International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

J.902
(01/2012)

SERIES J: CABLE NETWORKS AND TRANSMISSION OF TELEVISION, SOUND PROGRAMME AND OTHER MULTIMEDIA SIGNALS

Transmission of 3-D TV services

Multilayered data structure for scalable view-range representation

Recommendation ITU-T J.902

ITU-T

Recommendation ITU-T J.902

Multilayered data structure for scalable view-range representation

Summary

Recommendation ITU-T J.902 specifies a data structure within the scope of Recommendation ITU-T J.901, in order to provide a scalable view-range representation. This Recommendation does not specify view generation schemes, but defines the generic data structure to be used to achieve an efficient view generation.

History

Edition	Recommendation	Approval	Study Group
1.0	ITU-T J.902	2012-01-13	9

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at http://www.itu.int/ITU-T/ipr/.

Table of Contents

		Page
1	Scope	1
2	References	1
3	Definitions	1
	3.1 Terms defined in this Recommendation	1
4	Abbreviations and acronyms	1
5	Conventions	2
6	A hypothetical system configuration based on Recommendation ITU-T J.901	2
7	Scalable view-range representation	2
	7.1 Limitation of view generation by a single data set	2
	7.2 Multilayered data structure	2
Appendix I – Walk-through view illustration		4
	I.1 View generation set-up	4
	I.2 Simple inter-view generation	4
	I.3 Walkthrough view generation	5
Appendix II – Examples of data format		8
	II.1 MPEG-3DV	8
	II.2 Multilayer multi-view video plus depth	9
Bibliography		10

Recommendation ITU-T J.902

Multilayered data structure for scalable view-range representation

1 Scope

Free viewpoint television (FTV) is an innovative technology that allows one to view a three-dimensional (3D) world by freely changing the viewpoint. The transmission aspect of FTV is published as Recommendation ITU-T J.901.

[ITU-T J.901] defines the reference system configuration and shows the allocation of depth estimation and the interpolation module in the configuration. Then, [ITU-T J.901] specifies the requirements for the protocols and data format that are needed, in accordance with the configuration.

The most favourable feature of FTV is its ability to offer an audience a selection of viewpoints. However, as a reproducible range of viewpoints can give rise to a trade-off with the amount of data to be transmitted, [ITU-T J.901] demands data scalability as an optional requirement.

This Recommendation specifies the data structure within the scope of [ITU-T J.901], where the data structure enables scalability in the sense of a reproducible view-range and the amount of data. This Recommendation utilizes existing data representation and view generation schemes specified by ITU and other related standardization organizations, as well as future view generation schemes.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T J.901] Recommendation ITU-T J.901 (2008), *Requirements for the free viewpoint television (FTV) video transmission system*.

3 Definitions

3.1 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.1.1 depth: Distance from the capturing camera to a surface of an object in the scene.

3.1.2 view range: The position and direction of a viewpoint in a three-dimensional (3D) scene, where a virtual view can be generated.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

3D Three-Dimensional

FTV Free Viewpoint Television

5 Conventions

None.

6 A hypothetical system configuration based on Recommendation ITU-T J.901

[ITU-T J.901] bases view generation technology using multi-view images with their depth information. This also applies to this Recommendation.

There are several configurations to place the depth search functionality. This Recommendation uses the configuration in Figure 1, which originates from Figure 2 of [ITU-T J.901].

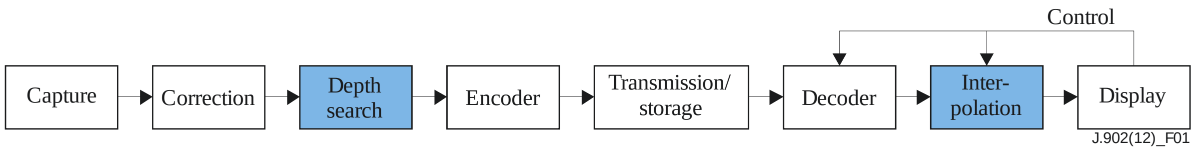

Figure 1 – System configuration based on Recommendation ITU-T J.901

In this configuration, a depth search is performed at the sender side, while interpolation is performed at the receiver side. The sender transmits multi-view images as well as depth information with some additional parameters (ex. camera parameters). The computational load at the receiver side is reduced because the depth information significantly helps the interpolation process.

7 Scalable view-range representation

7.1 Limitation of view generation by a single data set

In this Recommendation the data set is comprised of images, depth information, and additional parameters.

In general, it is possible to generate virtual view images at the viewpoint between video cameras using the data set. The virtual viewpoint can be slightly distant from the line connecting the cameras. However, there are limitations to moving the viewpoint freely due to occlusion and a limited resolution of images.

The former arises with the existence of objects in front of the cameras. The scene behind the object is not captured by the cameras, and cannot be generated.

The latter affects the generated picture quality. When the viewpoint moves forward and gets close to the objects, the generated images are degraded due to excessive image enlargement.

7.2 Multilayered data structure

In the above case, the substantial problem is that the data set does not have sufficient information to generate the scene. One solution is to provide the data set for any position of the scene, however it causes a data size problem.

