FTV (Free Viewpoint Television) Creating Ray-Based Image Engineering

Masayuki Tanimoto
Graduate School of Engineering, Nagoya University
Furo-cho, Chikusa-ku, Nagoya, 464-8603 Japan

ABSTRACT

We have been developing ray-based 3D information systems that consist of ray acquisition, ray processing, and ray display. FTV (Free viewpoint TV) based on the ray-space method is a typical example. FTV will bring an epoch-making change in the history of television because it enables us to view a distant 3D world freely by changing our viewpoints as if we were there. We constructed a real-time FTV including the complete chain from capturing to display. A new algorithm was developed to generate free viewpoint images. We need a new user interface for FTV to make full use of 3D information of FTV. Two types of user interface were made using 2D and automultiscopic displays with a head tracking system. We are creating "Ray-Based Image Engineering" through these works.

1. INTRODUCTION

Television has a long and successful history since it realized a human dream of seeing a distant world in real-time. It keeps the position of the most important visual information system. However, TV shows us the same scene even if we move our viewpoints in front of the display. It is quite different from what we experience in the real world. In TV, users can get only a single view of a real 3D world they want to see. The view is determined not by users but by a camera placed in the 3D world. Although many important technologies have been developed, this function of TV has never changed.

TV has been developed by pixel-based technologies. However, the most essential element of visual systems is not pixel but ray. We have been developing ray-based 3D information systems that consist of ray acquisition, ray processing, and ray display. FTV (Free Viewpoint Television) [1], [2] is a typical example of such systems.

FTV will bring an epoch-making change in the history of television because it has a new function to view the 3D world freely as if we were there. The ray-space method [3], [4] enables it.

We proposed the concept of FTV and verified its feasibility by the world's first real-time experiment.

FTV brings a new frontier to the field of signal processing since the signal processing of FTV such as coding and view image generation is performed in the ray-space that is a new domain with higher dimension than that of the current TV. A new user interface is also needed for FTV to make full use of 3D information of FTV.

FTV makes a large market for industry since it is ultimate 3D TV that can transmit all visual information of 3D space, a new tool for art and content creation, and an information infrastructure for secure society. FTV can be a platform for various important 3D systems and the FTV signal can be a common signal format for the 3D systems.

We proposed FTV to MPEG [1] and have been making contributions. It is considered the most challenging scenario in 3DAV of MPEG and the standardization of multiview video coding has started.

It should be noted that FTV is not a conventional pixel-based system but a ray-based system. We have been developing the technologies of ray acquisition, ray processing, and ray display.

2. RAY-SPACE REPRESENTATION

2.1. Ray-space method

We have been developing FTV based on the ray-space representation [3], [4]. In the ray-space representation, one ray in the 3D real space is represented by one point in the ray space. The ray-space is a virtual space. However, it is directly connected to the real space. The ray-space is generated easily by collecting multi-view images with the consideration of camera parameters.

As defined in Fig.1, the ray-space is 4-dimensional and 5-dimensional including time. If we place cameras at limited region, the obtained rays are limited and the ray-space constructed from these rays is a subspace of the ray-space. For example, if we place cameras on a line or a circle, we have only one part of data of the ray-space. In such cases, we define smaller ray-space. One is ray-space in the orthogonal coordinate and another is ray-space in the spherical coordinate. The orthogonal ray-space is used for the linear camera arrangement and the spherical ray-space is used for the circular camera arrangement.

For the linear camera arrangement, the ray space is constructed by placing many camera images vertically in parallel as shown in Figure 2, forming the FTV signal. The FTV signal consists of many camera images and the horizontal cross-section has line structure as shown in Figure 3. Vertical cross-sections of the ray space give view images at the corresponding viewpoints.

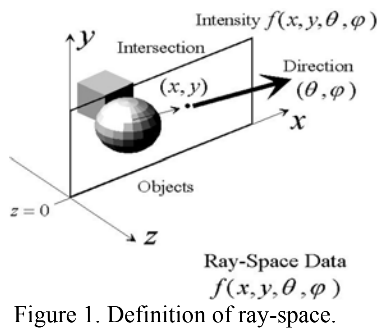

Figure 1. Definition of ray-space.

0-7803-9134-9/05/$20.00 ©2005 IEEE

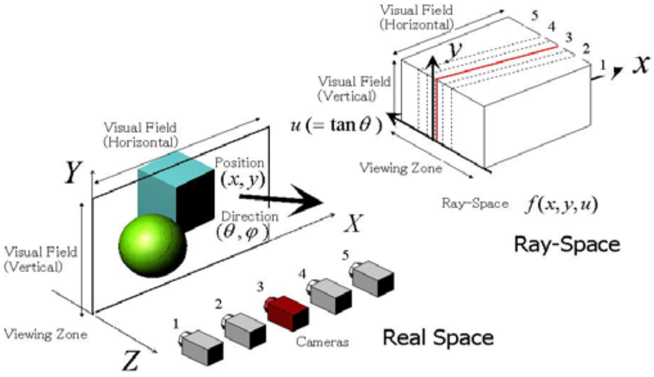

Figure 2. Acquisition of FTV signal.

