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

ISO/IEC JTC1/SC29/WG11
MPEG2013/N14104
Geneva, Switzerland
October 2013

Source: **Requirements**
Title: **Use Cases and Requirements on Free-viewpoint Television (FTV)**
Editors: **Mehrdad Panahpour Tehrani, Shinya Shimizu, Gauthier Lafruit, Takanori Senoh, Toshiaki Fujii, Anthony Vetro, Masayuki Tanimoto**
Status: **Approved**

1 Introduction

This document describes the use cases and requirements of the third phase of the FTV (Free-viewpoint Television) [1]-[4] initiative that is underway in MPEG. FTV is a framework that allows viewing of a 3D world by freely changing the viewpoint.

MPEG has been engaged in various aspects of FTV standardization since 2001. MVC (Multiview Video Coding) was the first phase of FTV, which enabled the efficient coding of multiple camera views. 3DV (3D Video) is the second phase of FTV, which enables viewing adaptation and display adaptation of multiview displays. Based on recent development of 3D technology, MPEG started the third phase of FTV in August 2013, targeting super multiview and free navigation applications [5][6]. Our vision of this third phase is to establish a new FTV framework that revolutionizes the viewing of 3D scenes. This document is made by referring to the previous documents [7][8].

Super multiview 3D displays are emerging, anticipated as the next generation of auto-stereoscopic display, providing ultra-realistic 3D visualization at acceptable cost. Several multiview displays have been or are currently being developed, such as 3DTV by MERL[9], REI (Ray Emergent Imaging) by NICT [10], Holovizio by Holographica [11], SMV256 by Tokyo University of Agriculture and Technology [12], Light Field 3D Display by Samsung[13], IP (Integral Photography) by NHK [14][15] and 3D VIVANT[16], Projection type IP by Seoul National University display [17], 360 degree LFD by USC [18], Seelinder by Nagoya University and The University of Tokyo [19], Holo Table by Holy Mine [20], fVisiOn by NICT [21], Walk-through by KDDI[22][23], etc

Since these super multiview 3D displays require a huge number of multi-view images to be rendered in real-time, synthesized from a lower number of input camera views (for cost reasons), a new coding standard is essential to realize their services and products in the market [5][6]. During the 105th MPEG, a new AHG on FTV was established [24][25]to address these new use cases, and investigate the standardization of related technologies.

This document provides application scenarios, use cases, and other aspect of requirements for super multiview, and free navigation targeted by third phase of FTV.

2 FTV Structure

The FTV structure is illustrated in Fig. 1. The input is 3D space, and the output is a single view with varying viewpoint, multiple views or super multiview, depending on the particular application. Applications are categorized into types, where each type of application has a distribution format and codec that is appropriate to satisfy the application needs. View synthesis is considered one important function of all FTV applications. Many target use cases can be realized in this structure.