Figure 2 shows the concept of a multilayered data structure. Here the data is classified into each layer. In the case of occlusion by the object at Layer 1, it can be solved by Layer 2. Hence, the data set has more ability to generate view images using additional layers. That means that such layer structure provides scalability to the data set.

Figure 3 shows another configuration for spatial segmentation. This configuration can describe a wide space in an efficient manner. It allows a video server to transmit only the necessary parts of the whole data when the required view range is limited to a part of the whole range.

A multilayered data structure is specified as follows:

– Each layer has a set of images, depth information, and associated parameters.

– Each layer has an indicator to distinguish it and the reproducible view range information.

It is noted that the image data is not limited to a real camera image, but can include a virtually generated image; depth information describes a kind of 3D object model; associated parameters including optical and geometrical parameters such as the projection matrix of each image.

A reproducible view range tells a view rendering system an area of virtual viewpoint to be displayed, so that the system is able to know which layer data is necessary for rendering the requested viewpoint. It improves the system performance regarding response time and transmission bandwidth, since the system can avoid the excessive and/or deficient transmission of data.

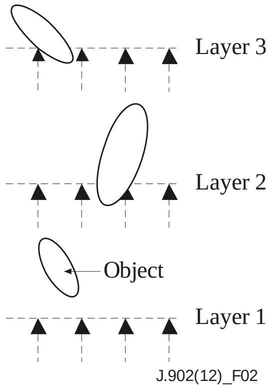

Figure 2 – Conceptual figure of multilayered data structure
(Arrow denotes a camera position.)

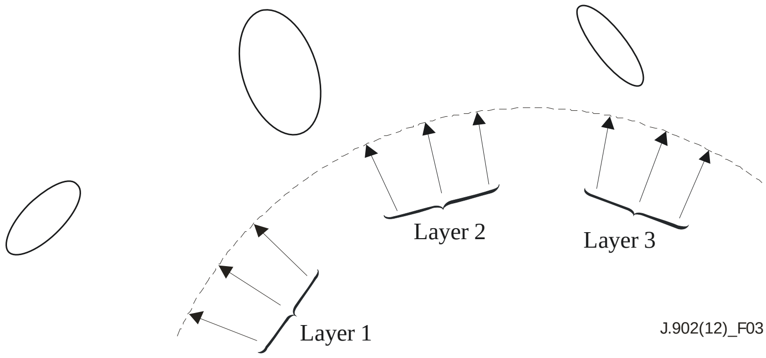

Figure 3 – Another configuration of multilayered data structure

Examples of data formats are presented in the appendices according to each depth-based image generation scheme.

Appendix I

Walk-through view illustration

(This appendix does not form an integral part of this Recommendation.)

This appendix illustrates an example of view generation that can achieve walk-through experience which solves the occlusion problem by utilizing a layered structure.

I.1 View generation set-up

In the simulation, multi-view videos are virtually generated using the multiple local ray-space method in [b-Tehrani]. Depth information is also estimated using the 3-D model information that is available in the multiple local ray-space method.

– Multi-view video plus depth:
  - 1 layer, 180 views per layer (maximum disparity =50 pixels)
– Multilayer multi-view video plus depth:
  - 3 layers, 180 views per layer (maximum disparity =50 pixels)

I.2 Simple inter-view generation

Three examples of free viewpoint generation using single multi-view video plus depth are provided below in Figures I.1, I.2 and I.3.

Figure I.1 – Example 1 of free viewpoint generation using single multi-view video plus depth

Figure I.2 – Example 2 of free viewpoint generation using single multi-view video plus depth

Figure I.3 – Example 3 of free viewpoint generation using single multi-view video plus depth

I.3 Walkthrough view generation

Figure I.4 represents the free viewpoint within the first layer, in the multilayer multi-view video plus depth representation. The result in this layer can also be achieved with single multi-view video plus depth.

Figure I.4 – Example of free viewpoint within the first layer in the multilayer multi-view video plus depth

Figure I.5 represents an example of walk-through view generation in the second layer of multilayer multi-view video plus depth.

Figure I.5 – Example of walk-through view generation in the second layer of multilayer multi-view video plus depth

Figure I.6 represents an example of walk-through view generation in the third layer of multilayer multi-view video plus depth.

Figure I.6 – Example of walk-through view generation in the third layer of multilayer multi-view video plus depth

Appendix II

Examples of data format

(This appendix does not form an integral part of this Recommendation.)

This appendix gives examples of multilayer representation based on existing data formats. Each base data format uses a view image and depth for view generation.

II.1 MPEG-3DV

MPEG-3DV is developing a 3D video format that enables both advanced stereoscopic display processing and improved support for auto-stereoscopic N-view displays. They use multiple video data for input, and generate an arbitrary number of views with the appropriate disparity for 3D presentation.

In this clause, the video format in the MPEG-3DV test sequence is used with spatial segmentation (Figure II.1). Therefore, the basic nature of each data is the same as those in MPEG-3DV[1].

