STEREOSCOPIC VIDEO GENERATION METHOD USING MOTION ANALYSIS

Donghyun Kim, Dongbo Min, and Kwanghoon Sohn

Dept. of Electrical and Electronic Eng., Yonsei University, Seoul, Korea
khsohn@yonsei.ac.kr

ABSTRACT

Stereoscopic video generation method can produce stereoscopic contents from conventional video filmed from monoscopic camera. We propose stereoscopic video generation method using motion-to-disparity conversion considering multi-user condition and characteristics of display device. Field of view, maximum and minimum values of disparity are calculated through initialization process in order to apply various types of 3D display. After motion estimation, we propose three cues to decide the scale factor of motion-to-disparity conversion which are magnitude of motion, camera movement and scene complexity. Subjective evaluation is performed by comparing videos captured from stereoscopic camera and generated from one view of stereoscopic video.

Index Terms— Multimedia systems, Stereoscopic video generation

1. INTRODUCTION

One of serious problems in 3D video technology is lack of 3D contents. There are many ways to generate 3D contents: Capture with stereoscopic camera, 3D graphics and manual conversion of 2D contents. However, these methods are expensive, time consuming and laborious tasks. In this paper, we propose an automatic stereoscopic conversion algorithm based on computer vision technique. Automatic stereoscopic conversion (2D/3D conversion) can provide a various 3D contents because it can make 3D contents from conventional 2D videos. 2D videos from Broadcasting, CATV and DVD can be converted into stereoscopic videos by automatic stereoscopic video conversion. Conversion technique allows people to enjoy 3D images in 3D display devices. Several stereoscopic convergence algorithms have been proposed. Modified time difference (MTD) method detects movements of the object and decides delay direction and time by the characteristics of movements. Then, stereoscopic images are selected according to time difference in 2D image sequences [1]. Computed image depth (CID) method uses relative position between the multiple objects in still image. Image depth is computed with contrast, sharpness and chrominance of the input images [2]. Motion-to-disparity conversion method generates stereoscopic images by converting motion from the movement of object to the horizontal disparity. In order to eliminate the effect of vertical disparity, norm of motion vector is converted into horizontal disparity [3][4]. There are several structure estimation methods calculating the depth of scene. Structures are estimated assuming that camera motion is restricted to translation [5]. Using the extended Kalman filter, camera motion is estimated in rotation and translation and the structure of scene is estimated [6][7]. As an alternative method, the method which uses detection of vanishing line was proposed [8]. There is a method using sampling density of spatial temporal interpolation in human visual characteristics [9]. [10] uses Pulfrichi effect which is the time delay effect by difference of the amount of light in both eyes. In this paper, we propose an automatic framework of stereoscopic video generation system which uses motion-to-disparity conversion method. Multi-user condition and characteristics of stereoscopic display are considered for general purpose of stereoscopic contents generation. We propose three cues to decide the scale factor of motion-to-disparity conversion which are magnitude of motion, camera movement and scene complexity.

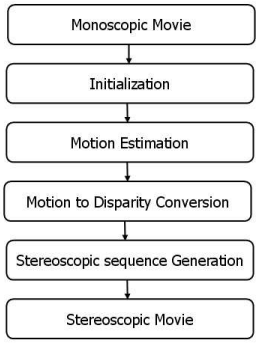


Fig. 1. Block diagram of overall system

2. PROPOSED ALGORITHM

The proposed algorithm is general methodology of stereoscopic video generation algorithm based on motion-to-disparity conversion method. Block diagram of the proposed stereoscopic conversion algorithm is shown in Fig. 1. At the initial-

"This research was supported by the MIC, Korea, under the ITRC support program supervised by the IITA" (IITA-2006-(C1090-0603-0011))

ization stage, field of view, maximum and minimum disparity are determined to consider multi-user condition and the characteristics of display device. Motion is estimated with bidirectional KLT feature tracker based on color segmentation. After motion estimation, scale factor of motion-to-disparity conversion is determined with multiple cues. Finally, stereo views are generated by using computed depth map and the original video.

2.1. Initialization

The conversion system must consider about circumstances where the contents are displayed. Circumstances include not only the place but the number of audience, lighting, sound and so on. If the place is a theater for hundreds of audience, glass type stereoscopic display device and non-real time conversion method should be chosen. Once display device and conversion method are determined, number of audience and field of view are decided to control the magnitude of disparity for stereoscopic conversion. At the initialization stage, maximum and minimum value of disparity is determined by adjusting various value of disparity and verifying success of stereoscopic fusion for each viewing angle and distance.

