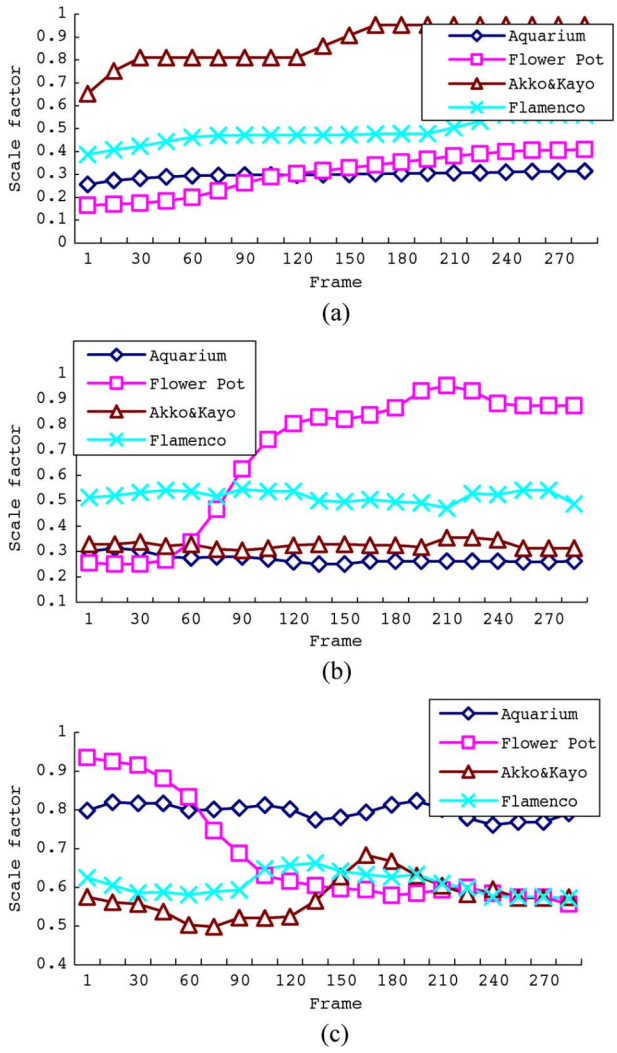

Fig. 15. Results of the three cues. (a) Magnitude of motion; (b) camera movements; (c) scene complexity.

Fig. 16. Interlaced stereoscopic video. (a) Aquarium; (b) Flower Pot; (c) Akko&Kayo; (d) Flamenco.

