3D Video Processing for 3DTV

Kwanghoon Sohn
Dept. of Electrical and Electronic Engineering, Yonsei University,
134 Shinchon-dong, Seodaemun-gu,Seoul 120-749, Korea
TEL/FAX: +82-2-2123-2879,
e-mail: khsohn@yonsei.ac.kr.

Keywords : 3DTV, 3D content generation, 3D codec, 3D display

Abstract

This paper presents the overview of 3D video processing technologies for 3DTV such as 3D content generation, 3D video codec and video processing techniques for 3D displays. Some experimental results for 3D contents generation are shown in 3D mixed reality and 2D/3D conversion.

1. Introduction

One of the most desirable ways of realizing high quality information and telecommunication services has been called "The Sensation of Reality", which can be achieved by visual communication based on 3D images. These kinds of 3D imaging systems have shown many potential applications in the fields of education, entertainment, medical surgery, video conferencing, etc. Especially, 3DTV (Three-Dimensional Television) is believed to be the next generation of TV technology. Fig. 1 shows how TV has evolved in terms of the development of display technologies.

Interest in the use of 3D images has provoked numerous research groups to report on 3D image processing and display systems.

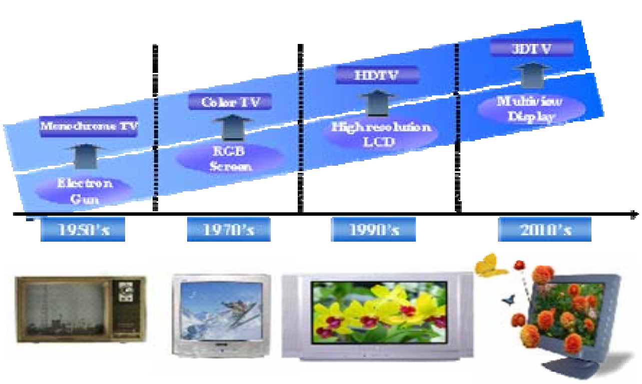


Fig.1. Historical evolution of TV technology.

3DTV projects have also been performed in many countries. In Europe, research on 3DTV was initiated by several projects such as the DISTIMA (Digital Stereoscopic Imaging and Applications) project. The goals of the ATTEST (Advanced Three-Dimensional Television System Technologies) project were to develop a flexible, 2D-compatible and commercially feasible 3DTV system [1]. Next, the objective of the NoE (Network of Excellence) was to produce a 3DTV system with high-end displays that could provide true 3D views [2]. Another noteworthy effort was NHK's 3D HDTV project in Japan [3]. In addition, 3D broadcasting with stereo displays was proposed for the FIFA 2002 World Cup in South Korea [4].

This paper is organized as follows. Section 2 presents the 3D video processing technologies for 3DTV such as 3D contents generation, 3D video codec and 3D displays. The experimental results of several 3D content generation methods are provided in Section 3. Finally, Section 4 concludes a summary of this paper.

2. 3D Video processing for 3DTV

2.1. 3D Contents Generation

To date, one of the most serious problems of 3D video technology has been a lack of 3D content. However, many ways of generating 3D content have recently been developed, such as using 3D CG (computer graphics), capturing images with multiple cameras and converting 2D content to 3D content.

In all 3D systems, the depth information of real scenes is an essential component. However, it is theoretically impossible to recover perfect depth information from images since this process represents an inverse problem. In order to recover 3D information from 2D image sequences, more than two cameras are usually necessary and all video cameras must be synchronized.

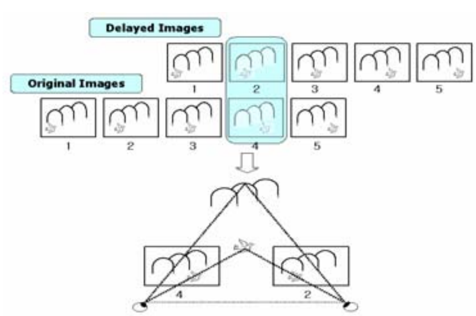
The 3D world is captured by two cameras, where a 3D point P (X,Y,Z) is projected into pixels $p_l(x_l, y_l)$ and $p_r(x_r, y_r)$ on each image plane. In the following equation, x_l and x_r denote the X-coordinates of the projected points on the left and right images, respectively. The level of disparity with respect to the left camera is defined as:

