International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

J.301

(10/2014)

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

Application for Interactive Digital Television – Part 2

Requirements for augmented reality smart television systems

Recommendation ITU-T J.301

International Telecommunication Union

Recommendation ITU-T J.301

Requirements for augmented reality smart television systems

Summary

Recommendation ITU-T J.301 specifies requirements that should be considered for augmented reality (AR) smart television system (AR-STV).

AR technology integrates 2D/3D graphics into real-world situations in order to enhance the user experience and enrich information. As smart devices have become more common, AR has been adopted in various applications such as AR-books, AR shopping, AR-SNS, etc. Depending on trends of AR technology, TV broadcasting can be positioned as a more advanced and smart service platform by applying this new technology. Usually, mobile devices such as smartphones and tablets have various sensors to determine the location, direction and speed of the device. However TVs do not have such sensors and cannot provide the same user experience as a smartphone, which can also offer automatic addition of virtual visual components. AR on TV focuses on overlaying visual components onto broadcast content in real time, optionally manipulating components by using a smart device. AR-STV is TV broadcasting using AR technology to blend augmented content (2D/3D graphic objects) with broadcast content in real time on receiving terminals. The augmented content overlaid in the augmentation region is fixed statically or can move along a trajectory similar to what could be provided by an AR-based location application on a smart phone that shows path information on the real location from a smartphone's camera. AR-STV provides that TV viewers can watch the mixed broadcasting contents whenever they want only the service.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T J.301	2014-10-29	9	11.1002/1000/12314

Keywords

Augmented Reality, Augmented Reality Smart Television.

---

* To access the Recommendation, type the URL http://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID. For example, http://handle.itu.int/11.1002/1000/11830-en.

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 not 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

Recommendation ITU-T J.301

Requirements for augmented reality smart television systems

1 Scope

This Recommendation defines functional requirements of metadata, its delivery and synchronization between broadcasting content and augmented content for augmented reality smart television services (AR-STV). This Recommendation also describes the use cases as the background of those requirements.

2 References

None.

3 Definitions

3.1 Terms defined elsewhere

None.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 augmented broadcasting: Broadcasting service or programme to realize augmented reality smart television (AR-STV).

3.2.2 augmented broadcasting metadata: Formatted data describing augmented content.

3.2.3 augmented broadcast provider: A broadcaster who provides augmented broadcasting services, or an entity allowed to develop and provide augmented broadcasting service as a broadcast provider.

3.2.4 augmented content: A binary object, such as 2D images, 3D animated models or audio/video streaming files, to be augmented into a predefined augmentation region.

3.2.5 augmented content provider: An entity allowed to develop and provide augmented content.

3.2.6 augmented reality (AR): A type of mixed reality where graphical elements are integrated into the real world in order to enhance user experience and enrich information.

3.2.7 augmented reality smart television (AR-STV): AR-based TV broadcasting service to implement the mixed content of a broadcast programme and augmented object at a targeted position and time in real time on a receiving terminal such as a TV or set-top box.

3.2.8 augmentation region: The targeted area to augment any object or region in a TV scene. It provides the parameters of a static or animated 2D region related to natural media.

3.2.9 augmentation time: The appointed time to augment any object or region in a TV scene. It provides the parameters of start presentation time and time duration for augmented content.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

AR Augmented Reality

AR-STV Augmented Reality Smart Television

STB Set-Top Box

5 Conventions

In this Recommendation:

The keywords "is required to" indicate a requirement which must be strictly followed and from which no deviation is permitted if conformance to this document is to be claimed.

The keywords "is recommended" indicate a requirement which is recommended but which is not absolutely required. Thus this requirement need not be present to claim conformance.

The keywords "is prohibited from" indicate a requirement which must be strictly followed and from which no deviation is permitted if conformance to this document is to be claimed.

The keywords "can optionally" indicate an optional requirement which is permissible, without implying any sense of being recommended. This term is not intended to imply that the vendor's implementation must provide the option and the feature can be optionally enabled by the network operator/service provider. Rather, it means the vendor may optionally provide the feature and still claim conformance with the specification.