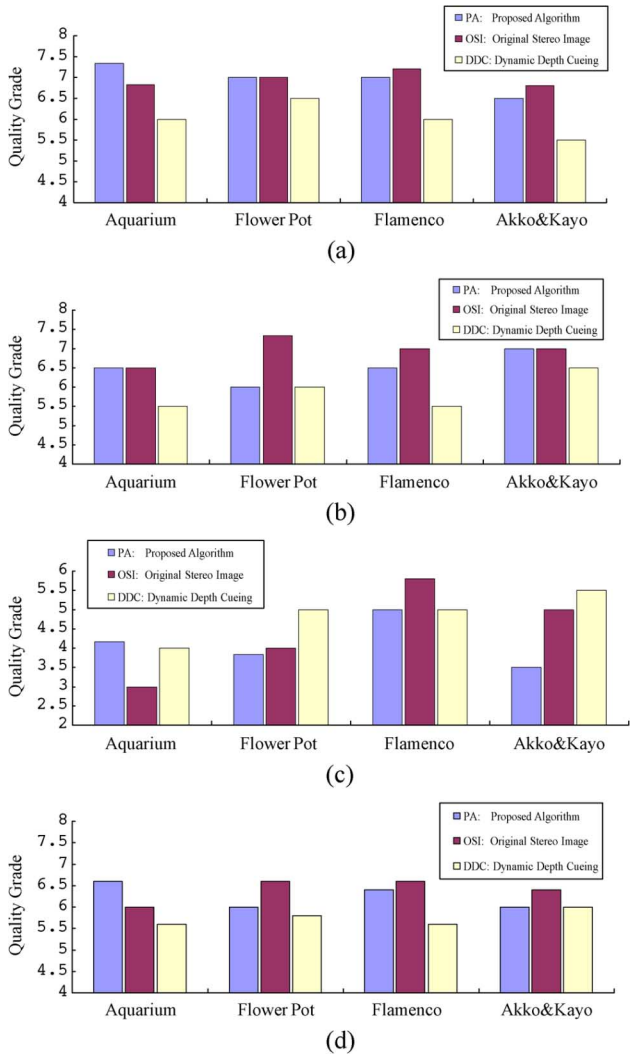

Fig. 17. Subjective evaluation. (a) Sense of presence; (b) protrusion; (c) fatigue; (d) evaluation with weighted sum.

IV. Conclusion

In this paper, we proposed a stereoscopic video generation method using motion-to-disparity conversion. In order to consider multi-user conditions and stereoscopic display devices, a stereoscopic display characterization process was performed. We obtained the field of view and the maximum and minimum disparity values for a stereoscopic display device in the stereoscopic display characterization process. Motion vectors were estimated by using color segmentation and the KLT feature tracker. After motion estimation, motion-to-disparity conversion was performed by scale factors computed by several proposed cues. Subjective evaluation showed that the generated stereoscopic videos were stable and comfortable. The proposed algorithm can be improved by additional conditions of the scale factor decision method. Also, the fast motion estimation method can be used to make the use of the proposed system in real-time applications possible.

In future work, we will work on scene change detection. For example, it is hard to detect blending between shots, because current algorithms simply compare the mean of image sequences for scene change detection. Moreover, present research

is targeted to stereoscopic content generation, but this research can be further extended to multi-view content generation using depth map scaling.

REFERENCES

[1] T. Okino and H. Murata, "New television with 2D/3D image conversion technologies," *SPIE*, vol. 2653, pp. 96–103, 1996.

[2] H. Murata and Y. Mori, "A real-time 2D to 3D image conversion technique using computed image depth," in *SID 98 DIGEST*, 1998, pp. 919–922.

[3] W. J. Tam and L. Zhang, "3D-TV content generation: 2D-to-3D conversion," in *2006 IEEE International Conference on Multimedia and Expo*, 2006, pp. 1869–1872.

[4] M. B. Kim and M. S. Song, "Stereoscopic conversion of monoscopic video by the transformation of vertical to horizontal disparity," *SPIE*, vol. 3295, pp. 65–75, 1998.

[5] M. B. Kim and S. H. Park, "Object-based stereoscopic conversion of MPEG4 encoded data," in *PCM*, 2004, pp. 491–498.

[6] Y. Matsumoto and H. Terasaki, "Conversion system of monocular image sequence to stereo using motion parallax," *SPIE*, vol. 3012, pp. 108–115, 1997.

[7] E. Rotem, K. Wolowelsky, and D. Pelz, "Automatic video to stereoscopic video conversion," in *SPIE Conference on Stereoscopic Displays and Applications XVI*, 2005, vol. 5664, pp. 198–206.

[8] L. Zhang, B. Lawrence, D. Wang, and A. Vincent, "Comparison study on feature matching and block matching for automatic 2D to 3D video conversion," in *The Second IEE European Conference on Visual Media Production (CVMP 2005)*, London, UK, 2005, pp. 122–129.

[9] A. Azarbayejani, "Recursive estimation of motion, structure, and focal length," *Pattern Analysis and Machine Intelligence, IEEE Trans.*, vol. 17, no. 6, pp. 562–575, 1995.

[10] T. Jebara, "3D structure from 2D motion," *Signal Processing Magazine, IEEE*, vol. 16, no. 3, pp. 66–84, 1999.

