I n t e r n a t i o n a l   T e l e c o m m u n i c a t i o n   U n i o n

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

J.901
(06/2008)

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

Transmission of 3-D TV services

Requirements for the free viewpoint television (FTV) video transmission system

Recommendation ITU-T J.901

ITU-T

Recommendation ITU-T J.901

Requirements for the free viewpoint television (FTV) video transmission system

Summary

Recommendation ITU-T J.901 shows FTV system model, and then defines requirements for FTV video transmission system.

Source

Recommendation ITU-T J.901 was approved on 13 June 2008 by ITU-T Study Group 9 (2005-2008) under Recommendation ITU-T A.8 procedure.

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/.

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	1
6	FTV System Configuration	2
	6.1 Acquisition	2
	6.2 Correction	2
	6.3 Coding	2
	6.4 Generation	2
7	View generation detail in transmission aspect	3
8	Requirements for FTV transmission system	4
	8.1 Protocol requirements	4
	8.2 Transmission data format requirements	4
Appendix I – Definition of related video services		6
Appendix II – FTV service example		7
	II.1 VOD service for common streaming players	7
	II.2 Streaming FTV service for displays with viewpoint control	7
	II.3 Streaming FTV service for 3D displays without viewpoint control	8
	II.4 Download FTV service for displays with viewpoint control	9
	II.5 Download FTV service for 3D displays without viewpoint control	10

Recommendation ITU-T J.901

Requirements for the free viewpoint television (FTV) video transmission system

1 Scope

FTV is an innovative technology that allows one to view a distant 3D world by freely changing the viewpoint. FTV will open a new era in the history of television since such a function has not yet been achieved by conventional TV technology.

To achieve free navigation functionality, additional requirements are posed other than the conventional video transmission. This Recommendation specifies requirements for transmission of the FTV video under the structure that are also defined in this Recommendation. This Recommendation only defines a system aspect in order to avoid overlap work in the coding and view generation technologies.

NOTE – The structure and content of this Recommendation have been organized for ease of use by those familiar with the original source material; as such, the usual style of ITU-T Recommendations has not been applied.

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 H.264] Recommendation ITU-T H.264 (2007), *Advanced video coding for generic audiovisual services*.

3 Definitions

3.1 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.1.1 depth map: Distance from the capturing camera to a surface of an object in the scene measured per each pixel on the captured image.

3.1.2 FTV: Video media system that can provide audiences of freedom to choice their viewpoint.

4 Abbreviations and acronyms

This Recommendation does not use any particular abbreviations and acronyms.

5 Conventions

This Recommendation does not use any particular notation, style, presentation.

6 FTV System Configuration

The configuration of FTV based on ray space theory is shown in Figure 1. The major processing of FTV is acquisition, correction, coding and generation. Interpolation is a key technology used in compression and generation.

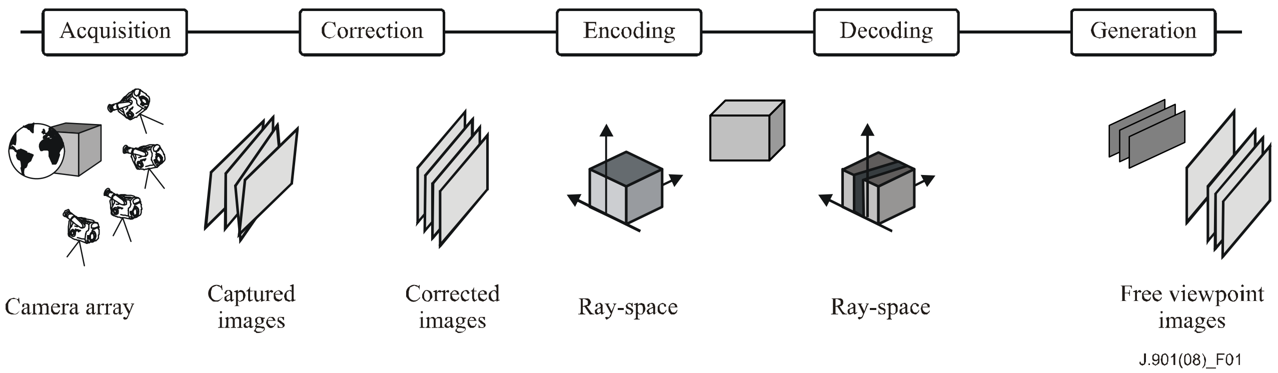

Figure 1 – Configuration of the FTV system

6.1 Acquisition

The FTV signal is acquired by a camera array. There are several types of camera arrangement depending on how freely we want to see the scene. For example, if we want to see the scene from one side only, the cameras are placed on a line. If we want to see the scene from the backside, they are placed on a circle. If we want to see the scene from the top, they are placed on a hemi-spherical dome.