In the body of this Recommendation and its annexes, the words *shall*, *shall not*, *should*, and *may* sometimes appear, in which case they are to be interpreted, respectively, as *is required to*, *is prohibited from*, *is recommended* and *can optionally*. The appearance of such phrases or keywords in an appendix or in material explicitly marked as *informative* are to be interpreted as having no normative intent.

6 Augmented reality smart television (AR-STV) system

6.1 System conceptual model

Figure 1 shows a conceptual model for AR-STV. By this novel TV broadcasting service, the user is provided with additional information alongside the regular broadcast TV programme, and with a more interactive TV experience. The broadcast provider defines an augmentation region so that audio/video content and additional information can be added to that region. The information, formatted as augmented broadcast metadata, is generated by an authoring tool. The specific form of metadata is from MPEG-4 BIFS [b-ISO/IEC 23000-13] with some modification, etc. The metadata includes the unique name, object type, position, presentation/life time and resource location of the augmented objects and contains rendering attributes for mixing augmented objects according to the augmented broadcasting scenario. The service scenario for augmented broadcasting is similar to an acting script for the augmented objects, describing what to augment, when to present, where to locate, how to move and so forth.

Content providers manage an augmented content database harmonized with the scenario. A broadcast programme is transmitted together with metadata to set-top boxes or smart TVs. When viewers wish to view the augmented broadcast service while watching TV, they can launch/control the service via a TV remote control or an application on a second smart mobile device. Therefore, viewers can watch the mixed broadcast contents on demand and only when they want only the service. The system will still provide the original broadcasting content. Moreover, multiple content providers can exist and viewers can be given the choice to select the augmented content from a preferred content provider.

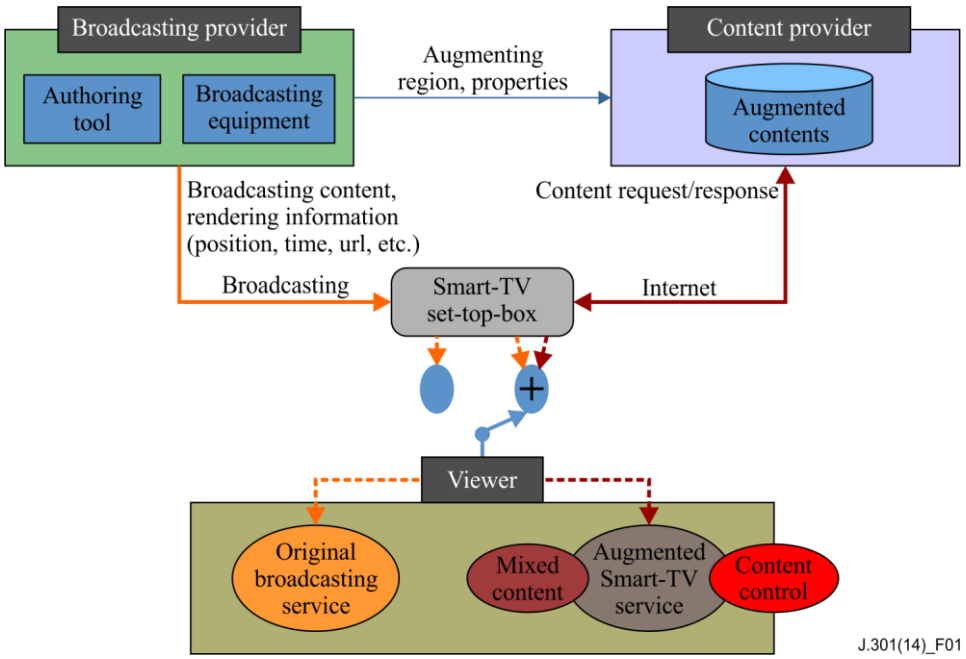

Figure 1 – Conceptual model for AR-STV

6.2 System architecture

The AR-STV system is composed of four essential parts: an authoring tool, the transport multiplexer, a content-providing server and an AR-STV platform, such as a TV or TV set-top box equipped with a 2D/3D-graphics rendering engine. A second smart device on which an AR-STV mobile application is installed is needed to allow interactions between the user and the augmented content. In addition, a broadcasting network is needed to transmit TV programmes and information for the real-time rendering of an augmented content. Also, augmented content will be delivered over a telecommunication network. Figure 2 shows the overall system framework of an AR-STV service.