$$d = x_l - x_r = \frac{fX}{Z} - \frac{f(X-B)}{Z} \qquad (1)$$

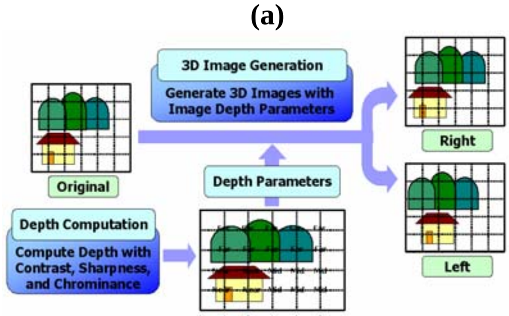
MR (Mixed Reality) systems refer to environments in which both virtual and real environments exist [5]. In MR systems, users can immerse themselves and interact in spaces composed of real objects as well as computer-generated objects.

In most conventional MR systems, virtual objects are simply overlaid onto images of real objects as if the virtual objects were placed in front of the real ones. When real objects are placed in front of virtual ones, the virtual object images have to be pruned before display. Moreover, when virtual objects collide with real objects, the result is the same as usual.

The principle of 2D/3D conversion stems from Ross' psychophysics theory, in which stereoscopic perception can be generated using the Pulfrich effect, which refers to the time delay effect caused by the difference of the amount of light in both eyes [6].



(a)



(b)

Fig. 2. 2D/3D conversion methods: (a) modified time difference (MTD) method, (b) computed image depth (CID) method.

[6] generated a stereo image pair by combining current and delayed images appropriately and then displaying them to both human eyes. Since this research, several 2D/3D conversion methods have been developed to allow people to enjoy 3D experiences with formerly created 2D content.

The MTD (Modified Time Difference) method shown in Fig. 2(a) detects movements of objects and then decides delay directions and times according to the characteristics of the movements. Then, stereoscopic images are selected according to the time differences in 2D image sequences [7]. The CID (Computed Image Depth) method, as shown in Fig. 2(b), uses the relative position between multiple objects in still images. Image depth is computed by using the contrast, sharpness and chrominance of the input images [8].

2.2. 3D Codec

Stereoscopic and multi-view videos can provide more vivid and accurate information about scene structures than monocular videos, since they provide depth information. Transmitting stereo sequences requires twice the amount of bandwidth as that required in conventional 2D sequences. However, stereoscopic sequences have high correlations with each other. This property can be used to reduce a considerable number of coding bits. The quantity of multi-view data increases according to the number of views. This has been a serious problem when implementing multi-view video systems. Therefore, an efficient compression technique is necessary for multi-view video coding. For this, both the spatial redundancy in the view axis and the temporal redundancy in the time axis have to be efficiently exploited and reduced. Decoded multi-view sequences can be properly displayed on various types of displays including 2D, stereo and 3D/multi-view TVs as shown in Fig. 3.

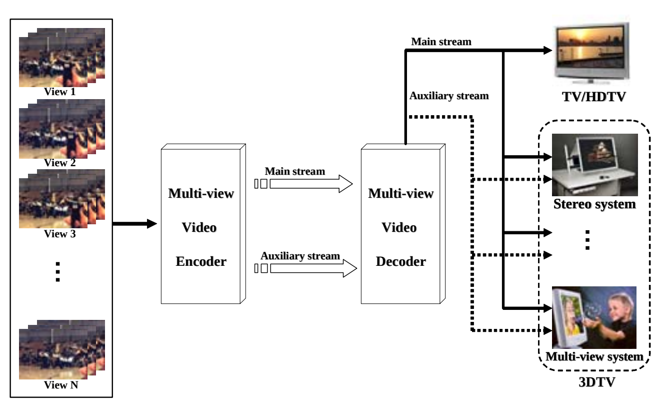


Fig. 3. Multi-view video codec with view scalability.

The MPEG-3DAV AHG (Ad-hoc Group) was organized at the 58th MPEG meeting. After determining the MVC reference software at the 75th MPEG meeting, several CEs (Core Experiment) were used to standardize the MVC in detail. The issues of MPEG-MVC were moved to the JVT (Joint Video Team) at the 77th MPEG meeting. JVT-MVC is expected to be standardized by April 2008.

