INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11 MPEG2012/M30229
July 2013, Vienna, Austria

Source **NISRI, NICT, KDDI, NTT, 3Dragons LLC, Poznan University of Technology, MERL, Telecom SudParis, Fraunhofer HHI**
Status **Proposal**
Title **Proposal on a New Activity for the Third Phase of FTV**
Author Masayuki Tanimoto
Takanori Senoh
Sei Naito
Shinya Shimizu
Hideyoshi Horimai
Marek Domański
Anthony Vetro
Marius Preda
Karsten Mueller

Abstract

MPEG has been conducting the standardization of MVC and 3DV successfully. MVC was the first phase of FTV (Free-viewpoint Television) and 3DV is the second phase of FTV. Based on the recent development of 3D technology, this document proposes a new 3D standardization activity targeting super mul-tiview and free navigation applications for the third phase of FTV.

1 Introduction

MPEG has been conducting the standardization of MVC and 3DV successfully. MVC was the first phase of FTV (Free-viewpoint Television) [1]-[4] and 3DV is the second phase of FTV. Based on the recent development of 3D technology, this document proposes a new 3D standardization activity targeting super multiview and free navigation applications for the third phase of FTV.

2 History of FTV Standardization in MPEG

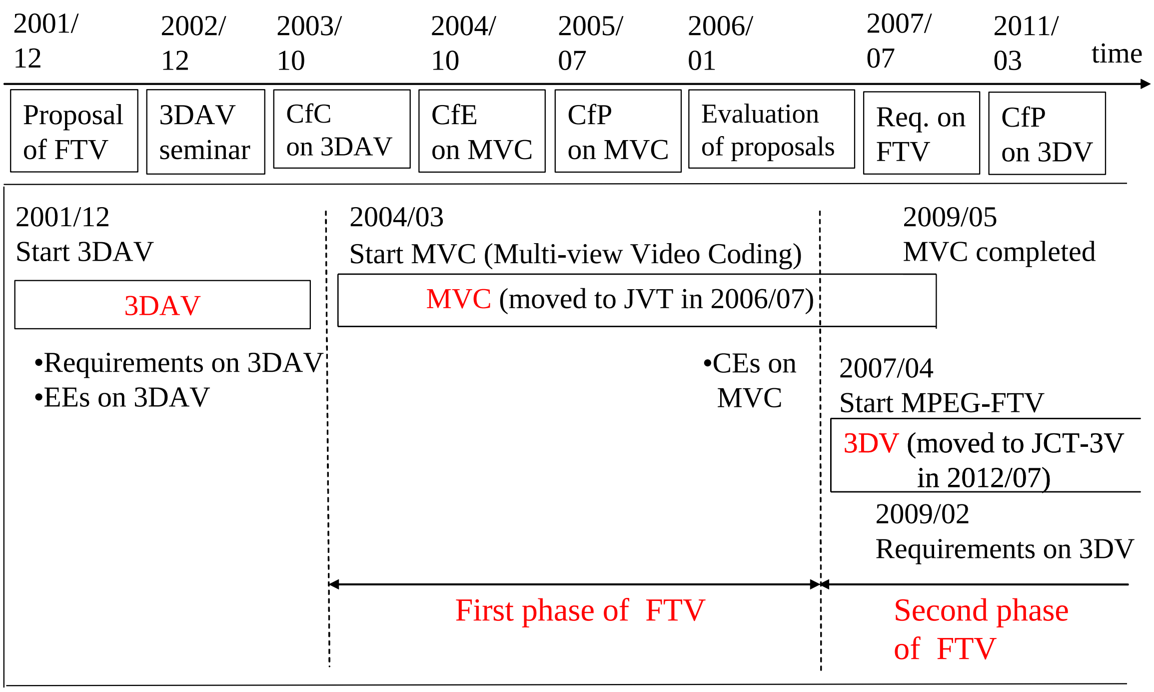

Fig. 1 History of FTV standardization in MPEG.