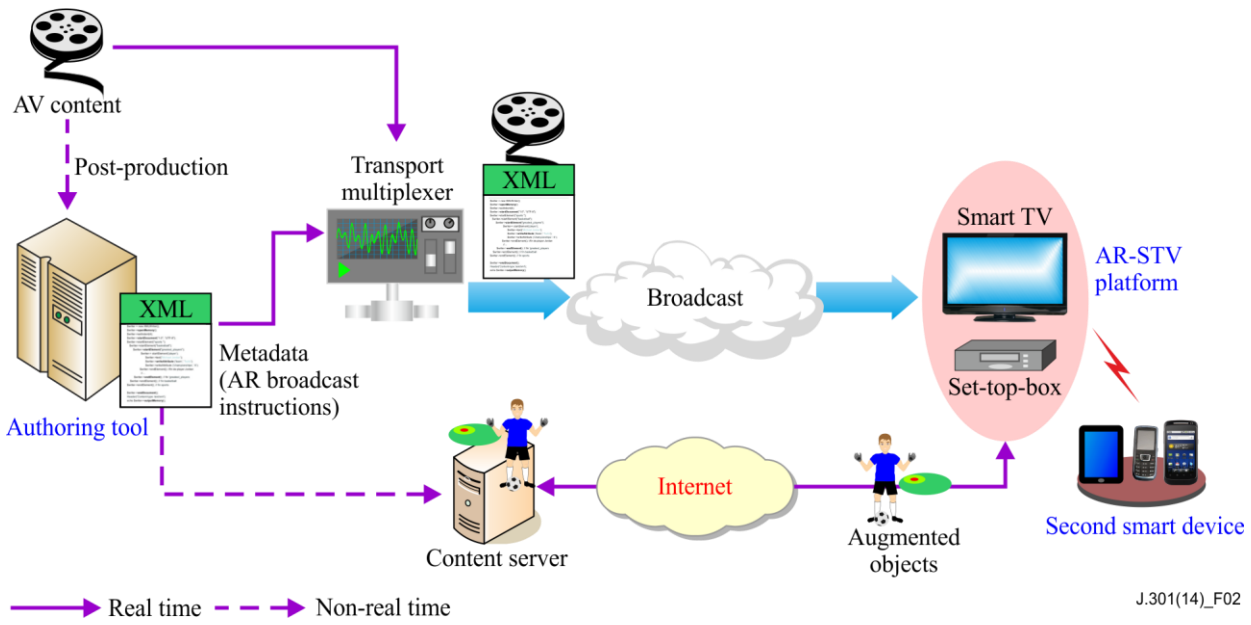

Figure 2 – System architecture of AR-STV

Authoring tool

An authoring tool is used for generating information for real-time rendering of augmented content on a TV or set-top box. The information is generated as metadata, which includes the regions where the

augmented content will be identified and presented; the time schedules for when the augmented content will be downloaded; the uniform resource identifiers (URIs) for downloading the augmented content; the types of augmented content to be presented; and the attributes of the augmented content. This information allows a seamless blending of AR graphic images and video.

The following figure shows the augmentation region, augmentation time, its related attributes and their relationships. The augmentation region defines a targeted area for the augmented object to be overlaid onto background video. The augmentation time defines the appointed time for the presentation of augmented content to start and stop on the TV screen. The augmentation region during the augmentation time can move or rotate, and each event affects the presentation of the augmented content. The rendering environment defines rendering attributes (e.g., lighting position and colours) for naturally matching the augmented content with a background video. The TV viewer's interaction defines is shown to represent a user generated event for interaction, for example, turn on service, drag object, zoom, etc.