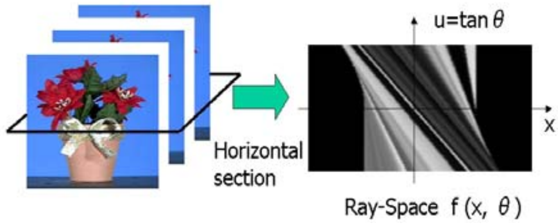

Figure 3. An example of FTV signal.

Table 1. Representation methods for free viewpoint and their features

Method	Data Acquisition	Data Conversion	View Generation	Quality of Generated View Image	Transparency
Image Domain	Direct acquisition	No	Warping /Projection	Not precise	High
Integral Photography	Direct acquisition (precisely aligned)	Optically	Optically	Precise	High
Ray-Space	Calibration & Registration	Coordinate Transform	Memory access/ Interpolation	Precise	High
Surface Light Field	Calibration & Registration	Decomposition Approximation	Texture mapping /pixel-by-pixel multiplication	Not precise in the case of approximation	Low
Model-Based	Calibration & Registration	3D model Texture	Texture mapping	Not precise	Low

2.2. Characterization of representation methods for free viewpoint

There are several kinds of representation methods for free viewpoint. They are characterized in Table 1.
In the image domain methods, the data are kept as a set of images captured by cameras and free viewpoint images are generated by their switching, warping or projection. "EyeVision" [5] belongs to this category.
If the camera density is very dense as in Integral Photography [6], view generation is performed simply by selecting camera image or by collecting pixels from camera images. If the camera density is not very dense but still dense, intermediate views can be generated precisely by interpolation of camera images using camera parameters. This is referred to Ray-Space in the table. Light field [7] belongs to this category. If the camera density is sparse, intermediate views can be generated by detecting objects in the scene. This is referred to Model-Based in the table. Generated views are not precise in this method. Surface Light Field [8] is a hybrid method of Ray-Space and Model-Based methods. It uses both the surface model of object and the information of rays.
Transparency in the table denotes scene independency. It is an important factor because if a method has low transparency, its application is very limited.

3. FTV SYSTEM

3.1. Real-Time System

We constructed a real-time FTV system covering the complete chain from video capturing to display [1], [9]. It consists of 16 cameras, 16 clients, one server and Ethernet. Each client has one camera and connected to the server by Gigabit Ethernet. The function of FTV was successfully demonstrated by generating photo-realistic free viewpoint images of the moving scene in real-time.
Figure 4 shows an experimental system of FTV. Examples of the generated free viewpoint images are shown in Figure 5. Complicated natural scenes including sophisticated objects

Figure 4. Experimental system of FTV.

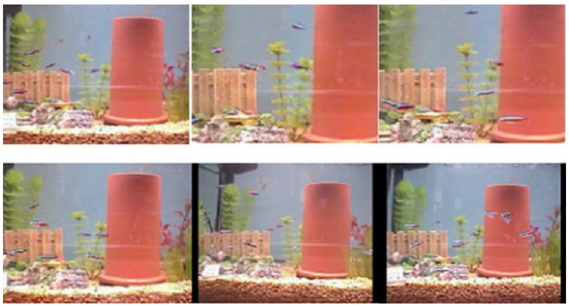

Figure 5. Examples of generated FTV images.

such as small moving fishes, bubbles and reflection of light from the aquarium glass are reproduced very well.

"100-camera system" has been developed to capture larger space by Nagoya University (Intelligent Media Integration COE and Tanimoto Lab.). The system consists of one host server PC and 100 client PCs (called 'nodes') that are equipped with JAI PULNiX TM-1400CL cameras. The interface between camera and PC is Camera-Link. The host PC generates a synchronization signal and distributes it to all the nodes. This system is capable of capturing not only high-resolution video with 30 fps but also analog signal up to 96 kHz. The camera setting is flexible as show in Figure 6.