MPEG has been conducting FTV standardization as shown in Fig. 1. In 2001, FTV was proposed to MPEG and the 3DAV (3D Audio Visual) activity started. FTV is an innovative 3DTV that allows users to see a 3D scene by freely changing the viewpoint. As shown in Fig. 2, many 3D topics such as omni-directional video, FTV, stereoscopic video and 3DTV with depth disparity information were discussed in 3DAV. The discussion converged on FTV in 2004.

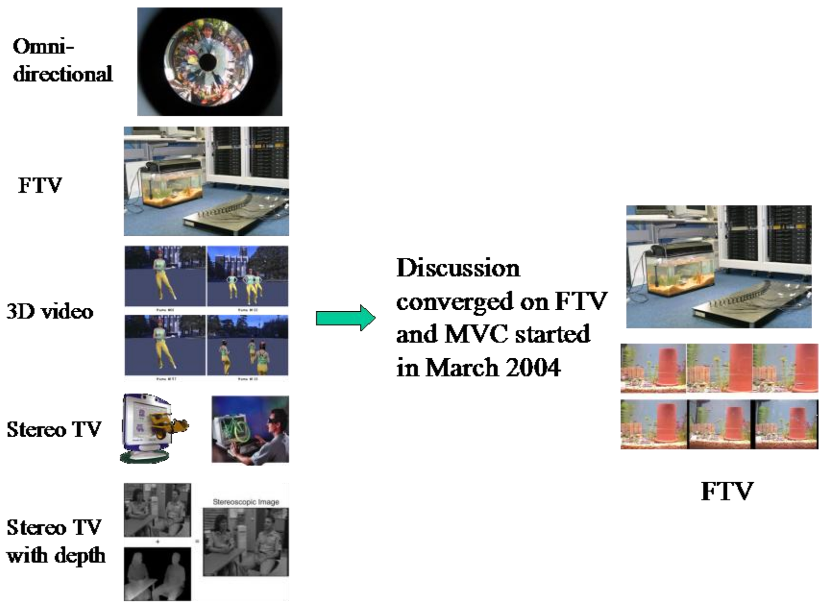

Fig. 2. Discussion at 3DAV.

Many technologies are needed to realize FTV. MPEG has been conducting the standardization of FTV step by step. The first phase of FTV was MVC (Multi-view Video Coding). The framework of MVC is shown in Fig. 3. MVC targeted efficient coding of multi-view video. MVC started in

2004 and was completed in 2009. MVC has been adopted by Blu-ray 3D. In MVC, the number of input views is the same as that of output views. The view synthesis function of FTV is not included in MVC.

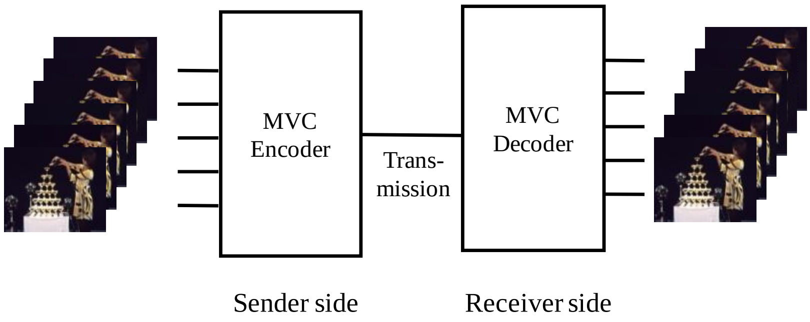

Fig. 3 Framework of MVC.

Before completing MVC, 3DV started as the second phase of FTV in 2007. 3DV targets multiview 3D display applications. The framework of 3DV is shown in Fig.4. 3DV sends small number of views and generates large number of views at the receiver side for multiview 3D displays. Thus, view synthesis is introduced into 3DV. This view synthesis is assisted by depth information. The standardization of multiview plus depth coding and data format has been in progress in JCT-3V.

However, the view synthesis of 3DV has large limitation. Views are synthesized only along a horizontal base line of linear camera setup. Multiview displays with views less than about 30 are considered in 3DV. These limitations have to be reduced to introduce further FTV applications.