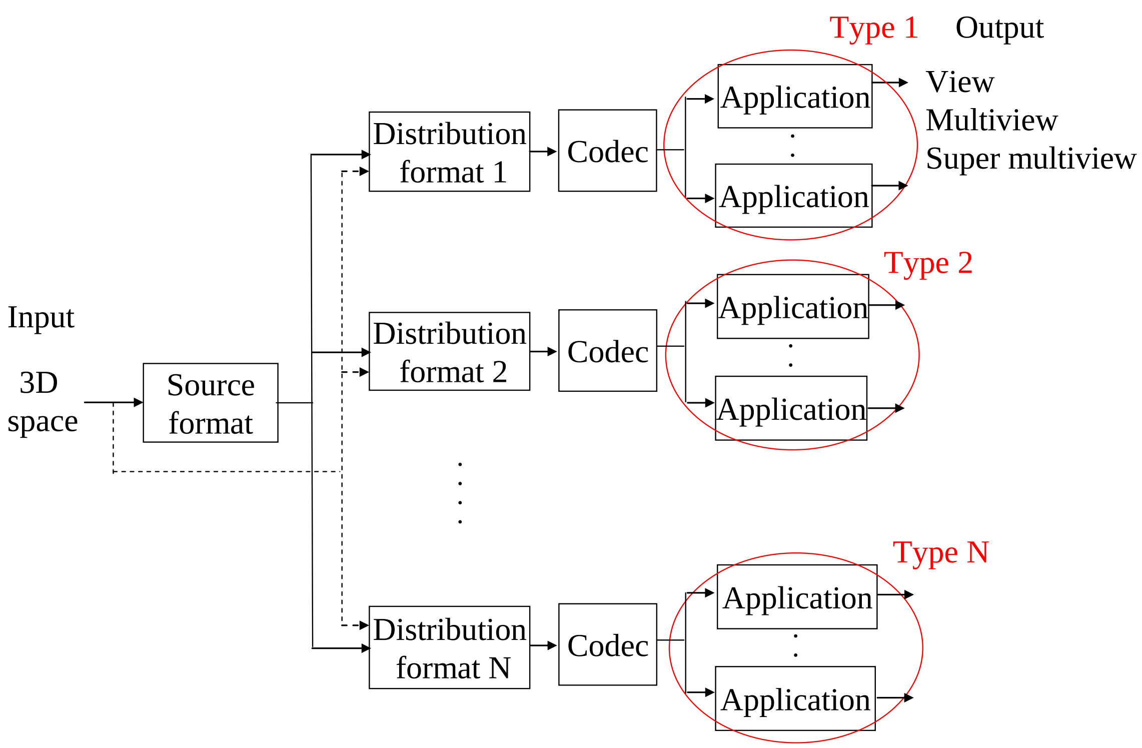

Fig. 1. FTV Structure.

3 Application Scenarios

The third phase of FTV targets three specific application scenarios [26]-[28]:

- Super Multiview: Support for high quality super multiview displays with Horizontal Parallax Only (HPO), where users have a natural 3D depth impression without the need to wear special glasses, i.e. through auto-stereoscopic displays. Moreover, the viewpoint can be freely chosen, positioning oneself in front of the display, as if we would move around the object. The new format enables the generation of many high-quality views from a limited amount of input data.
- Integral Photography: Support for integral photography displays where there is not only horizontal, but also vertical motion parallax. Therefore this application scenario can be considered as a super multiview case. In integral photography, the elemental images representing horizontal and vertical parallax may have a wide

range of resolution (from very low to very high resolution), depending on the application requirements.

- Free Navigation: Support for application scenarios where the user can select the viewpoint and can freely navigate or fly through the scene, along a different pathway than the topology of the input cameras. The display technology can be conventional 2D, stereoscopic, auto-stereoscopic, super multiview or integral photography.

Due to limitations in the production environment (e.g. cost, minimum distance between cameras, etc.), the 3D format is assumed to be based on a limited set of input cameras, placed at the best convenience of the producer (not necessarily 1D linear array). The challenge for FTV ultra-realistic visualization is that – independently of the input and output device architectures – all captured and rendered rays (the intensity and direction of all rays in space, i.e. the Ray-Space) should be identical/symmetrical between input and output views, even with a large ratio of output vs. input views, and/or a completely different topology of camera vs. display devices.

Three types of application scenarios are illustrated in Fig 2.