2.2. Motion estimation

Stereoscopic video generation using motion-to-disparity conversion is the method that assigns depth feeling to moving objects which audiences are supposed to be interested in. Therefore, acquiring dense and accurate motion map is one of the most important parts. In this paper, motion map is calculated by using color segmentation and KLT (Kanade-Lucas-Tomasi) feature tracker. Color segment based method is used for robust estimation in textureless region and boundary of objects, and it is also used in stereoscopic matching algorithm [11]. We assume motion is uniform in each segments. With this assumption, tracking a few features in segments by KLT feature tracker enables to generate accurate and dense motion map. Mean shift algorithm (MSA) is utilized for color segmentation [12]. Mean shift algorithm estimates density gradient of feature space and do not require multiple parameters which is important characteristics for robust color segmentation. After color segmentation, labeling and feature selection is performed. Motion is estimated with KLT feature tracker[13]. In the proposed method, feature points are selected from boundary of color segmented area and features are tracked by KLT feature tracker. Bidirectional tracking is performed to increase accuracy of feature tracking. When feature tracking is failed or features are not extracted from segment, interpolation is performed with neighbor segment's color and distance information.

2.3. Motion-to-disparity conversion

It is important to classify scale factor of motion-to-disparity conversion without occurring reverse depth or fatigue, because converted disparity vectors are pseudo depth that assign the moving objects 3D feeling in image sequences. We propose three cues to decide the scale factor of motion-to-disparity conversion. They are magnitude of motion, camera movement and scene complexity. Maximum disparity is assigned when each cue indicates maximum value. Eq. (1) shows the maximum disparity for motion-to-disparity conversion using the proposed three cues. Maximum disparity is calculated by multiplying three cues to maximum disparity value of stereoscopic display device.

$$D_{max} = D_{max\ for\ disparity} \times Cue1 \times Cue2 \times Cue3 \quad (1)$$

,where D_{max} is maximum disparity for motion-to-disparity conversion, D_{max} for display is maximum disparity value allowed for characteristics of display.

2.3.1. Magnitude of motion

Motion is an important factor that can divide an image into static background and dynamic foreground. If the moving object has a large portion in the image and different motion tendency with surroundings, we can assign maximum disparity, assuming that users are interested in the moving object. Eq. (2) shows the cue of magnitude of motion.

$$Cue1 = \alpha_1 \times \frac{M_{max}}{Search\ range_{motion}} \quad (2)$$

,where M_{max} is the mean of upper 10 percent in estimated motion vectors and α_1 is weighting factor of Cue 1.

2.3.2. Camera movement

Stereoscopic conversion can generate disordered results when camera is moving because it is difficult to distinguish background and foreground with motion information captured from moving camera. User cannot perceive 3D feeling in foreground object when using same algorithm which is used for fixed camera case because motion of foreground is smaller than motion of background. We propose a cue that recognizes the movement of camera to fixed, panning and zooming camera. Fig. 2 shows the recognition method for camera panning and zooming. Directions of camera panning are classified by checking motion tendency in 3 boundaries except bottom part of boundary in motion map. Bottom part of boundary is not suitable because probability of error caused by foreground object is higher than other boundaries. Zoom-in and zoom-out are classified by checking motion tendency in 4 corner of motion map. When camera movement is panning, smaller disparity is assigned because reverse depth effect occur when the motion of background is larger than that of foreground. In zoom-in and zoom-out case, minimum disparity is assigned to reduce eye fatigue. Eq. (3) shows the second cue which controls scale factor for panning and zooming camera.

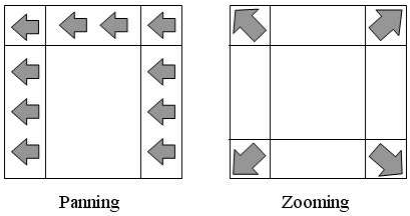


Fig. 2. Camera movement recognition

$$Cue2 = 1 - \left(\alpha_2 \times \frac{Block_{panning}}{Block_{boundary}} + \alpha_3 \times \frac{Block_{zoom}}{Block_{corner}} \right) \quad (3)$$

,where α_2, α_3 are weighting factor for panning and zooming in Cue 2. $Block_{boundary}$ and $Block_{corner}$ are arrowed areas of panning and zooming in Fig. 2, respectively.