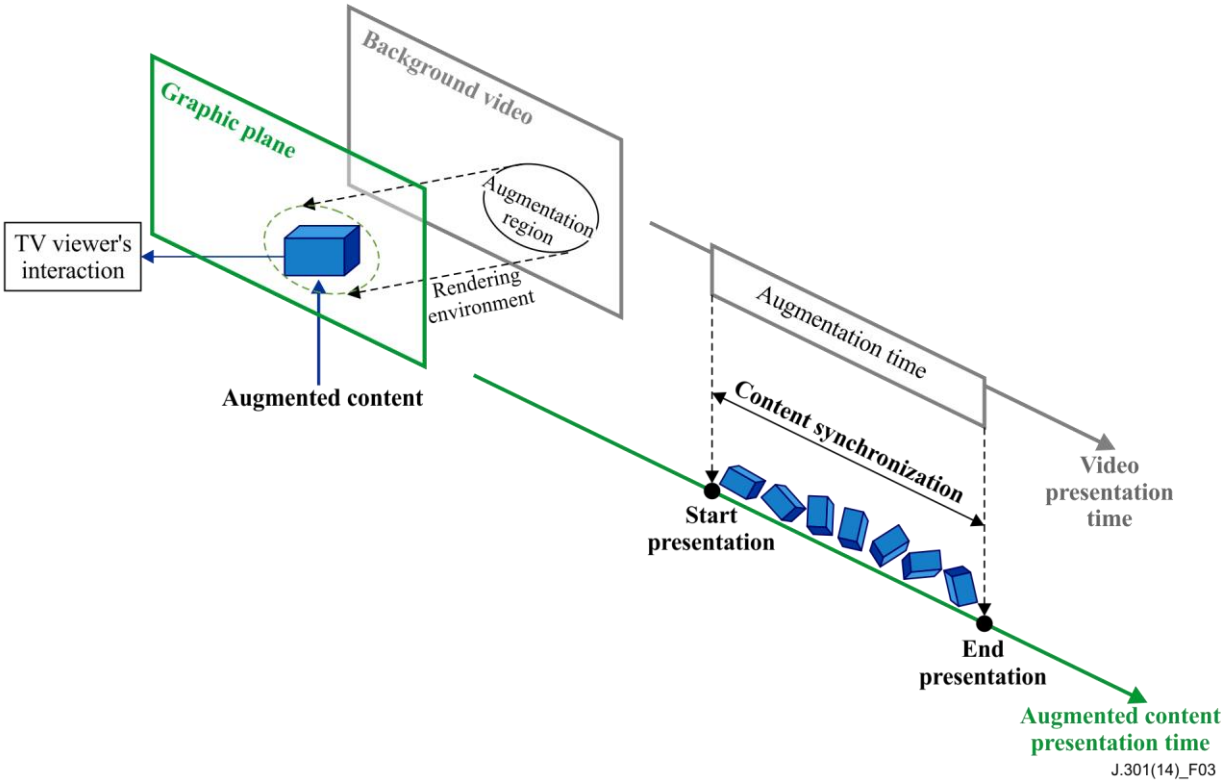

Figure 3 – Definition of augmentation region and augmentation time

The metadata is fragmented into constant-size frames in order to effectively transmit via the broadcast channel. In addition, the reference point of the fragmentation is the time during which the regions or scenes used for presenting augmented content are created, or the time during which the attributes of the augmented content or the regions or scenes, become altered. Each unit of fragmentation, or "instruction", may include all or part of the information used for the presentation of an augmented content. To create a region or scene for presenting an augmented content, a reference instruction is generated, which includes all of the information needed for the presentation of the augmented content. Subsequent instructions may include only information on variations of the region, scene or augmented content. Each time an augmentation region or scene is created, a reference instruction for the region or scene is generated.

Transport multiplexer

In an AR-STV system, a process for rendering augmented content in real-time and blending it into a broadcast video is required. The augmented broadcasting metadata can be multiplexed with a video stream based on an MPEG-2 transport stream. In this process, the packet ID (PID) for the augmented broadcasting metadata needs to be defined, and the programme-specific information (PSI) indicating the relation between the TV programme and the metadata is required to be transmitted through the broadcasting network. In addition, synchronization of the presentation of the augmented content and the video content is required. Therefore, information for synchronization is required to be transmitted within broadcasting programme whenever AR-STV service is enabled.