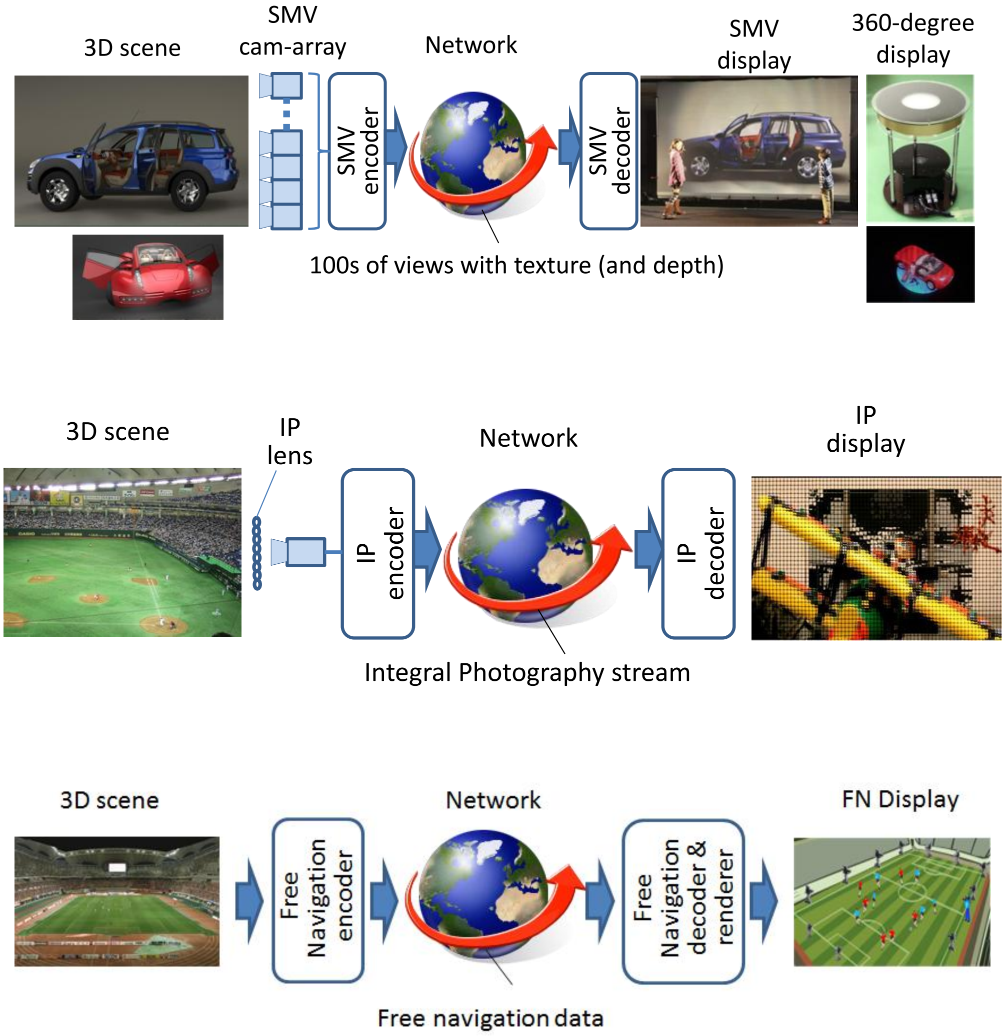

Fig. 2. Three types of application scenarios

Several prototypes of semi-commercial super multiview displays have been recently developed [9]-[13],[17]-[21]. These displays can be categorized into two types which are front view displays and all around displays. Both types of displays only provide horizontal parallax. Source format for super multiview with front view display can be 1D-parallel and convergent multiview, whereas circular multiview topologies are used for all-around displays. The density of the input views in such use cases is considered nearly sparse. To support the super multiview display the number of views shall be increased at the display side.

Integral photography [14]-[16] is recently emerging widely and its commercial broadcast is anticipated in the near future. Integral photography not only has horizontal, but also vertical parallax. Spatial and angular information from the light rays is captured in so-called elemental images (images underneath the lenticular or micro-lenses in front of the imager). Currently the resolution of elemental images is small. However, it is predicted that the resolution will increase in commercial applications. Therefore, the original source format for Integral Photography is a 2D array of multiview images that are easily transformed into the Ray-Space common format. To support the integral photography display the number of views shall be increased at display side.

Free navigation has already been demonstrated in sport applications [22][23], where the user experience approaches the feeling of being present in the field. The input cameras are placed around the field, in a set of approximate 1D linear arrays and/or cylindrical arrays. At the display side one or more views are generated for rendering (mono, stereo or auto-stereoscopic).

4 Use Cases

In order to further understand use cases for super multiview and free navigation, below we summarize the use cases based on the different type of user interests.

4.1 Public Use Cases

Such a use case addresses the events that have general user interest.

1. **Virtual Stadium:** In such a scenario, an empty stadium at the other side of the world can be used to experience the same match, by projecting the contents on a large screen with the same size as the stadium. The displayed SMV 3D image is as large as the real size. It has not only binocular disparity for depth sensation, but it has also very wide motion parallax, which has not been realized by the conventional auto-stereoscopic 3D displays. Wide motion parallax realizes deep immersive sensation. The SMV is a truly ultra-realistic 3D image. The audiences will feel just as if they were standing in the Olympic game stadium.