Camera parameters are needed in addition to images for FTV signal.

6.2 Correction

Correction of captured camera images is needed for efficient coding and interpolation. In practical systems, it is difficult to align many cameras with the same performance at the desired places precisely. Therefore, the difference of camera performance and the misalignment of cameras should be corrected. For example, a large improvement of PSNR is achieved in both coding and interpolation by transforming the captured images.

6.3 Coding

The simplest way to compress the FTV signal is to apply the conventional video coding technology to each camera signal independently. However, considering FTV coding is not a simple multi image coding, there is more efficient way to compress the signal by such as multi-view coding (MVC) in H.264 extension.

Not only coding efficiency but also functionality is important in FTV coding depending on the usage models described later. Assume the FTV On Demand. Many users want to see the scene from different viewpoints. In this case, random access by local decoding is required.

In this Recommendation, coding technology itself is out of scope and refers to relevant coding standards.

6.4 Generation

Free viewpoint images are generated according to viewpoint control by an audience using video data acquired by cameras. In this Recommendation, depth map information and camera parameters are utilized for efficient signal processing in the generation stage. In addition, display specifications may also be required in order to achieve best quality by adaptation to the display characteristics such as 2D/3D, the number of views of the display, and the distance of each view. Note that the view generation method is out of scope of this Recommendation.

7 View generation detail in transmission aspect

The view generation module is further decomposed as follows.