Augmented content server

Augmented content is made based on TV programme content. Thus, augmented content may be distributed mainly by TV broadcast systems. However, to provide various forms of augmented content, a wide range of content providers would be involved. For this, through the AR-STV system, the chance to deliver various forms of advertisements or additive information to TV viewers could be provided to content providers. The content server is managed by a broadcasting provider or content providers.

In the augmented content server, the various types of augmented content are uploaded by content providers. Thus, the augmented broadcasting metadata can be configured such that only specific augmented contents are selected automatically; or the augmented broadcasting metadata can be configured such that they contain only URL information, and the viewers can select the augmented content themselves.

The broadcast provider, before providing an augmented broadcasting service, gives permission to the augmented contents provider to use an augmentation region. A list of content providers is managed by the broadcast provider, and a mapping table of these providers to the actual server location is interpreted by the AR-STV terminal. Therefore, the AR-STV terminal requests augmented content for the indicated augmentation region from the augmented content provider. The augmented content provider will prepare augmented content for each augmentation region pre-defined by the broadcasting provider.

6.3 The relationship between broadcast provider and content provider

Case A: Use case providing augmented content by broadcasting provider

A broadcast provider can provide both the broadcast programme and augmented content, and service signalling to AR-STV terminal. Current AR broadcasting transmits a one-sided multiplexed stream after blending the visual components, while AR-STV broadcasting separates visual component as augmented content. An AR-STV terminal can provide augmented content over the broadcast programme by the TV viewer's selection. A broadcast provider produces and distributes augmented content and plays the role of a content provider. A broadcast provider could optionally provide the augmented content only for allowed AR-STV terminals as a value-add service model.

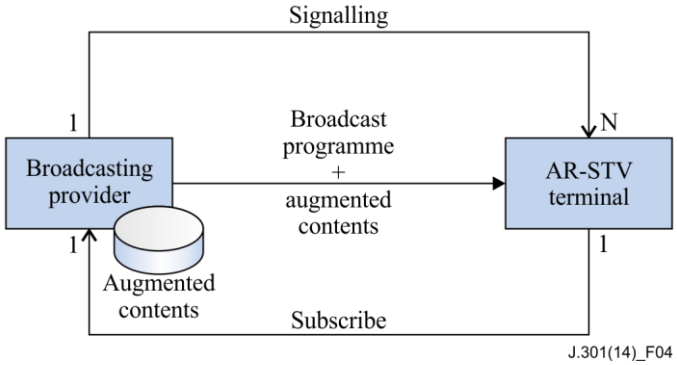

Figure 4 – Broadcasting provider as content provider

Case B: Use case providing augmented content by a content provider

A broadcast provider can make contracts with one or more content providers on business relationships to provide augmented content delivery. Content providers supply a broadcast provider with their URL information in order for AR-STV terminals to access content servers. A broadcast provider can provide the broadcast programme and programme signalling as usual, and content providers provide augmented content to AR-STV terminal respectively. A broadcasting provider generates augmented broadcasting metadata and publishes to content providers for pre-production of augmented content before a scheduled broadcasting programme. The whole or partial video content will be released to content providers. The content providers produce components based on the augmentation region referred to by the metadata and will upload them to a content server with relevant information. Then the broadcast provider is required to provide to AR-STV terminal a linkage between the broadcast programme and the augmented content. An AR-STV terminal can subscribe to one or more content providers and provide a different augmented viewing experience by downloading augmented content from a preferred content provider by a TV viewer.