[11] S. Diplaris, "Generation of stereoscopic image sequences using structure and rigid motion estimation by extended Kalman filters," in *IEEE International Conference on Multimedia and Expo*, 2002, pp. 233–236.

[12] K. Moustakas, "Stereoscopic video generation based on efficient layered structure and motion estimation from monoscopic image sequence," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 15, no. 8, pp. 1065–1073, 2005.

[13] S. Battiato, "3D stereoscopic image pairs by depth-map generation," *3dpvt*, pp. 124–131, 2004.

[14] B. J. Garcia, "Approaches to stereoscopic video based on spatio-temporal interpolation," *SPIE*, vol. 2653, pp. 85–95, 1990.

[15] J. Ross and J. H. Hogben, "The Pulfrich effect and short-term memory in stereopsis," *Vision Research*, vol. 15, pp. 1289–1290, 1975.

[16] D. C. Burr and J. Ross, "How does binocular delay give information about depth?," *Vision Research*, vol. 19, pp. 523–532, 1979.

[17] H. Tao and H. Sawhney, "A global matching framework for stereo computation," in *Proc. ICCV*, 2001, pp. 532–539.

[18] L. Hong and G. Chen, "Segment-based stereo matching using graph cuts," in *Proc. IEEE CVPR*, 2004, pp. 74–81.

[19] D. Comaniciu, "Robust analysis of feature spaces: Color image segmentation," in *CVPR*, 1997, pp. 750–755.

[20] C. Tomasi and T. Kanade, Detection and Tracking of Point Features Technical Report, CMU-CS-91-132, 1991.

[21] G. Sudhir and J. C. M. Lee, "Video annotation by motion interpretation using optical flow stream," *J. Vis. Commun. Image Represent*, vol. 4, pp. 354–368, 1996.

[22] A. Akutsu, Y. Tonomura, and H. Hashimoto, "Video indexing using motion vectors," in *Proc. SPIE VCIP*, 1992, vol. 1818, pp. 1522–1530.

[23] L. Zhang, "Stereoscopic image generation based on depth images for 3D TV," *IEEE Trans. Broadcasting*, vol. 51, no. 2, pp. 191–199, 2005.

[24] [Online]. Available: http://www.dimen.co.kr/sub01/02.html

[25] [Online]. Available: http://www.ddd.com/technology/tech_tridefreal-time.html

[26] A. Woods, "Image Distortions in Stereoscopic Video System," *SPIE*, 1993.

Donghyun Kim (S'07) received the B.S., M.S., in electrical and electronic engineering from Yonsei University, Seoul, Korea, in 2004 and 2007, respectively. He is currently pursuing the Ph.D. degree at Yonsei University.

His research interests include 2D to 3D video conversion, 3D computer vision and 3D video quality assessment.

Dongbo Min (S'07) received the B.S., M.S., in electrical and electronic engineering from Yonsei University, Seoul, Korea, in 2003 and 2005, respectively. He is currently pursuing the Ph.D. degree at Yonsei University.

His research interests include stereo vision, 3D modeling, view synthesis, and hybrid sensor system.

Kwanghoon Sohn (M'92) received the BE degree in electronics engineering from Yonsei University, Seoul, Korea, in 1983, the MSEE degree in electrical engineering from University of Minnesota in 1985, and the PhD degree in electrical and computer engineering from North Carolina State University in 1992. He was employed as a senior member of the research staff in the Satellite Communication Division at Electronics and Telecommunications Research Institute, Daeduk Science Town, Korea, from 1992 to 1993. Also, he was employed as a postdoctoral fellow at the MRI Center in the Medical School of Georgetown University. He was a visiting professor of Nanyang Technological University from 2002 to 2003. He is currently a professor in the School of Electrical and Electronic Engineering at Yonsei University. His research interests include three-dimensional image processing, computer vision, image communication, and neural networks. Dr. Sohn is a member of IEEE, the Korean Institute of Communications Science, and the Korean Institute of Telematics and Electronics.