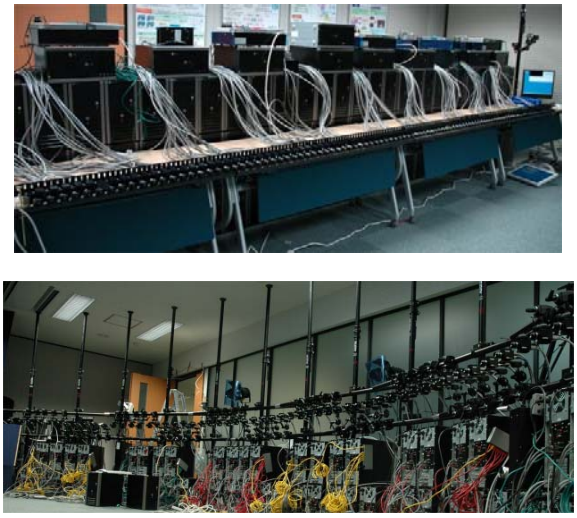

Figure 6. Linear camera setup (upper) and half-round camera setup (lower) of 100-camera system.

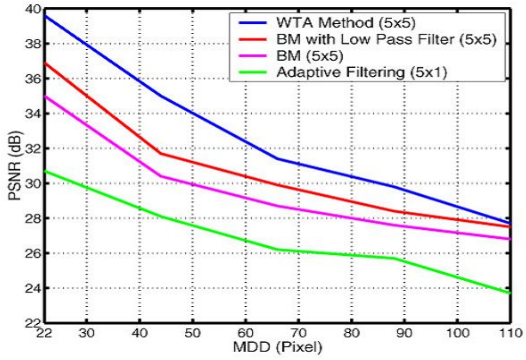

Figure 7. PSNR of interpolated view image as a function of maximum disparity difference.

3.2. Processing of FTV

The signal processing of FTV is performed in the ray-space. The hierarchical ray-space [10] is also used for processing.

The quality of generated view images strongly depends on the interpolation. We proposed several interpolation schemes of the ray-space [11]-[13]. PSNR of the interpolated image is shown as a function of MDD (Maximum Disparity Difference) for various interpolation schemes in Figure 7 [12]. For a wide range of MDDs, the WTA method gives about 1-4 dB higher PSNR than simple BM (Block Matching). Here, WTA denotes the block-based Winner-Takes-All scheme. It uses an energy function that forms the relation between similarity measures and neighborhood regularization of the disparity map. Then the energy function is minimized using a simple winner-takes-all scheme.

Further 2dB improvement of PSNR is achieved by loopy belief propagation (LBP) method [13].

The developed algorithm is simplified and implemented on a note PC. Thus, real-time FTV on a note PC is realized.

4. DISPLAY OF FTV

FTV needs a new user interface to display free viewpoint images. Two types of user interface were made using 2D and automultiscopic displays with a head tracking system.

Many head tracking systems have been proposed using magnetic sensor, various optical marker, infrared camera, retroreflective light from retina, etc. Our head tracking system uses only a conventional 2D camera and detects the position of a user's head by image processing. The user doesn't need to attach any markers or sensors.

The head tracking system detects the location of the viewer's head as follow: The camera placed on the display captures the viewer with the background. Then, the region of moving objects is obtained by calculating the difference between the previous and present frames. For each horizontal position, we calculate the length of the vertical line within the moving area. The place with the longest length is regarded as the location of the viewer. Repeating this process for a few frames, we calculate the time-average of the location of the viewer. This averaging prevents the detection of small unwilling movement of the head. The

distance between the user and display is estimated by the size of the moving area.
In the user interface using a 2D display, the location of the user's head is detected with the head tracking system and the corresponding view image is generated as described before. Then, it is displayed on the 2D display.
Automultiscopic displays enable a user to see stereoscopic images without special glasses. However, it has two problems: limited viewing zone and discreteness of motion parallax. Because the width of viewing zone for each view equals the interpupillary distance approximately, the view image does not change by viewer's movement smaller than the width. On the other hand, when the viewer moves over the zone, the view image changes suddenly.
The reason of limited viewing zone is that the display has only a few number of views. However, since each view has many sidelobes, moving out of original viewing zone repeats exactly the same perspective progression. Therefore, we extend the viewing zone by feeding the view channel corresponding to viewer's position with FTV image.
In the user interface using the automultiscopic display as shown in Figure 8, the function of providing motion parallax is extended by using the head tracking system. The feeded images change according to the movement of head position for providing small motion parallax, and the view channel to feed image is switched for handling large motion. It means that binocular parallax for naked eyes is provided by automultiscopic display and motion parallax is provided by head tracking and changing image adaptively.
These two systems are for a single user. We use another type of ray reproducing display [14] for multi users.