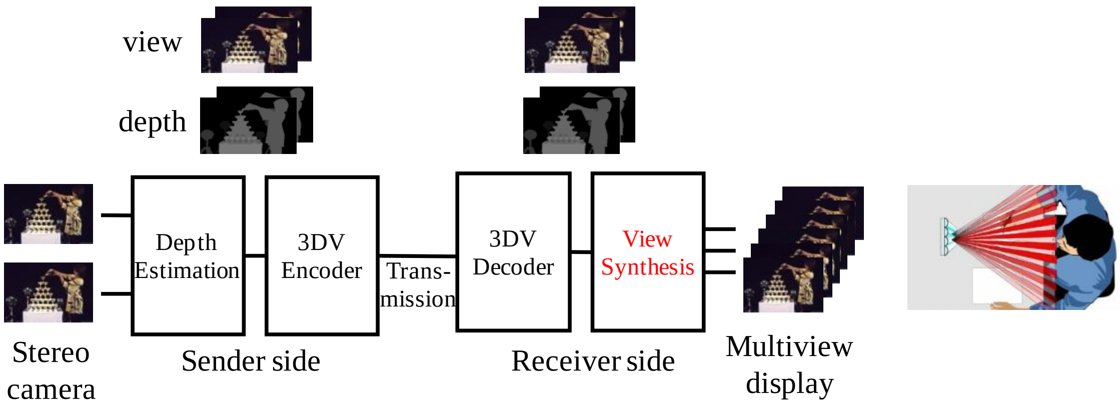

Fig. 4 Framework of 3DV.

3 Background and Motivation

2022 FIFA World Cup Japan Bid Committee planed to deliver the excitement on soccer stadium to the world by FTV. It aimed to revolutionize the viewing of the soccer game by free navigation and realistic 3D viewing as shown in Fig.5. Super mul-tiview displays with hundreds views will be needed for the realistic 3D viewing.

Fig. 5 Revolutionizing viewing by FTV.

Fig. 6 shows the progress of 3D capture and display capabilities. In this figure, the ability of 3D capture and display is expressed by a factor of the pixel-view product, defined as "number of pixels" times "number of views". It is seen that the pixel-view product has been increasing rapidly year after year in both capture and display. This rapid progress of 3D capture and display capabilities enables the introduction of super multiview and free navigation applications.

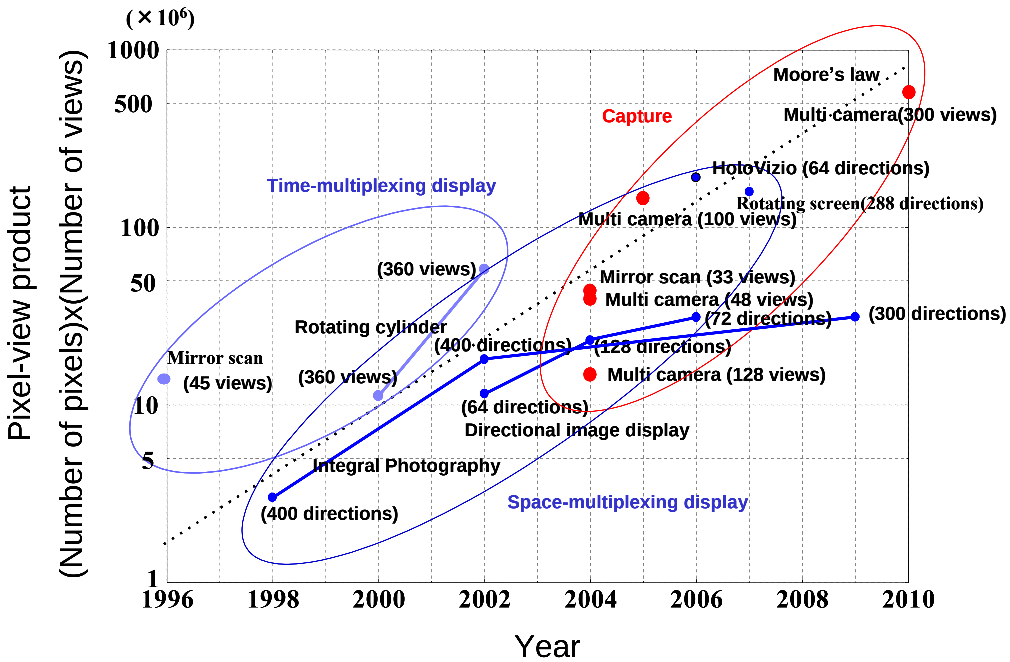

Fig. 6 Progress of 3D capture and display capabilities.