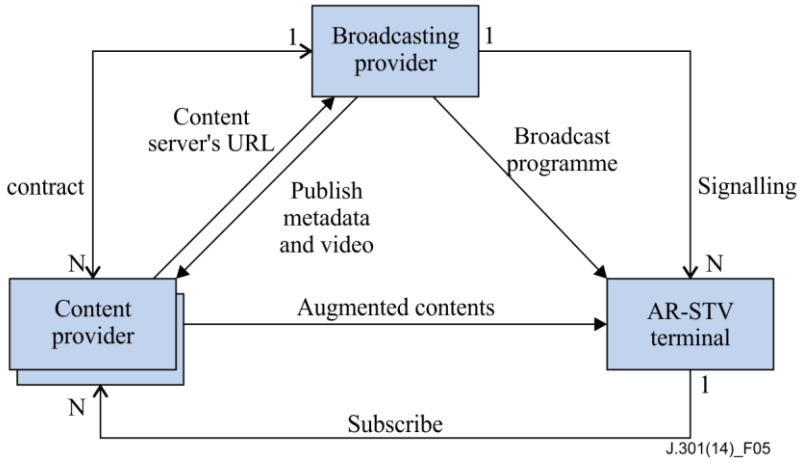

Figure 5 – Broadcasting provider's contract with content providers

7 Functional requirements

7.1 General requirements

[AR_STV-GREQ-1] An AR-STV is required to generate metadata for AR-STV before transmitting the broadcast programme.

[AR_STV-GREQ-2] The AR-STV is required to include the following information in the metadata:

– augmentation region;

– augmentation time;

– download URL of the augmented object.

[AR_STV-GREQ-3] The AR-STV can optionally include GPS information in metadata.

[AR_STV-GREQ-4] The AR-STV is required to present augmented objects as intended in the augmentation region and at the augmentation time.

[AR_STV-GREQ-5] AR-STV is required to have functionality to allow TV viewers to choose whether they watch AR-STV or not (namely, to only watch the original broadcast content).

[AR_STV-GREQ-6] AR-STV is recommended to have functionality to allow TV viewers to manipulate the augmented object through a smart screen device.

[AR_STV-GREQ-7] The TV broadcast programme of AR-STV is required to be compatible with existing broadcasting service. In other words, user can still receive the TV broadcast without augmented content if the user's STB does not support AR-STV.

[AR_STV-GREQ-8] AR-STV is required to enable TV viewers to choose a preferred augmented content provider even though they are in the middle of watching TV.

[AR_STV-GREQ-9] An AR-STV broadcaster is required to inform a receiving terminal of linkage information, including a list of augmented content providers, if the augmented content providers is different from the broadcaster.

[AR_STV-GREQ-10] AR-STV broadcasters/content originators are required to have proprietary rights to generate and publish the metadata.

7.2 System requirements

[AR_STV-SYSREQ-1] The AR-STV receiving terminal is required to be capable of extracting metadata related to the augmented TV broadcast programme from the broadcast channel.

[AR_STV-SYSREQ-2] The AR-STV system is required to transmit AR-STV associated metadata ahead of the related video content to enable synchronization of the TV broadcasting programme and augmented content.

[AR_STV-SYSREQ-3] The AR-STV receiving terminal is required to synchronize broadcast video and augmented content within a tolerance of less than 200 ms.

Appendix I

Scenarios for augmented reality smart television

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

In mobile AR, there are some use cases for example, when a user with an AR-enabled mobile application is trying to find local landmarks. He/she uses a smartphone's camera and walks toward the graphic icon rendered/augmented onto a location map on the smartphone until arriving at the point indicated by the graphic icon. The viewer may want to get information from scenes on the TV screen and expects the augmented objects to be displayed instantly without any intention such as lifting up a camera. In this appendix, some service scenarios for augmented reality smart television are described.

I.1 Augmentation region based application

The augmentation region is a coordinate space in the video image that 2D/3D objects shall be overlaid. The objects shall be synchronized with the movement of the augmentation region. Figure I.1 shows an example of augmentation region based AR. The augmentation region and its tracking information should be described beforehand because a current STB may not have enough power to process tracking from the video stream automatically, in real time. Therefore, it is necessary to prepare the trajectory of the movement and transmit the trajectory together with the video data. Moreover, it would be possible to transmit a prepared mask, depth-information, 3D object or other additional data such as illumination resource for natural composition of 3D objects.