2.3.3. Scene complexity

The cue for scene complexity analyze image and compute the complexity of scene. This cue assumes that image with complex motion pattern is hard to assign large amount of disparity. For real-time implementation, image is divided into macro blocks and the numbers of blocks that have larger difference than threshold with neighbor blocks are counted as following equation.

$$Cue3 = 1 - \alpha_4 \times \frac{Block_{complex}}{Total\ \#\ of\ Blocks} \quad (4)$$

,where $Block_{complex}$ is the number of block where the difference between current block and previous block is larger than threshold.

2.4. Stereo generation from depth map

To generate stereoscopic image pair based on disparity vectors we used the algorithm proposed by [14]. This algorithm provided a solution for occlusion problem in depth image based rendering with depth smoothing method. They generate stereoscopic image with original view and corresponding depth image. However, we generate both left and right images from reference image and depth image which enable stable and seamless results in same disparity. This approach can be extended to multi-view video generation when appropriate disparity for multi-view display can be acquired.

3. EXPERIMENTAL RESULTS

In order to evaluate the proposed algorithm, several sequences were used. We used two stereoscopic sequences "Aquarium" and "Flower Pot" and two multi-view sequences "Akko&kayo" and "Flamenco." Our test platform is 17 inch polarized stereoscopic display device which offers resolution of 1280×512 in stereoscopic mode. We empirically set field of view to 50 to 130, and maximum and minimum values of disparity for display to 1.06 cm to -0.53 cm by stereoscopic fusion experiments. Fig. 3 shows results of stereoscopic conversion for four test sequences. In this figure, we could find shape of objects are well represented enough to assign depth feeling to the moving objects. Note that in the "Flower Pot" captured by panning camera, the result shows that there is reverse depth of background and foreground. It verifies camera movement recognition is essential process. Errors may occur when the original images are roughly segmented or illumination variation is occurred. Fig. 4 shows the results of three cues which are magnitude of motion, camera movement, and scene complexity. Large disparity can be assigned when there is larger magnitude of motion, less camera movement, and simpler scene complexity. Performance of generated stereoscopic video was evaluated in subjective manner by comparing the conventional stereoscopic video with generated stereoscopic video from one view. We provided three types of video sequences: The videos acquired by stereoscopic camera, generated video by proposed algorithm and generated video by conventional stereoscopic conversion algorithm using Dynamic Depth Cueing [15]. Subjective evaluation was performed for 30 persons. They watch random ordered stereoscopic video twice and give a grade from 1 to 10 point according to three evaluation terms which are sense of presence, protrusion, fatigue. Fig. 5 is weighted sum of three subjective evaluation terms. Higher weighting factor is assigned to evaluation term with lower variance which is considered as reliable term. Stereoscopic video captured from stereoscopic camera obtain highest score, and the proposed algorithm is superior to the conventional algorithm in sense of presence and protrusion terms. However, stereoscopic camera did not obtain highest score in fatigue term because camera arrangement is not similar to human visual system.

4. CONCLUSION

In this paper, we proposed the stereoscopic video generation method using motion-to-disparity conversion. In order to consider multi-user condition and stereoscopic display characteristic, initialization process is performed. We obtain field of view, maximum and minimum disparity for stereoscopic display device in initialization process. Motion vectors are estimated by using color segmentation and KLT feature tracker. After motion estimation, motion-to-disparity conversion is performed by scale factors computed by the proposed several cues, which are magnitude of motion, camera movement and scene complexity. Subjective evaluation shows the generated stereoscopic videos are stable and comfortable. The proposed algorithm can be improved by additional conditions of scale factor decision method. Present research is targeted to stereoscopic content generation, but the research can be extended to multi-view contents for 3D display by depth map scaling.