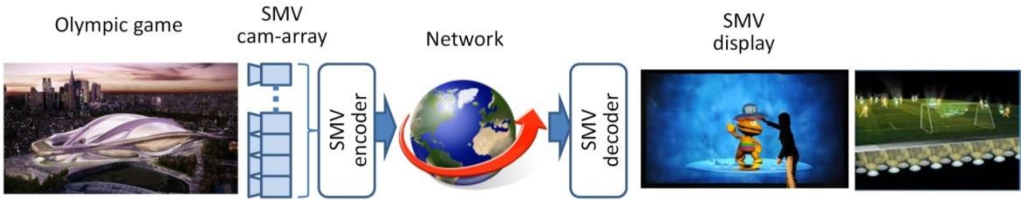

Fig. 3. SMV use case for Olympic scene

2. **Theme Park:** Theme park can be the very initial way to experience this technology. Since the audience in such theme parks are children and adolescents, computer generated contents can be easily displayed and practiced by the super multiview 3D display technology.
3. **Future Cinema:** It is fairly expected that in the near future high quality camera array can be used to shoot the scene. Given the current 3D display technologies, we expect that future cinema would give different experiences to audiences, as if they would have observed during the shooting.
4. **Exhibitions in public areas**: Several contents such (Cultural and Industrial contents) can be provided for commercial purposes, and shown in public areas such as airports and city centers.
5. **Virtually Joint Restaurants, Cafés or Bars:** People in a restaurant/café/bar can join people at the other side of the world while they are having a good time in their local restaurant. The wall of the restaurant, café or bar is the 3D display that can show life size objects from the other side.
6. **Science Signage:** This application is emerging recently, and we can see displays that are located in public areas, transferring common knowledge or virtually guide/educate the pedestrians on their way. The content can be 3D and even can project different contents at the same time, but from different angle, depending on the position of the pedestrian.

4.2 Private Use Cases

In such a use case, the display is provided for one or a limited number of users.

1. **Games**: A new experience can be achieved while the users can see the objects in their real size. In this application, the display shall be large enough to project the real size object. Gamer(s) interaction requires physical movement since gamer(s) is virtually inside the world that is created by the game.
2. **Virtual Class Room:** A student in a remote location can place him/her-self in the classroom as he/she wishes. The teacher can see the students in different places according to their location they chose to see the lecture, and with natural eye contact.
3. **Tele-videoconference:** It can replace current telecommunication means, such as telephone, 2D video teleconference, internet based 2D video teleconference. In such a use case, the participant at the other side can be seen by a large display in real size, with natural eye contact.
4. **Future Commercialized 3DTV:** Displays that are used for private use cases (1~3) can be also be our future TV that shows movies, concerts, TV shows, internet contents, etc.
5. **Automated Driving:** This technology can be used to enable the cars to drive without driver, when it is needed, or being remotely derived to a destination. The driver can remotely control/modify the action of the car from a super multiview display at the office/home.

Fig. 4. Image of a private use case (Tele videoconference)

4.3 Professional/Authority Use Cases

Authorities use this technology to provide higher / safer quality of life.

1. **Monitoring:** Organizations such as IOC, FIFA, FIBA, IBAF, can use this technology to monitor the live contents and provide fair judgments. By in-cooperating such a technology and having a HMD (Head Mounted Display) with the referee in the field, he can freely navigate and see the scene from his desired viewpoint and make a fair judgment.
2. **Video Contents Editing:** Content provider such as Disney, Universal Studios, TV / Network broadcasters, can use this technology and edit the content of animations, movies, and TV contents with a larger degree of freedom.
3. **Industry Inspections**: Currently automated inspection of products using image processing technology is used in industries. By using super multiview 3D display technology, better QC (quality control) can be expected by the manufactures.
4. **Security and Low Enforcement:** Tracking the suspects/criminal and monitoring the traffic, etc would be more efficient.
5. **Product Design:** The same object can be seen from several sides and angles and experts or professionals can virtually interact with the future product and optimize the design.