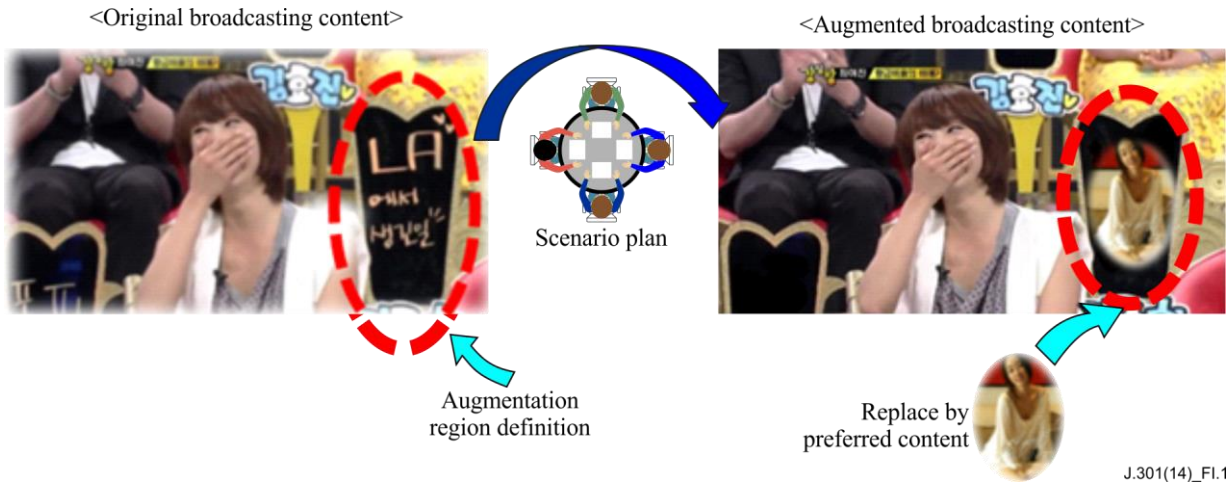

Figure I.1 – Content replacement in augmentation region

I.2 Interactive application with manipulation of the 3D object

Enabling a user to manipulate 3D object augmented over the background video can enrich the user's experience. For example, a student may be watching a TV programme about the organs of the human body in which a teacher explains the functionality of each part of the human heart and a 3D model of a heart can appear. The student could be able manipulate the 3D object by handling a remote controller equipped with motion sensors. Another option could be to manipulate 3D objects by running recognition algorithms for tracking the marker image. This will be possible in the near future if the cost of the set-top-box drops and more computational power is available to the STB.

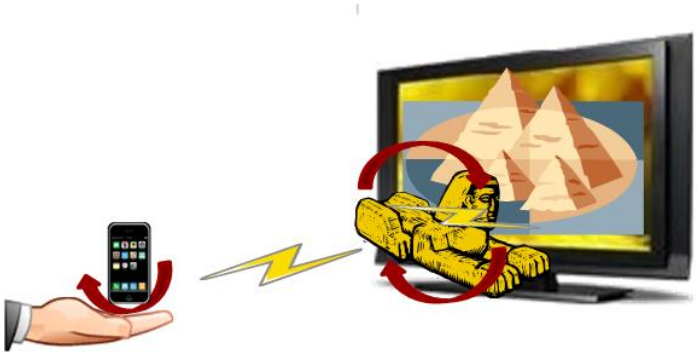

Figure I.2 – Interactive broadcasting with manipulation of the 3D object

I.3 Location based application

Location based AR services are currently very popular in mobile devices. Mobile devices usually have location and orientation sensors. However, TVs do not have such sensors and the video stream is not related with the location of user's viewing point. Then, what if the video stream contains the location and orientation signal obtained from the camera? TV cameras equipped with location and orientation sensors may already be used in AR advertisements, for example in sports game broadcasts. Not only the advertisement, but other varied useful information based on location information provided by the cameras could be possible. For example, in golf game broadcasts with location information, the AR service provider may give the users the ground shape of the field. Another service providers could give is information on the golf club. Similarly, in tourist broadcasts, location information allow users to obtain more useful information such as transportation guidance or restaurants near the location in the scene.