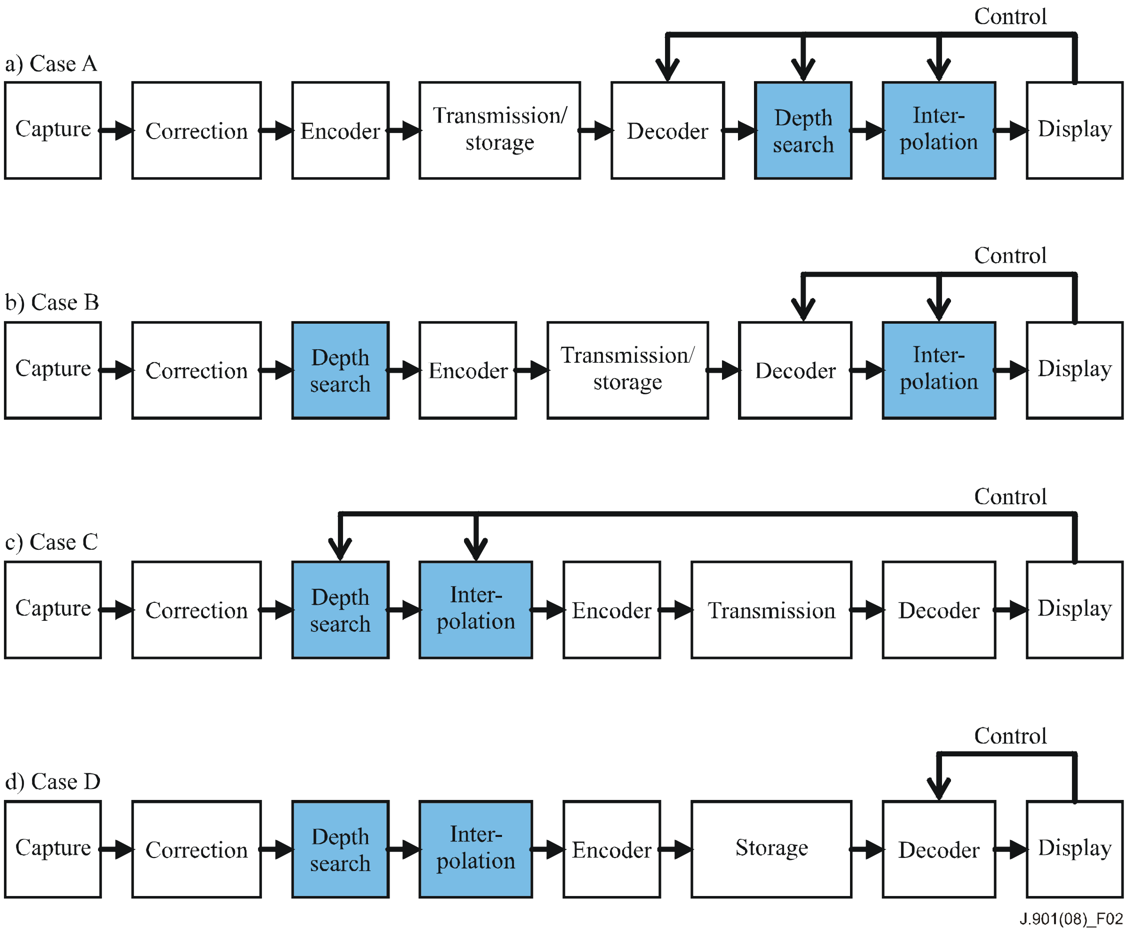

Figure 2 – Position of depth search and interpolation in FTV

The basic assumption here is that view generation is divided into depth search and interpolation. Then the FTV system can be constructed in various ways as shown in Figure 2.

In case A, both depth search and interpolation are performed at the receiver side. Video data is transmitted by MVC, however only the portion of the data is decoded in accordance with the necessity of view generation by depth search and interpolation. In this case, much computation is required at the receiver side, while the viewpoint control will be fast.

In case B, depth search is performed at the sender side for all captured images and interpolation is performed at the receiver side using the minimum depth information. In this case, computational load at the receiver side is reduced, because the depth information helps interpolation process greatly.

In case C, both depth search and interpolation are performed at the sender side. In this case, the receiver only receives the requested view data according to the viewpoint control. However, the response of the view control might be delayed.

In case D, both depth search and interpolation are performed at the sender side, and all interpolated data is stored into a media. Then, the receiver decodes the view data inside the storage. It should be noted that the storage may be at the receiver side as well as at the sender side like a remote disk. In this case, the storage data size is huge.

Cases A and B can apply to both real-time transmission and non-real time operation. Case C assumes only real-time transmission, while Case D is only for stored media.

8 Requirements for FTV transmission system

The standardization items for FTV transmission system are a protocol aspect and a data format for transmission that relates transmission of FTV data. FTV data are defined as view images, camera parameters and depth maps. Though other functional elements in Figure 3 are out of scope of this Recommendation, operations envisioned in the FTV model should also be considered in the transmission requirements.

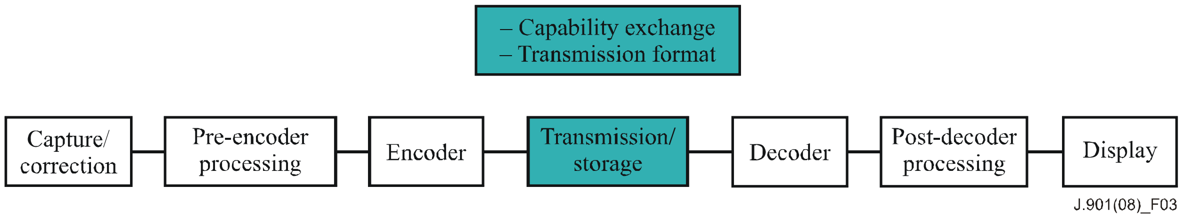

Figure 3 – FTV reference model and standardization items in this Recommendation

8.1 Protocol requirements

Scalability

Various types of scalability, e.g. picture resolution, frame rate, SNR, and range of possible viewpoint, may be supported.

View control

Viewpoint control message should be transmitted, if required. Play back control message such as play, pause and stop may be transmitted.

Audiovisual synchronization

Video and audio representation should be synchronized in time even if the viewpoint is changed.

Capability exchange

Encoder and decoder capability exchange should be performed. In addition, view generation capability and display characteristics may be exchanged.

8.2 Transmission data format requirements

Video data

Video and audio data should be supported in the data format. Video data may be corrected from captured images by real multiple cameras, so that misalignment of camera geometry and colours should be removed.

Depth maps

Depth maps should be supported in the data format. Depth maps may be generated by conversion from captured images by real multiple cameras or through special depth cameras or by other means. Types of depth information should be defined that are able to derive the original object distance.

Camera parameter

Camera parameters, both extrinsic and intrinsic parameters, should be supported in the data format.

Appendix I

Definition of related video services

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

FTV: Free-viewpoint TV is the presentation of the images obtained by multiple camera sources such as those in a ring configuration, enabling the viewer to change viewpoint freely by generating the images at virtual viewpoint from real images.

Stereoscopic TV: Stereoscopic TV is the presentation of the image normally using two sensors spaced at the typical separation of human eyes and displayed in such a way that the images are seen by left and right eye respectively to the position of the sensors.