2.3. 3D Displays

The most important goal of 3D imaging systems is to create realistic images of dynamically changing scenes. The ways to obtain arbitrary view generation from multiple images can be subdivided into two classes, IVS (Intermediate View Synthesis) and virtual view rendering as shown in Fig. 4 [9].

IVS performs disparity-compensated interpolation or projection of image textures to synthesize realistic images. This straightforward approach represents the interpolation of a pair of stereo images, where the scaling of disparities allows the selection of a viewpoint on the interocular axis between the cameras. The approach is very simple and works very fast, but the main problem is occlusion, which means that no reasonable disparity data is available within areas that are only visible from one of the cameras. The second problem is that the viewpoint is limited to the position of the baseline between the cameras.

On the other hand, the second approach also allows the building up of a 3D model represented by a 3D voxel based on the silhouette and disparity fields of objects. For virtual view generation, the 3D surface model can be rotated towards the desired place and orientation, and the texture data extracted from the original camera can be projected onto this surface. Generated 3D objects can easily be manipulated, e.g., rotated, to render images from different viewpoints.

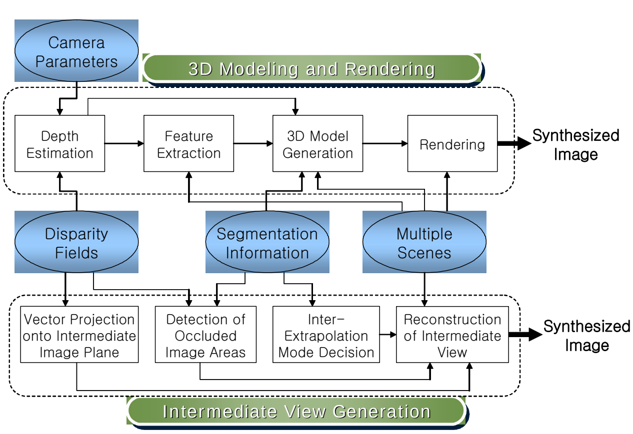


Fig. 4. Two approaches for arbitrary view generation.

3. Experimental Results

3.1. Mixed Reality Content

Z-keying is the simplest and popular method today of synthesizing two images when the depth information is known. In real 3D images, disparity vectors can be converted into depth information. Since virtual objects rendered by computer graphics already have depth information, the virtual objects can be mixed with the real world image by comparing the depth of virtual objects with the estimated depth values of real scenes. The pixels which are closer to the user's viewpoint are displayed. As a result, virtual objects can be placed in any desired positions in real world scenes. Fig. 5 shows the process of the Z-keying and its results.

We can reconstruct 3D models of scenes by using estimated depth maps. Fig. 6(a) shows some scenes of reconstructed models from several viewpoints. With reconstructed 3D models, we can realize the interaction of virtual objects with real objects. Fig. 6(b) shows snapshots of interaction with a virtual ball. In the first row, the ball disappears naturally behind the front board when it falls to the ground. The second row shows the active interaction between the real and virtual worlds. The virtual ball is pitched to the front board. When the ball strikes the board, it bounces back in the direction of the mirror.