4 Application Scenarios

The third phase of FTV targets two application scenarios.

(1) Super multiview

3D displays with hundreds views have become available. Fig.7 shows a portable 360-degree viewable 3D display, “Holo-Table” [5]. It displays more than 500 views with 1024x768 pixels at 30-60 fps. The feature of this display is not only large number of views but also high density of views. Because of these features, users can see a 3D scene from any directions with smooth motion parallax. Fig.8 shows examples of displayed images in different directions.

Fig. 7 Portable 360-degree viewable 3D display, “Holo-Table”.

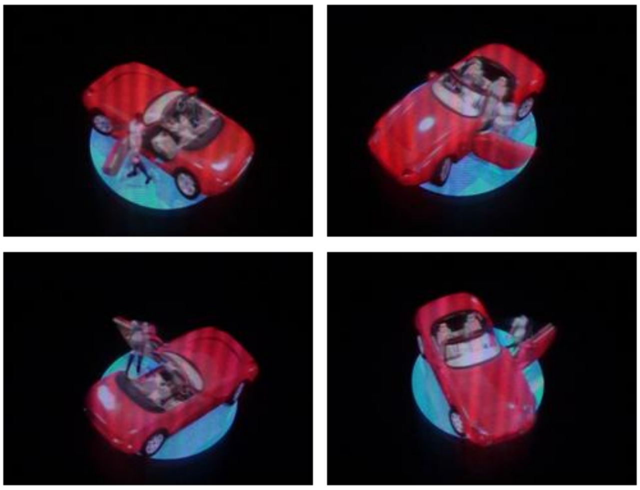

Fig. 8 Examples of displayed images in different directions.

Super multiviews with both horizontal parallax and vertical parallax are realized by IP (Integral Photography) displays [6], [7].

Super multivew provides useful applications in such areas as medical, robotics, design, education and so on.

(2) Free navigation

The free navigation applications are considered as a potential new market for the communication companies. Delivery of free navigation video with accompanying spatial audio could be delivered by internet as a new service. The rendering of the virtual view is foreseen as an external service for the customer that may even use low-power mobile devices as shown in Fig.9.