3D TV: In 3D TV, the images derived from the multiplicity of sensors such as those arranged in planar matrix are displayed in such a way to create pseudo-holographic images in space which enables the viewer to see one side or the other side of the object (such as a head for example) by moving the viewing position from one side to the other of the virtual image in space. Such an image might be some distance in front of the display device.

Appendix II

FTV service example

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

II.1 VOD service for common streaming players

This clause describes one usage of FTV system for video-on-demand services for common streaming players, where a simple video and audio stream decoder can be used as a receiver with an external control unit.

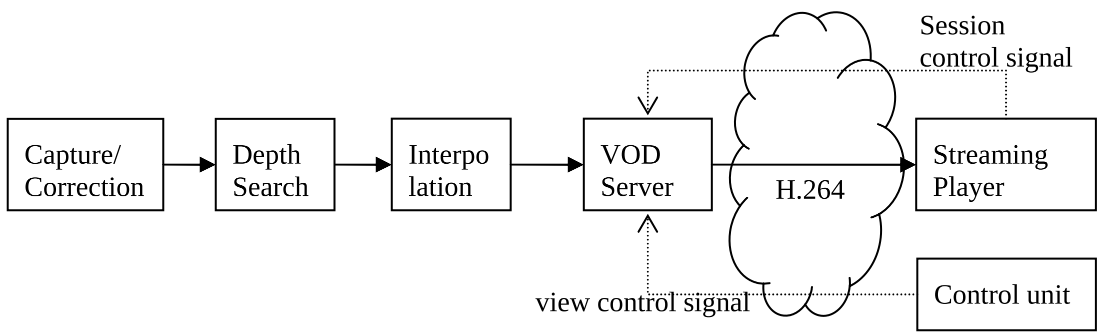

Figure II.1 – VOD service for streaming player with external control unit

The video encoding for the streaming can be H.264 or other coding standards. The streaming control protocol can be RTSP or other protocols. In this case, the player works as a common video streaming player.

For viewpoint control, a special controller is needed outside of the player. The controller will send a viewpoint control message to the VOD server.

The VOD server is able to provide a video stream to the streaming player as well as to change the viewpoint according to the message from the controller. All possible viewpoints can be rendered beforehand. Or the scene generation can be performed in real time according to the request, where VOD server should be coupled with interpolation module. The data format in the VOD server and the view generation method are out of scope of this Recommendation.

One drawback of this approach is operational latency. When an audience operates some controller to change a viewpoint, it takes some time for alteration of the scene, while moving picture from the old viewpoint may be presented for a while.

II.2 Streaming FTV service for displays with viewpoint control

This clause describes one usage of FTV system for streaming FTV service for displays with viewpoint control, where viewpoint-controlled images are displayed on a 2D/3D display.

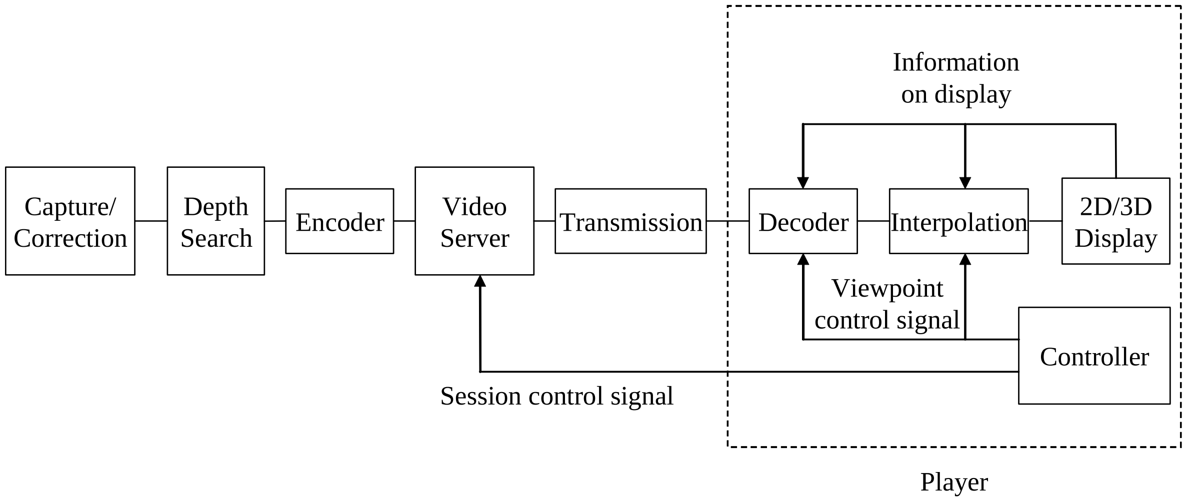

Figure II.2 – Streaming FTV service for displays with viewpoint control

The multi-view video and depth information are encoded and stored at the video server. The encoding can be an extension of MVC. The video server is able to provide the stream of the encoded multi-view video and depth information to the player according to the session control signal from the controller.