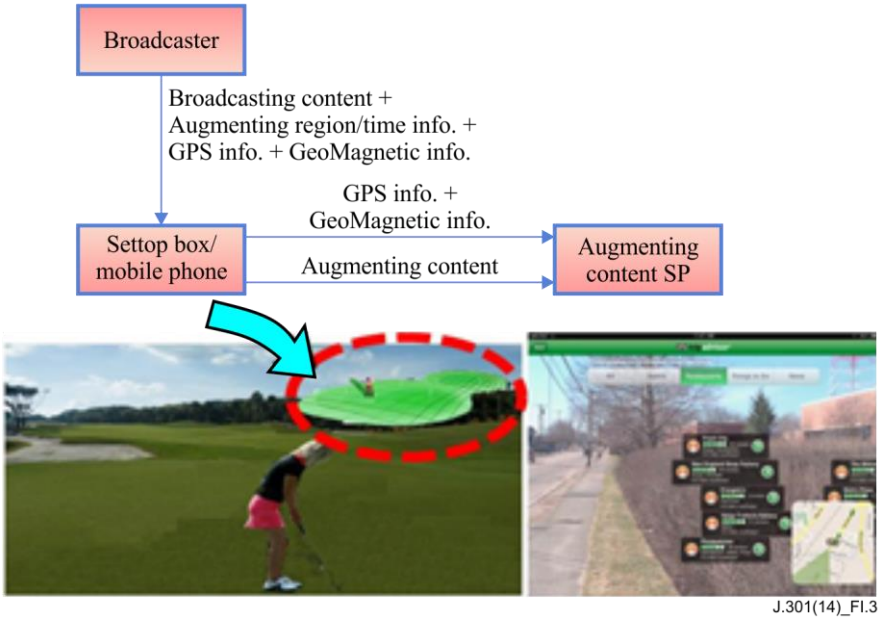

Figure I.3 – Location based broadcasts. Golf or tour broadcasting programme can give viewers the additional information acquired through location based sensors

Appendix II

AR-STV service platform with 2D/3D graphics rendering engine

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

The AR-STV service platform for rendering augmented broadcast content is shown in Figure II.1. The augmented broadcast metadata and broadcast video can be separated using a de-multiplexer. The metadata also will be parsed to download and graphically render augmented content. According to the synchronization information associated with the metadata, each rendered image of augmented content should be blended with its counterpart broadcast video frames immediately. In some use cases, augmented content may contain animated content, and if the animated content is an avatar, for example, its appearance can be presented with the attribute parameters (e.g., emotions, behaviours, or circumstances) included in the metadata.

Augmented content is typically a 2D image or 3D graphic model generated using computer graphics tools. In addition, an augmented content can be generated into various types of formats. Thus, a TV or set-top box should support various graphic application programming interfaces (APIs) to render the various formats of augmented content. Otherwise, it is necessary to generate an augmented content in a standardized way.

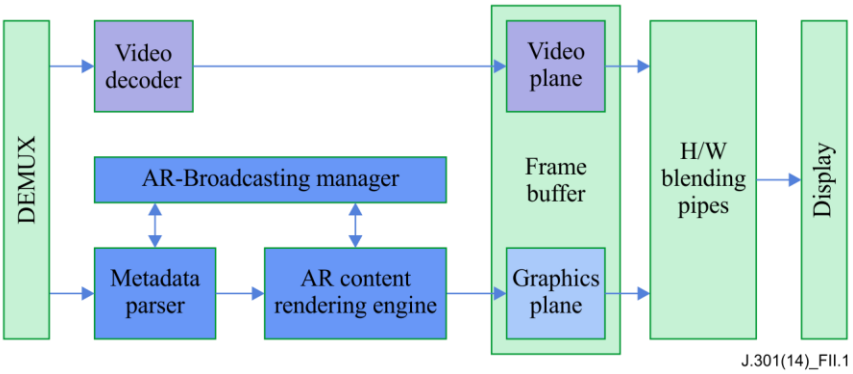

Figure II.1 – The structure of AR-STV service platform

Bibliography

[b-ISO/IEC 23000-13] ISO/IEC 23000-13, *Information Technology – Multimedia Application Format (MPEG-A)* – Part 13: *Augmented Reality Application Format*, May 2014.

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, 2015