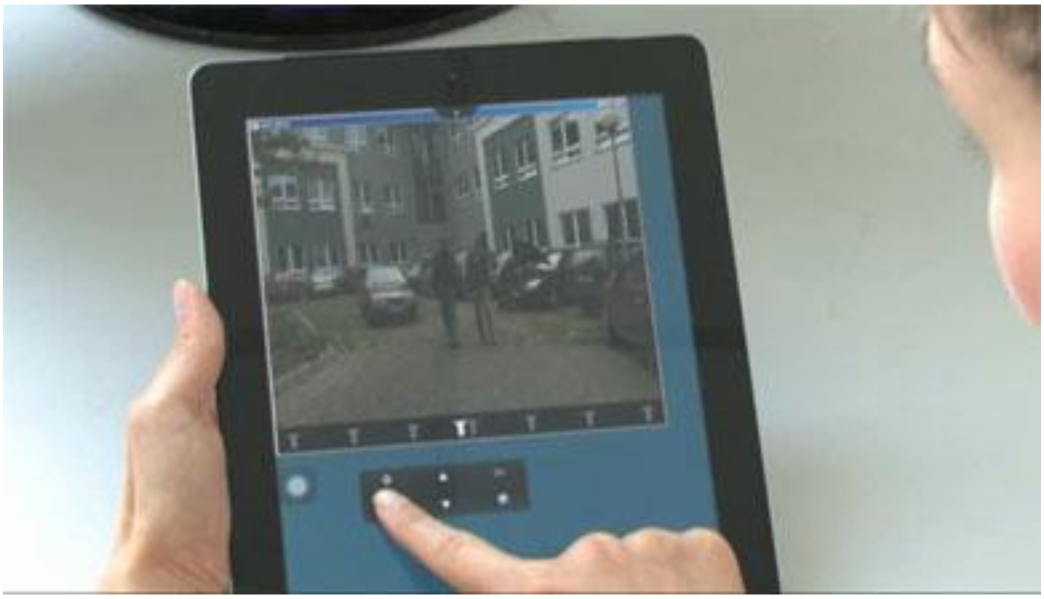

Figure 9: A FTV mobile application with external virtual view rendering (courtesy of Orange Labs Poland).

Walk-Through and Fly-Through Experience has been developed using 3D model plus texture mapping toward Free Viewpoint TV [8], [9].

This technology enables audiences of TV programs to see soccer games from the viewpoint of players on the ground, where conventional TV camera cannot be mounted. Furthermore, the viewpoint can be moved even through specific players as shown in Fig.10 [8]. This technology is based on “3D model plus texture mapping” [9], [10]. The 3D model of each object such as human region is reconstructed from multiview images as shown in Fig.11, and the virtual viewpoint is reproduced using the 3D model and textures acquired from the appropriate cameras.

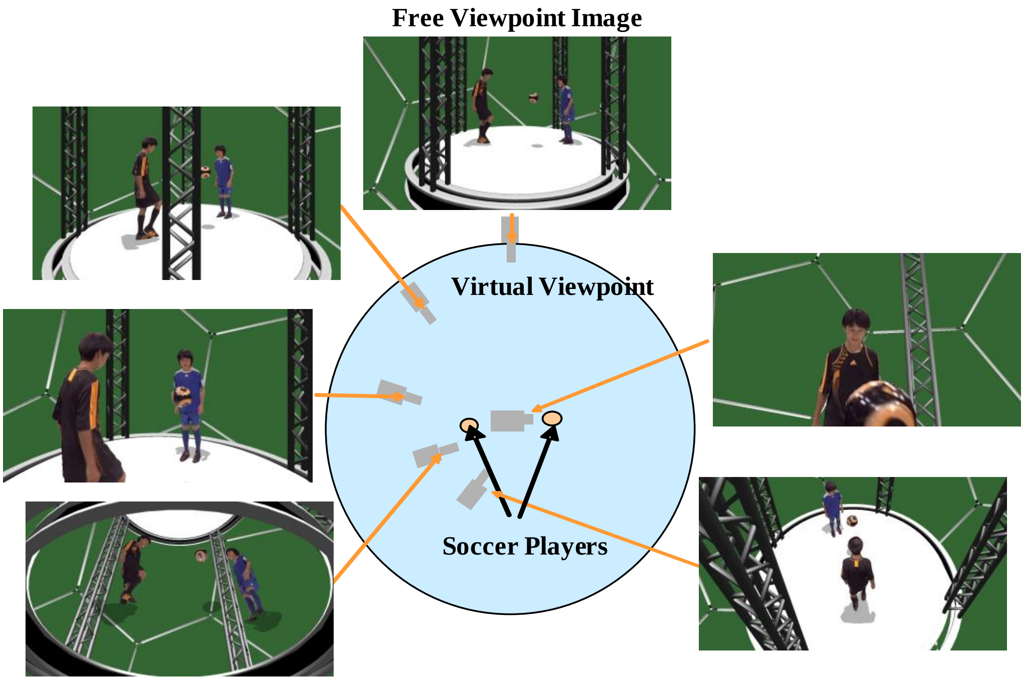

Fig. 10 Walk-through and fly-through experience.