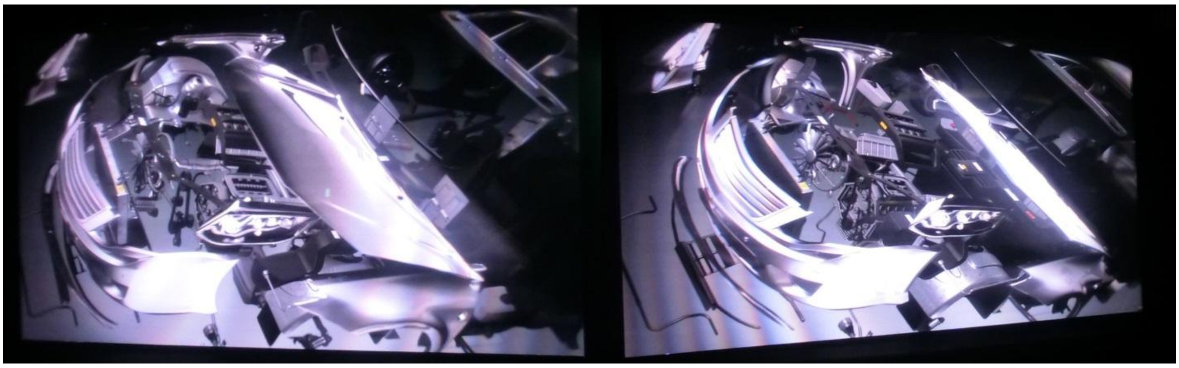

Fig. 5. Image of a professional use case (product design and manufacturing)

4.4 Scientific Use Cases

Such use case can facilitate the progress in different fields of science.

1. **Astronomy**: Such a technology can be used to monitor astronomical changes with realistic viewpoints by scientist in the field. A virtual universe can be created and users can navigate in the universe.
2. **GIS (Geographic Information System):** For example, natural disasters, e.g. hurricane, typhoon, tornado and earthquake, can be predicted more precisely by observing changes in details, in 3D and from different viewpoints.
3. **Medicine:** Currently remote surgery is nearly possible, but the viewing device is HMD and if the doctor is not trained for using such a system, his valuable experience cannot save a life of a person at another place over the world. A high quality 3D display, with HCI would be more natural and can be used for such an application.

Fig. 6. Images of scientific use cases, (expansion of universe, medicine, rain radar)

5 Requirements on FTV

5.1 Application requirements

5.1.1 Application specs

Table 1 categorizes the application scenarios of section 3 into subcategories, according to their technical specifications (the numbers should be interpreted in order of magnitude, not as an absolute number).

Desktop Auto-Stereoscopic Displays with their 2 to 3 input views is the reference point, fully supported by 3DV.

Application groups (I) to (IV) expose new, challenging technical specifications, in particular the dense input and/or output views with different from 1D linear camera topologies.

The Handheld Camera Free Navigation (I) [23] acquiring nearby input views (e.g. a couple of tens at least) along a non-linear and unpredictable pathway, creates a similar number of new views along a user-specified direction. Image registration cannot be performed once at setup time.

Sports Free Navigation (II), e.g. in soccer games, synthesizes new views from up to tens of cameras placed at fixed (not necessarily linear) positions around the terrain.

Super-Multiview Horizontal Parallax Only (HPO) Displays (III) render hundreds of views from tens of input views.

Integral Photography/Videography (IV) supports Full Parallax using lens arrays in front of the camera and display, effectively capturing and rendering the light ray directions in the scene.

Application categories (I), (III) and (IV) convey sufficient angular information (i.e. many inputs from different viewpoints) to enable efficient Ray-Space data representations with good View Synthesis quality if properly fine-tuned.

This Ray-Space Source Data Format also supports device-independent Integral Photography (IV), where input and output devices do not need to have the same lens array specifications.