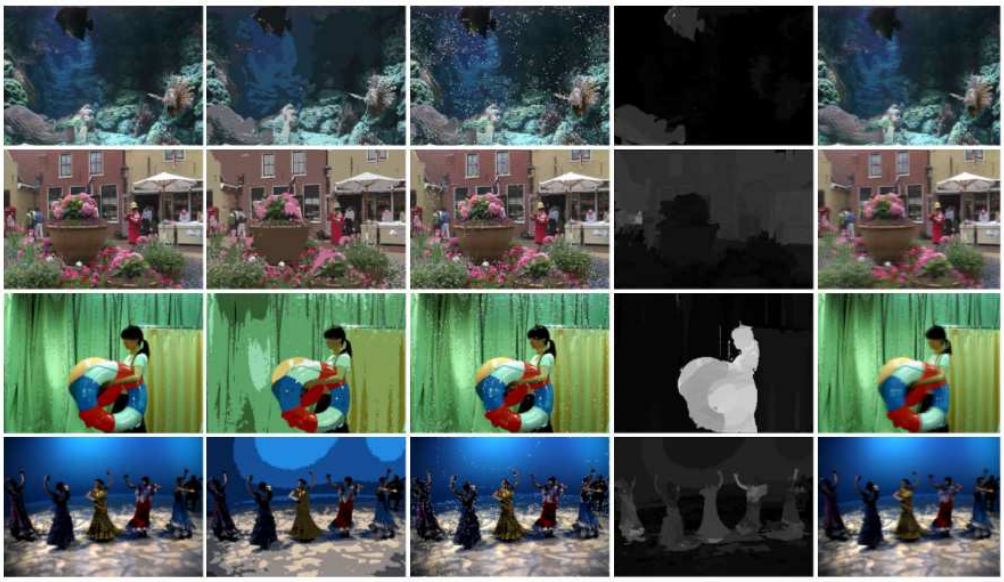
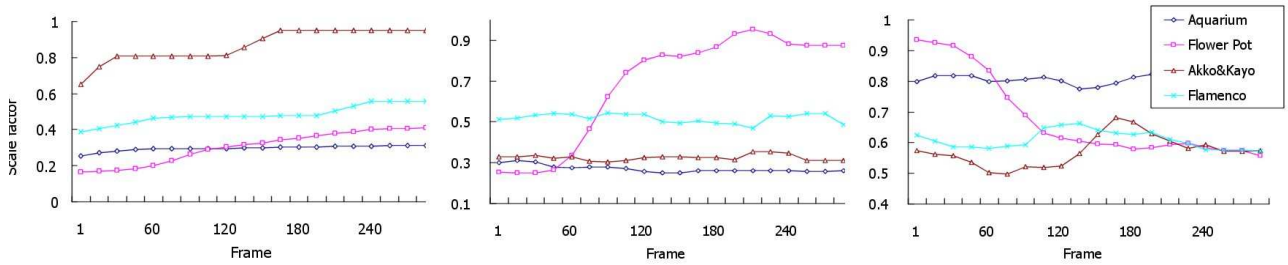


Fig. 3. Results of stereoscopic conversion



(a) Magnitude of motion (b) Camera movement (c) Scene complexity

Fig. 4. Results of three cues

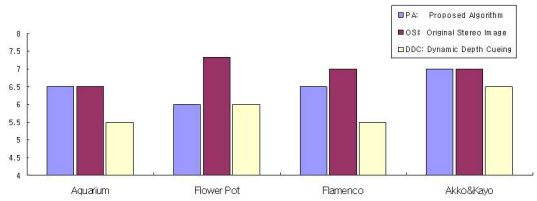


Fig. 5. Results of subjective evaluation

5. REFERENCES

[1] Toshiyuki Okino, Haruhiko Murata, "New television with 2D/3D image conversion technologies," *SPIE*, Vol. 2653, pp.96-103, 1996

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

[3] Man-Bae Kim, Mun-Sup Song, "Stereoscopic conversion of monoscopic video by the transformation of vertical to horizontal disparity," *SPIE* Vol. 3295, pp.65-75, 1998

[4] Manbae Kim, Sanghoon Park, "Object-based stereoscopic conversion of MPEG4 encoded data," *PCM*, pp.491-498, 2004

[5] Y. Matsumoto, H. Terasaki, "Conversion system of monocular image sequence to stereo using motion parallax," *SPIE* Vol. 3012, pp.108-115, 1997

[6] Sotiris Diplaris, "Generation of stereoscopic image sequences using structure and rigid motion estimation by extended kalman filters," *IEEE International Conference on Multimedia and Expo*, pp.233-236, 2002

[7] Konstantinos Moustakas, "Stereoscopic video generation based on efficient layered structure and motion estimation from monoscopic image sequence," *IEEE Transaction on Circuits and Systems for Video Technology*, Vol. 15, No. 8, pp.1065-1073, 2005

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

[9] Baxter J. Garcia "Approaches to stereoscopic video based on spatio-temporal interpolation," *SPIE* Vol. 2653, pp.85-95, 1990

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

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

[12] Dorin Comaniciu, "Robust analysis of feature spaces: color image segmentation," *CVPR*, pp.750-755, 1997

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

[14] Liang Zhang, "Stereoscopic image generation based on depth images for 3D TV," *IEEE Transaction on Broadcasting*, Vol. 51, No. 2, pp.191-199, 2005

[15] http://www.ddd.com/technology/tech_tridefrealtime.html