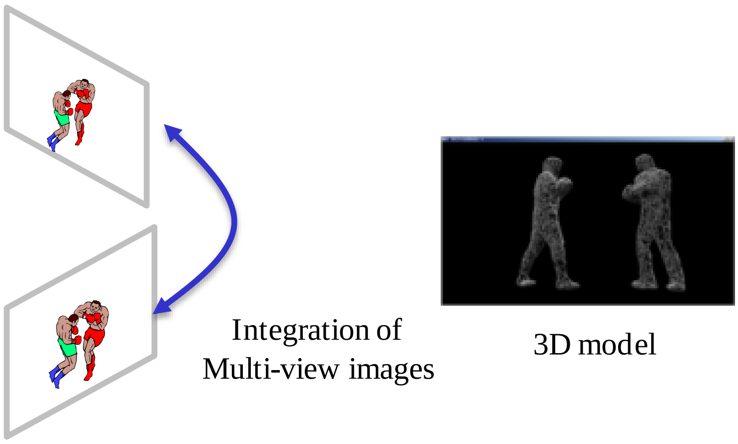

Fig.11 3D modelling based on multiview images.

Free navigation applications provide virtual views from arbitrary points on a trajectory of navigation. The views might be monoscopic but stereoscopic view will provide more realistic navigation through a scene.

Possible applications include:

- Interactive courses – the student may watch a demonstration from an arbitrary viewpoint,
- Interactive manuals,
- Teaching resources,
- Interactive movies – the viewers are virtually walking though e scene,
- Interactive theatre performance,
- Interactive circus performances,
- Sport transmissions,

- In particular: boxing, sumo, judo, wrestling – ability too watch the fight from various viewpoints
- Tourism
- Cultural heritage

5 Requirements

(1) Super multiview
camera arrangement: linear, planar, circular
camera density: dense
view synthesis: rendering from linear, planar, circular camera setup
coding: efficient real-time coding of hundreds views

(2) Free navigation
camera arrangement: circular, arbitrary
camera density: sparse
view synthesis: rendering from sparse camera setup, rendering from circular and arbitrary camera setup
coding: efficient real-time coding of distributed cameras

(3) Data structure
image-based structure: direct rendering, multiview+depth map conversion to a canonical format
3D model-based structure: 3D model construction from multi-view images, 3D image rendering
efficiency: feasible bitrate, CPU time, memory size.
versatility: linear/arc/circular/2D array/spherical/random camera arrangement

These requirements have not yet been tested in MPEG.
Production of test sequences is already well advanced as shown in Fig. 12.

Figure 12: Production of a test sequence with circular arrangement of cameras (widok = view, czas = time)

Production of test sequences needs precisely synchronized cameras that can be controlled from one location – an arrangement is shown in Fig. 13.

Figure 13: Test sequence production.

Some preliminary test sequences have been produced as shown in Fig.14. The results demonstrate ability to estimate the location of objects in the space using cameras that are only roughly located around a scene.