Table 1: Application categories in FTV third phase

	Input			Output		Camera Calibration Registration		View Synthesis on Source Format		Distribution format
Application	Sparse in	Dense in	Topology in	Sparse out	Dense out	Once Offline	Perframe Online	Depth Map	Ray Space	Codec
3DV = Desktop Auto-Stereoscopic Display	2 or 3		Linear	< 10		X		X		3DV Compatible
I) Handheld Camera Free Navigation		> 32	Free		> 32		X	X	X	
II)Sports Free Navigation (fixed cams)	< 10		4xLin.& Cylindrical		> 32	X		X		
III) Super-MultiView Display HPO		> 32	1D Linear		> 100 to 300	X		X	X	
IV) Integral Photography, Full Parallax		> 50x50	2D array		> 50x50	X			X	

5.2 Requirements for Source Data Format

5.2.1 Video data

The uncompressed data format shall support multiple camera input and multiple camera output configurations, with a higher number of output views than input views. The input camera views along pathways different from 1D linear arrays should be supported. The number of input and output views should vary between tens and hundreds of views (application examples given in Table 1) and View Synthesis should be robust against incorrectly acquired/calculated depth maps. Other input and output configurations beyond stereo should also be supported.

5.2.2 Supplementary data

Supplementary data shall be supported in the data format to facilitate high-quality intermediate view generation. Examples of supplementary data include depth maps, or 3D models, reliability/confidence of depth maps, segmentation information, transparency or specular reflection, occlusion data, etc. Supplementary data can be obtained by any means.

5.2.3 Metadata

Metadata shall be supported in the data format. Examples of metadata include extrinsic and intrinsic camera parameters, scene data, such as near and far plane, and others.

5.2.4 Low complexity for editing

The data format should allow for editing with low complexity.

5.2.5 Applicability

The data format shall be applicable for both natural and synthetic scenes.

5.3 Requirements for Compression

5.3.1 Compression efficiency

Compression efficiency should be comparable or better than the state-of-the-art video codding technology such as MV-HEVC or 3D-HEVC,

5.3.2 Synthesis accuracy

The impact of compressing the data format should introduce minimal visual distortion on the visual quality of synthesized views. The compression shall support mechanisms to control overall bitrate with proportional changes in synthesis accuracy.

Increasing the ratio of output/input views and/or the input views baseline should - below a reasonable threshold - introduce minimal distortion on the visual quality of synthesized views.

5.3.3 Compatibility

5.4 Requirements for Rendering

5.4.1 Rendering Quality

The visual distortion over adjacent, synthesized views should be gradual and robust against input noise, illumination variations and slight artifacts in supplementary data. The temporal visual distortion within one view should also remain within limits.

5.4.2 Rendering capability

The data format should support improved rendering capability and quality compared to existing state-of-the-art representations. The rendering range should be adjustable.

5.4.3 Low complexity

The data format shall allow real-time decoding and synthesis of views, required by any new display technology, with computational and memory power available to devices at realizable level.

5.4.4 Display types

The data format shall be display-independent. Various types and sizes of displays, e.g. stereo and auto-stereoscopic, super multiview, integral photography displays, etc of different sizes with different number of views shall be supported. The data format shall be adaptable to the associated display interfaces.

6 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] Masayuki Tanimoto, Takanori Senoh, Sei Naito, Shinya Shimizu, Marek Domański, Anthony Vetro, Marius Preda “Proposal on a New Activity for the Third Phase of FTV”, ISO/IEC JTC1/SC29/WG11/ M30229, July-Aug. 2013.

[6] Mehrdad Panahpour Tehrani, Takanori Senoh, Makoto Okui, Kenji Yamamoto, Naomi Inoue, Toshiaki Fujii, Hiroya Nakamura, “Proposal to Consider A New Work Item and

Its Use Case - REI : An Ultra-multiview 3D Display -”, ISO/IEC JTC1/SC29/WG11/m30022, July-Aug. 2013.

[7] Video and Requirements Group, “Vision on 3D Video,” ISO/IEC JTC1/SC29/WG11 N10357, Lausanne, CH, February 2009. Available online: http://www.chiariglione.org/mpeg/visions/3dv/index.htm

[8] Video and Requirements Group, “Applications and Requirements on 3D Video Coding”, ISO/IEC JTC1/SC29/WG11 N12035, Geneva, CH, March 2011

[9] W. Matusik, H. Pfister, “3D TV: A Scalable System for Real-Time Acquisition, Transmission and Autostereoscopic Display of Dynamic Scenes”, ACM SIGGRAPH, ISSN: 0730-0301, Vol. 23, Issue 3, pp. 814-824, August 2004.