The player is able to change the viewpoint according to the message from the controller. The necessary parts of the encoded multi-view video and depth information are decoded. Viewpoint-controlled images are generated by interpolating the decoded multi-view video using the depth information and displayed on a 2D/3D display. The number of generated view images depends on the type of the display.

Although this approach has to send more information to the player than the approach of clause II.1, it can generate new view images immediately after viewpoint change at the player.

II.3 Streaming FTV service for 3D displays without viewpoint control

This clause describes one usage of FTV system for streaming FTV service for 3D displays without viewpoint control, where the users can observe different views according to the positions of their eyes in front of the 3D display.

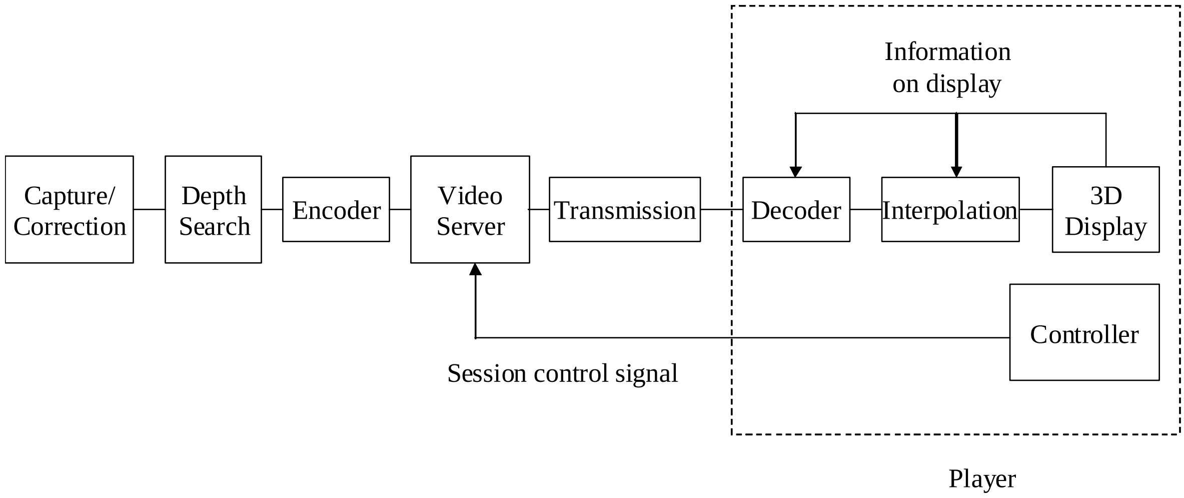

Figure II.3 – Streaming FTV service for 3D displays without viewpoint control

The multi-view video and depth information are encoded and stored at the video server. The encoding can be an extension of MVC. The video server is able to provide the stream of the encoded multi-view video and depth information to the player according to the session control signal from the controller.

The player is able to generate all view images needed by the 3D display. The view images are generated by interpolating the decoded multi-view video using the depth information and displayed on the 3D display. The users can view the 3D scene according to the positions of their eyes in front of the 3D display.

II.4 Download FTV service for displays with viewpoint control

This clause describes one usage of FTV system for download FTV service for displays with viewpoint control, where viewpoint-controlled images are displayed on a 2D/3D display at the player after downloading the encoded multi-view video and depth information.