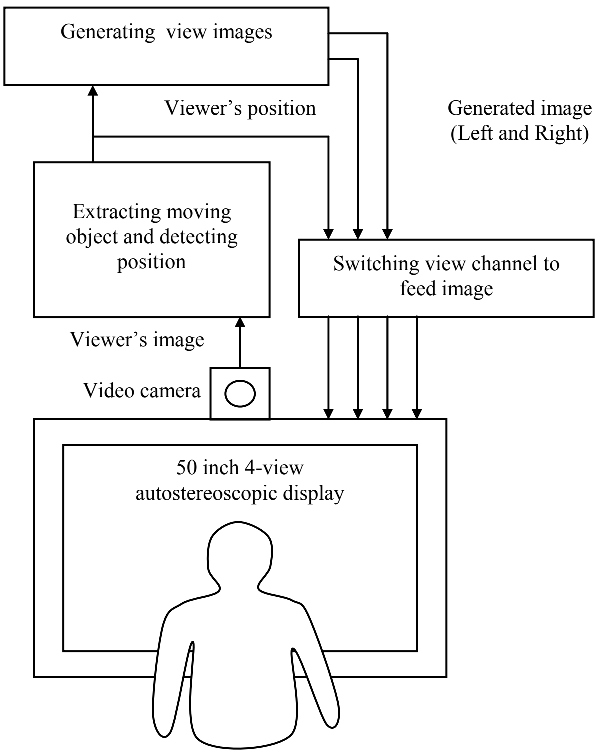

Figure 8. Configuration of FTV display using automultiscopic display.

5. CONCLUSION

FTV is a next generation television where users can determine their viewpoints freely. Unlike conventional TVs, where users can get only a single view of a real 3D world they want to see, FTV provides the users with unrestricted visual field of the 3D world. We have realized FTV based on the ray-space representation.
Ray is the most essential element of visual systems. We have been developing ray-based 3D information systems that consist of ray acquisition, ray processing, and ray display. FTV is one of such systems.
We have also been developing new types of ray acquisition [15] and display [14] systems. We are creating "Ray-Based Image Engineering" through these works.

6. REFERENCES

[1] M. Tanimoto and T. Fujii, "FTV－Free Viewpoint Television", ISO/IEC JTC1/SC29/WG11, M8595, July 2002.
[2] T. Fujii, M. Tanimoto, "Free-viewpoint Television based on the Ray-Space representation, " Proc. SPIE ITCom 2002, pp.175-189, August 2002.
[3] T. Fujii, "A Basic Study on the Integrated 3-D Visual Communication", Ph.D thesis of engineering, The University of Tokyo, 1994(in Japanese).
[4] T. Fujii, T. Kimoto, M. Tanimoto, "Ray Space Coding for 3D Visual Communication", Picture Coding Symposium '96, pp. 447-451, Mar. 1996.
[5] http://www.ri.cmu.edu/events/sb35/tksuperbowl.html.
[6] F. Okano, J. Arai, H. Hoshino, and I. Yuyama, "Three-dimensional video system based on integral photography", Opt. Eng., 38(6), 1072-1077, 1999.
[7] M. Levoy, and P. Hanrahan, "Light field rendering", SIGGRAPH '96, pp. 31-42, 1996.
[8] W. Chao Chen, J. Y. Bouguet, M. H. Chu, R. Grzeszczuk, "Light Field Mapping: Efficient Representation and Hardware Rendering of Surface Light Fields". ACM Transactions on Graphics. 21 (3), pp. 447-456, 2002.
[9] Na Bangchang, T. Fujii, M. Tanimoto, "Experimental System of Free viewpoint TeleVision", Proc. IST/ SPIE Symposium on Electronic Imaging, Vol. 5006-66, pp. 554-563, Jan. 2003.
[10] M. Tanimoto, A. Nakanishi, T. Fujii, T. Kimoto, "The Hierarchical Ray-Space for Scalable 3-D Image Coding", Picture Coding Symposium 2001, pp.81-84, April 2001.
[11] A. Nakanishi, T. Fujii, T. Kimoto, M. Tanimoto, "Ray-Space Data Interpolation by Adaptive Filtering using Locus of Corresponding Points on Epipolar Plane Image", The Journal of the Institute of Image Information and Television Engineers (ITE), Vol. 56, No. 8, pp. 1321-1327, 2002.
[12] M. Droese, T. Fujii and M. Tanimoto, "Ray-Space Interpolation based on Filtering in Disparity Domain", Proc. 3D Conference 2004, pp. 213-216, Tokyo, Japan, June 2004.
[13] M. Droese, T. Fujii and M. Tanimoto,"Ray-Space Interpolation Constraining Smooth Disparities Based On Loopy Belief Propagation", Proc. of IWSSIP2004, pp. 247-250, Poznan, Poland, Sept. 2004.
[14] T.Endo, Y.Kajiki, T.Honda, M.Sato, "Cylindrical 3-D Video Display Observable from All Directions", Proceedins of Pacific Graphics 2000, pp.300-306, 2000.
[15] T. Fujii, M. Tanimoto, "Real-Time Ray-Space Acquisition System", SPIE Electronic Imaging, vol. 5291, pp. 179-187, Jan. 2004.