[10]S, Iwasawa, M. Kawakita, and N. Inoue, “REI: an automultiscopic projection display,” Proc. 3DSA 2013, Selected paper 1, Jun (2013).

[11] http://www.holografika.com/

[12]Y. Takaki and N. Nago, "Multi-projection of lenticular displays to construct a 256-view super multi-view display," Opt. Express 18, 8824-8835 (2010).

[13]J. Lee, J. Park, D. Nam, S. Young Choi, D. Park, C. Y. Kim, “Optimal Projector Configuration Design for 300-Mpixel Light-Field 3D Display”, SID Symposium Digest of Technical Papers, Volume 44, Issue 1, pages 400–403, June 2013.

[14] 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 Imaging System,” JOURNAL OF DISPLAY TECHNOLOGY, VOL. 6, NO. 10, pp.422-430, OCT 2010.

[15]Jun Arai, Tomoyuki Mishina, Yuichi Iwadate, “Integral three-dimensional television”, ISO/IEC JTC1/SC29/WG11 M31052, Geneva, CH, Oct. 2013

[16]JCT3V-C0210: 3D Holoscopic Video Test Material, FP7 3D VIVANT Consortium, Jan. 2013.

[17] Hong J, Kim Y, Park SG, Hong JH, Min SW, Lee SD, Lee B., "3D/2D convertible projection-type integral imaging using concave half mirror array.", Opt Express. 2010 Sep 27;18(20):20628-37. doi: 10.1364/OE.18.02062

[18] http://gl.ict.usc.edu/Research/3DDisplay/

[19] T. Yendo, T. Fujii, M. Tanimoto, M. Panahpour Tehrani, “The Seelinder: Cylindrical 3D display viewable from 360 degrees”, Journal of Visual Communication and Image Representation (2010) ,Volume: 21, Issue: 5-6, Publisher: Elsevier Inc., Pages: 586-594

[20]http://www.holymine3d.com/prod/prod03.html

[21] Shunsuke Yoshida, “fVisiOn: glasses-free tabletop 3D display to provide virtual 3D media naturally alongside real media,” Proc. SPIE 8384, Three-Dimensional Imaging, Visualization, and Display 2012, 838411 (1-9), Apr. 25, 2012. http://urv.nict.go.jp/member/shun/fVisiOn/fVisiOn.html

[22]A. Ishikawa, M. Panahpour Tehrani, S. Naito, S. Sakazawa A. Koike, “Free Viewpoint Video Generation for Walk-through Experience Using Image-based Rendering”, Proceeding MM '08 Proceedings of the 16th ACM international conference on Multimedia Pages 1007-1008.

[23]K. Kawamura, H. Sankoh, S. Naito, “Requirements and use cases of free-navigation for fly-through experience,” ISO/IEC JTC1/SC29/WG11 M31052, Geneva, CH, Oct. 2013.

[24]“List of AHGs Established at the 105th Meeting in Vienna, Austria”, ISO/IEC JTC1/SC29/WG11/w13649, July-Aug. 2013.

[25]“AHG on FTV (Free-viewpoint Television)”, ISO/IEC JTC1/SC29/WG11/w13832, July-Aug. 2013.

[26]Mehrdad Panahpour Tehrani, Takanori Senoh, Makoto Okui, Kenji Yamamoto, Naomi Inoue, Toshiaki Fujii, “Use cases and Application Scenarios for Super Multiview Video and Free-Navigation”ISO/IEC JTC1/SC29/WG11 M31052, Geneva, CH, Oct. 2013.

[27]Mehrdad Panahpour Tehrani, Takanori Senoh, Makoto Okui, Kenji Yamamoto, Naomi Inoue, Toshiaki Fujii, “Introduction of Super Multiview Video Systems for Requirement Discussion”, ISO/IEC JTC1/SC29/WG11 M31052, Geneva, CH, Oct. 2013.

[28]Takanori Senoh, Mehrdad Panahpour Tehrani, Makoto Okui, Kenji Yamamoto, Naomi Inoue, “FTV AHG multiple aspects”, ISO/IEC JTC1/SC29/WG11 M31261, Geneva, CH, Oct. 2013.