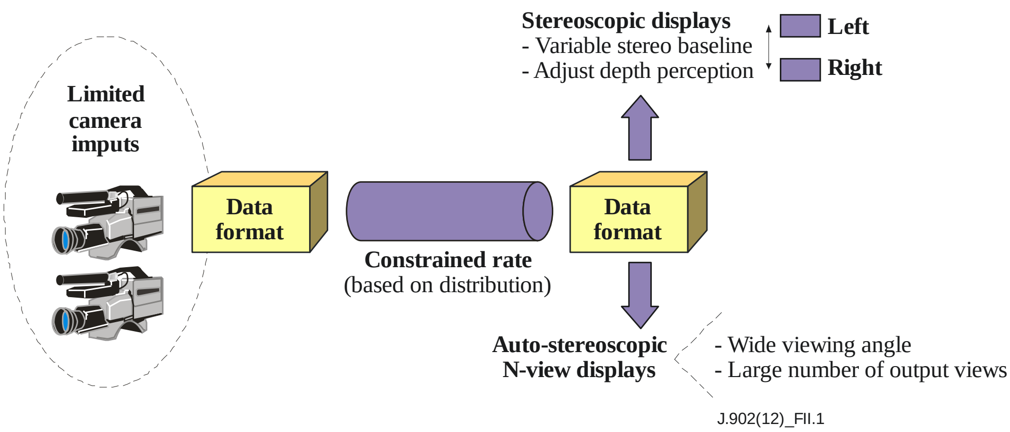

Figure II.1 – Target of the 3D video format [b-ISPA]

II.1.1 Data representation

a) Video, depth and associated information

View location: all view images are on a straight line with the same interval.

View direction: direction of all view images is the same and perpendicular to the image mounted line.

Focal length: common to all view images.

Image resolution: common to all view images, and rectangular shaped.

b) Reproducible view range

View location: on the same line and in between input view images.

View direction: Same as input view images.

Focal length: Same as input view images.

Image resolution: Same as input view images.

[1] It should be noted that the MPEG-3DV discussion is on-going, and the information included in the data format is tentative.

II.2 Multilayer multi-view video plus depth

As described in Appendix I, the multilayer multi-view video plus depth representation has the ability of a walk-through view generation (clause I.3). The key issue of the walk-through view generation is its resolution of the occlusion problem that prevents the rendering of an uncovered region that becomes visible as the virtual viewpoint moves. It is difficult to recognize the occurrence of occlusion until after rendering, which also incurs a computational cost. Therefore, the information of a reproducible view range in the data format is quite effective in simplifying the video delivery process.

II.2.1 Data representation

a) Video, depth and associated information

View location: View images can on arbitrary curve.

View direction: View images have the same direction on the curve.

Focal length: Common to all view images

Image resolution: Common to all view images, and rectangular shape

b) Reproducible view range

View location: View location is not limited on the same curve of the input view images, It is described by the length parameters defined per input view image as depicted in Figure II.2. A front line of view range is defined by straight lines connecting the points which locate ahead position of input images according to the length parameter. Then, the reproducible range is a region behind the front line.

View direction: in between adjacent input view images

Focal length: same as input view images

Image resolution: same as input view images

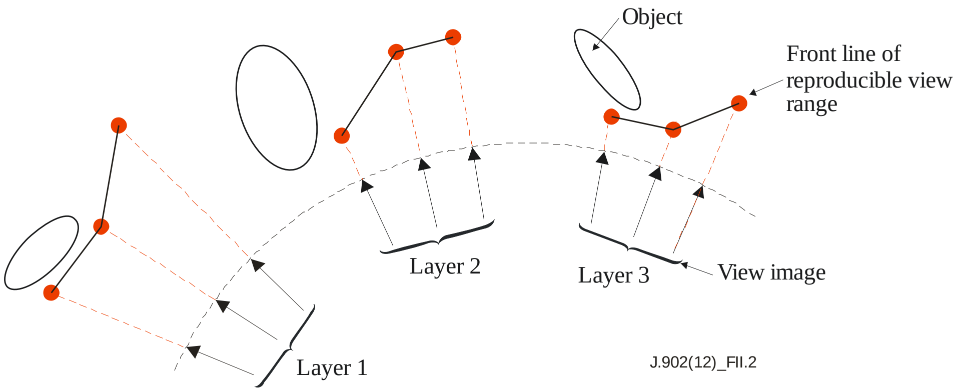

Figure II.2 – View range representation

Bibliography

[b-ISPA] Smolic, A., Mueller, K., Merkle, P., Vetro, A. (2009), *Development of a new MPEG standard for advanced 3D video applications*, Image and Signal Processing and Analysis. Proceedings of 6th International Symposium on Publication Year: 2009 , pp. 400-407.

[b-Tehrani] Tehrani, M., et al., (2008), *Enhanced Multiple Local Ray-spaces Method for Walk-through View Synthesis*, Second International Symposium on Universal Communication, pp.206-209.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	General tariff principles
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	**Cable networks and transmission of television, sound programme and other multimedia signals**
Series K	Protection against interference
Series L	Construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Terminals and subjective and objective assessment methods
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects and next-generation networks
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2012