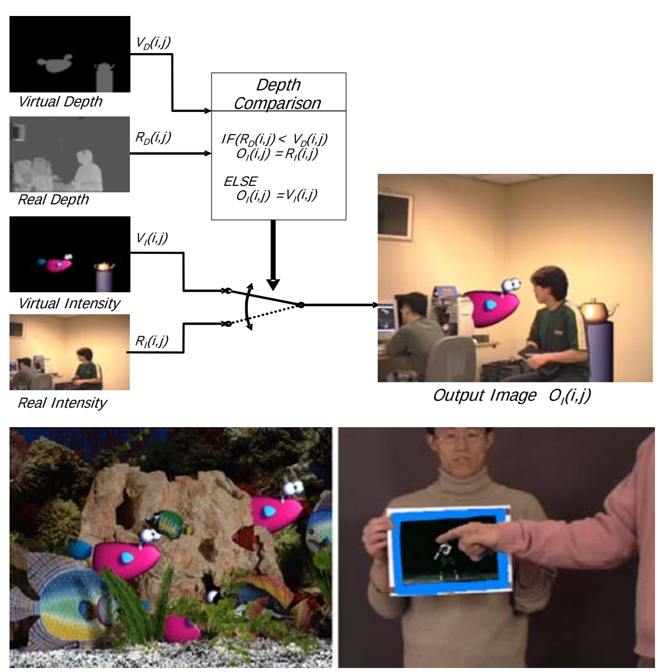
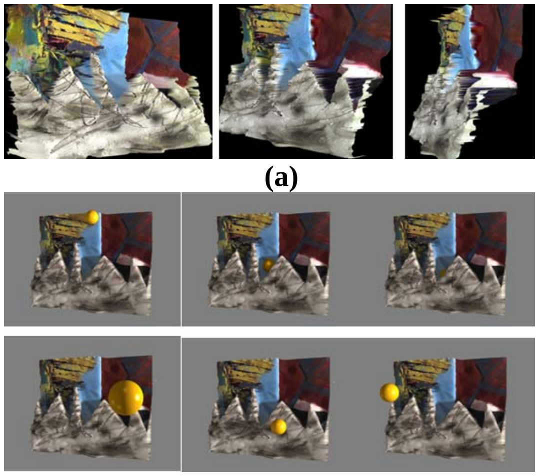


Fig. 5. Z-Keying.



(a) (b)

Fig. 6. Interaction in MR systems: (a) 3D reconstruction from a stereo pair, (b) interaction.

3.2. 2D/3D Conversion

MMD (Modified Motion-to-Disparity) method that considers multi-user conditions and stereoscopic display characteristics has been developed [10]. In this method, the scale factors of motion-to-disparity conversion are determined by considering several cues such as magnitude of motion, camera movement and scene complexity. Fig. 7 shows some 2D/3D conversion results obtained by the MMD method.

(a) (b) (c) (d)

Fig. 7. Results of 2D/3D conversion obtained by the Modified MD method: (a) original image, (b) motion fields, (c) synthesized stereo pair, (d) interlaced stereo image.

4. Summary

This paper has presented the overview of 3D video processing technologies for 3DTV such as 3D content generation, 3D video codec and video processing techniques for 3D displays.

One of the most serious problems of 3D video technology has been lack of 3D content. For this reason, many recent methods to generate 3D content have been developed. These methods include using 3D CG, capturing images with multiple cameras, and converting 2D-to-3D content. The quantity of multi-view data can be increased according to the number of views. To efficiently encode multi-view videos, both the spatial redundancy of the view axis and the temporal redundancy of the time axis have to be efficiently exploited and reduced. Current standard activities for MVC have been presented in this paper. We also showed some approaches for arbitrary view generation from multiple images, such as IVS and virtual view rendering.

5. References

1. A. Redert, I. Sexton and P. Surman, ATTEST – Advanced Three-Dimensional Television System Technologies, Proc. of 1st International Symposium on 3D Data Processing, Visualization and Transmission, pp.313-319, Padova, Italy, Jun. 2002.
2. L. Onural, T. Sikora and A. Smolic, An overview of a new European consortium integrated three-dimensional television-capture, transmission and display, EWIK, Semantics and Digital Media Technology, 2004.
3. C. V. Berkel and D. W. Parker, Multiview 3D-LCD, Proc. SPIE 2653, pp.32-39, 1996.
4. N. Hur and C. Ahn, Experimental service of 3DTV broadcasting relay in Korea, Proc. SPIE 4864, pp. 1-13, 2002.
5. Y. Ohta and H. Tamura, Mixed Reality, Springer-Verlag New York Inc., 1999.
6. J. Ross and J.H. Hogben, The Pulfrich effect and short-term memory in stereopsis, Vision Research 15, pp.1289-1290, 1975.
7. T. Okino and H. Murata, New television with 2D/3D image conversion technologies, SPIE, vol. 2653, pp.96-103, 1996.
8. H. Murata, Y. Mori, A real-time 2D to 3D image conversion technique using computed image depth, SID, IGEST, pp.919-922, 1998.
9. E.M. Izquierdo and S. Kruse, Image Analysis for 3D Modeling, Rendering, and Virtual View Generation, Computer Vision and Image Understanding, vol.71, no.2, pp.231–253, 1998.
10. D. Kim, Stereoscopic Image Conversion using Motion Parallax and Structure Estimation, MS Thesis, Yonsei Univ., 2006.