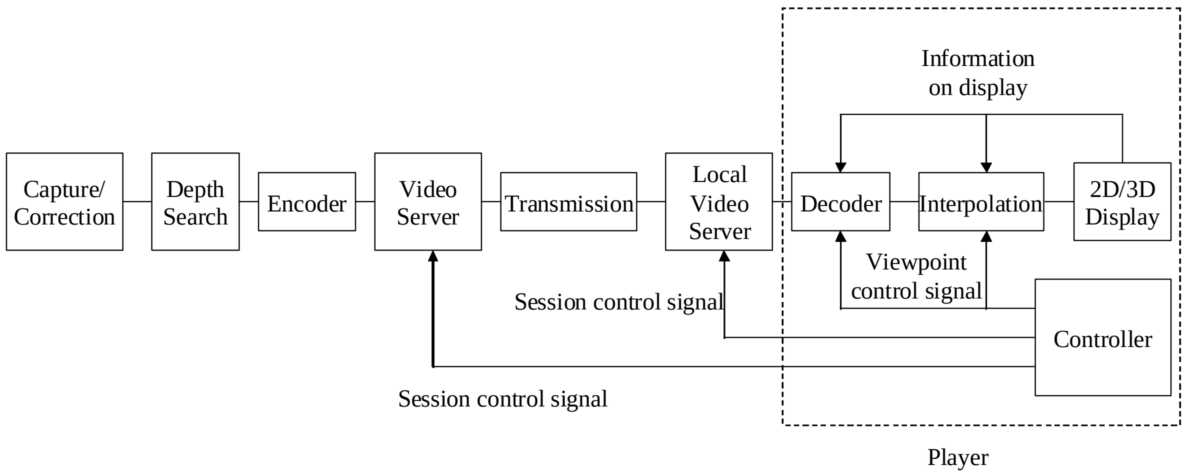

Figure II.4 – Download FTV service for displays with viewpoint control

The multi-view video and depth information are encoded and stored at the video server. The encoding can be an extension of MVC. The encoded multi-view video and depth information are downloaded by the local video server according to the session control signal from the controller.

The local video server is able to provide the encoded multi-view video and depth information to the player according to the session control signal from the controller.

The player changes the viewpoint according to the message from the controller. The necessary parts of the encoded multi-view video and depth information are decoded. Viewpoint-controlled images are generated by interpolating the decoded multi-view video using the depth information and displayed on a 2D/3D display. The number of generated view images depends on the type of the display.

The local video server and the player form a home network. It can be a PC.

II.5 Download FTV service for 3D displays without viewpoint control

This clause describes one usage of FTV system for download FTV service for 3D displays without viewpoint control, where the users can observe different views according to the positions of their eyes in front of the 3D display after downloading the encoded multi-view video and depth information.

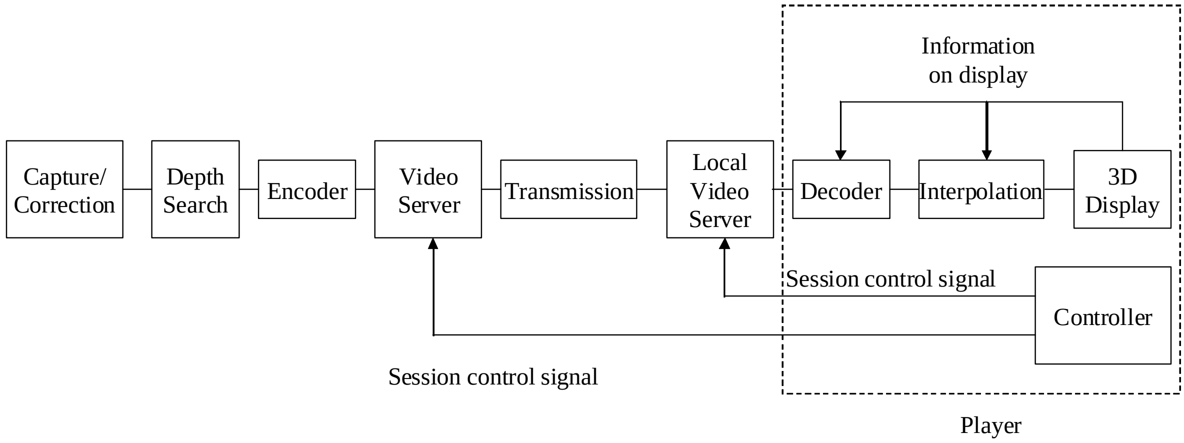

Figure II.5 – Download FTV service for 3D displays without viewpoint control

The multi-view video and depth information are encoded and stored at the video server. The encoding can be an extension of MVC. The encoded multi-view video and depth information are downloaded by the local video server according to the session control signal from the controller.

The local video server is able to provide the encoded multi-view video and depth information to the player according to the session control signal from the controller.

The player is able to generate all view images needed by the 3D display. The view images are generated by interpolating the decoded multi-view video using the depth information and displayed on the 3D display. The users can view the 3D scene according to the positions of their eyes in front of the 3D display.

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	Telephone transmission quality, telephone installations, local line networks
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, 2009