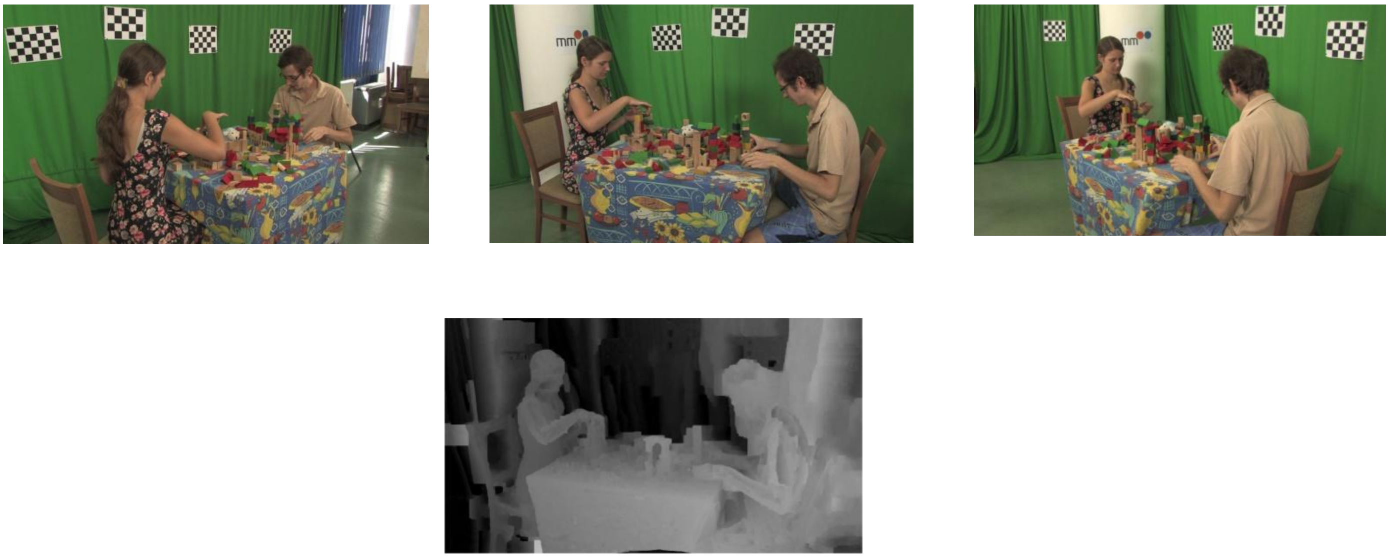

Figure 14: A test sequence produced with circular arrangement of camera, below – a depth map.

6 Conclusion

Super multiview and free navigation are very useful and challenging application scenarios. We request to start the third phase of FTV to investigate them.

References

[1] Masayuki Tanimoto, Mehrdad Panahpour Tehrani, Toshiaki Fujii, Tomohiro Yendo, "Free-Viewpoint TV", IEEE Signal Processing Magazine, vol.28, no.1, pp.67-76, January 2011.

[2] Masayuki Tanimoto, Mehrdad Panahpour Tehrani, Toshiaki Fujii, Tomohiro Yendo, "FTV for 3-D Spatial Communication", Proceedings of the IEEE, Vol. 100, No. 4, pp. 905-917 (April 2012). (invited paper)

[3] Masayuki Tanimoto, "FTV: Free-viewpoint Television", Signal Processing : Image Communication, Vol. 27, Issue 6, pp. 555-570 (June 2012). doi:10.1016/j.image.2012.02.016 (online 24 February 2012). (invited paper)

[4] Masayuki Tanimoto, "FTV (Free-viewpoint Television)", APSIPA Transac-tions on Signal and Information Processing, Vol. 1, Issue 1, e4 (14 pages) (August 2012). doi: 10.1017/ATSIP.2012.5 (invited paper)

[5] Hideyoshi Horimai, "360 degree viewable 3D Display "Holo-Table"", Joint Collaborative Team on 3D Video Coding Extension Development of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11 (MPEG), JCT3V-C0234, M28301, January 2013.

[6] J. Arai, F. Okano, M. Kawakita, M. Okui, Y. Haino, M. Yoshimura, M. Furuya, and M. Sato, "Integral Three-Dimensional Television Using a 33-Megapixel Im-aging System," JOURNAL OF DISPLAY TECHNOLOGY, VOL. 6, NO. 10, pp.422-430, OCTOBER 2010.

[7] Paulo Nunes, Luís Ducla Soares, Amar Aggoun, "3D Holoscopic Video Test Material", Joint Collaborative Team on 3D Video Coding Extension Development of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11 (MPEG), JCT3V-C0210, M28112, January 2013.

[8] Akio Ishikawa, Sei Naito, Shigeyuki Sakazawa, "Walk-Through and Fly-Through Experience using 3D model plus texture mapping toward Free Viewpoint TV", ISO/IEC JTC1/SC29/WG11, M14952, October 2007.

[9] H. Sankoh, A. Ishikawa, S. Naito, and S. Sakazawa, "Robust Background Sub-traction Method based on 3d Model Projections with Likelihood," In 12th Interna-tional Workshop on Multimedia Signal Processing (IEEE MMSP 2010), pp. 171-176, 2010.

[10] H. Sankoh, M. Sugano, and S. Naito,"Dynamic camera calibration method for free-viewpoint experience in sport videos," In 20th International Conference on Multimedia (ACM MM 2012), pp. 1125-1128, 2012.