INTERNATIONAL TELECOMMUNICATION UNION

TELECOMMUNICATION STANDARDIZATION SECTOR

STUDY PERIOD 2022-2024

SG16-TD249/WP3

STUDY GROUP 16

Original: English

Question(s): 8/16 Rennes, 15-26 April 2024

TD

Source: Editor H.ILE-3DIT

Title: H.ILE-3DIT "Functional requirements and frameworks of 3D model-based immersive telepresence services" (New): Initial draft (Rennes, 15-26 April 2024)

Contact: Hiroshi ANDO
NICT
Japan
E-mail: h-ando@nict.go.jp

Contact: Juan LIU
NICT
Japan
E-mail: juanliu@nict.go.jp

Contact: Hideo IMANAKA
NICT
Japan
E-mail: h.imanaka@nict.go.jp

Contact: Xiaozhi YUAN
China Telecom
China
E-mail: yuanxiaoz@chinatelecom.cn

Abstract: This document is the initial draft of H.ILE-3DIT "Functional requirements and frameworks of 3D model-based immersive telepresence services".

NOTE – Editorial adjustments performed by TSB. Issues needing editor's attention are identified in WinWord comments. Editors are requested to carefully read the Author's Guide for drafting ITU-T Recommendations, https://www.itu.int/oth/T0A0F000004/en.

This document is the output draft for ITU-T H.ILE-3DIT "Functional requirements and frameworks of 3D model-based immersive telepresence services" based on the discussion results of the following contributions:

Doc number	Sources	Proposals	Results
SG16-C419	NICT	Establish new WI	Accepted with some editor's notes.

Draft New Recommendation ITU-T H.ILE-3DIT

Functional requirements and frameworks of 3D model-based immersive telepresence services

Summary

The telepresence scenario described in H.430.3, ILE service scenario, is an ILE service in which remote participants can attend a meeting or an event in real time, where the participants feel as if they were in the same meeting room or event site by using ILE technologies.

This Recommendation describes the functional requirements and framework for 3D model-based immersive telepresence system as one of ILE services, in which a 3D model of each remote participant is constructed from image information and displayed in real time at appropriate locations with appropriate posture in a shared 3D space. This Recommendation also presents candidate technologies to realize 3D model-based immersive telepresence system.

Keywords

ILE; requirements; 3D model-based immersive telepresence services; framework

1 Scope

This Recommendation provides the functional requirements and frameworks of 3D model-based immersive telepresence services (3DIT) as the part of Immersive Live Experience (ILE). This document also provides the candidate technologies.

The scope of this Recommendation includes:

– Overview of 3DIT

– Functional requirements of 3DIT

– Functional framework of 3DIT

– Candidate technologies to realize 3DIT

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.430.1] Recommendation ITU-T H.430.1 (2018), *Requirements for immersive live experience (ILE) services*.

[ITU-T H.430.3~~ (V2)~~] Recommendation ITU-T H.430.3 (2023~~*~~), *Service scenario of immersive live experience (ILE)*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 Immersive Live Experience (ILE) [ITU-T H.430.1]: A shared viewing experience that stimulates emotions within audiences at both the event site and the remote sites, as if the audience at the remote sites had wandered into substantial event venue and had actually watched the events taking place in front of them. This impression is due to high-realistic sensations provided by a combination of multimedia technologies such as sensorial information acquisition, media processing, media transport, media synchronization and media presentation.

3.1.2 TBD

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 TBD: TBD

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

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 Recommendation 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 "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 this Recommendation.

6 Overview

ILE services provide immersive sensations to users by creating high-realistic environments which involve simulating visual, auditory, haptic and other sensory information. The telepresence scenario described in H.430.3, ILE service scenario, is an ILE service in which remote participants can attend a meeting or an event in real time, where the participants feel as if they were in the same meeting room or event site by using ILE technologies.

Commented [TSB1]: Editor: wrong citation style. Read the Author's Guide and correct. Also applicable in other parts of the document.

To construct a photo-realistic 3D model of a participant, conventional methods require the installation of a number of cameras or special 3D sensors around the person, but it is difficult for ordinary participants to prepare such a large facility and space. Therefore, it is desirable to construct a 3D model using only one web camera as used in current online meetings, and to faithfully reproduce non-verbal information such as participants' facial expressions and gestures in 3D to ensure smooth communication as in face-to-face physical meetings.

This Recommendation describes the functional requirements and framework for 3DIT system where a 3D model of each remote participant is constructed from image information and displayed in real time at appropriate locations with appropriate posture in a shared 3D space. This framework for realizing Immersive Live Meeting corresponds to Media/Signalling processing part in an immersive live experience architecture shown in Recommendation H.430.2 (Architectural framework for ILE) (Figure 1).

(Add more "immersive service", "3D modelling", and "telepresence" related information)

Commented [TSB2]: Editor: wrong citation style. Read the Author's Guide and correct. Also applicable in other parts of the document.

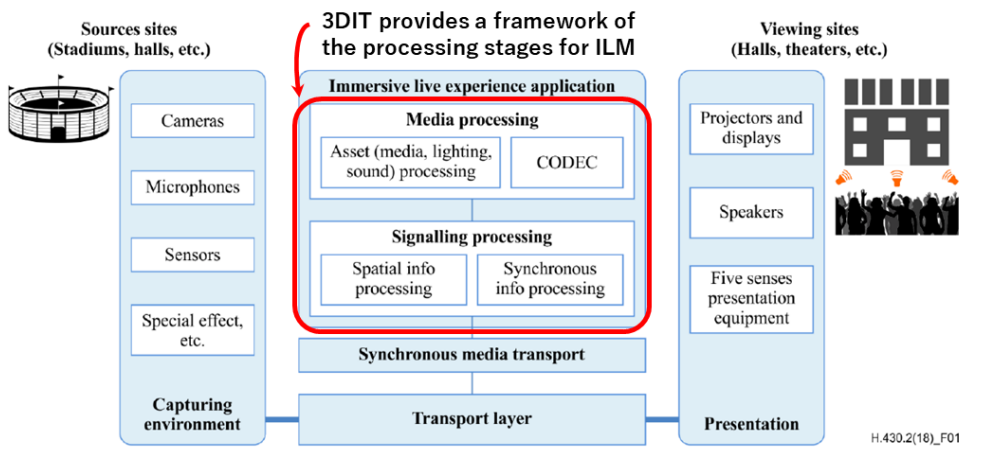

Figure 1: – In the architecture of ILE services, 3DIT provides a framework of the processing stages for Immersive Live Meeting (ILM).

Commented [TSB3]: Editor: The caption should be succinct, not have explanations. Explanations should go in the main text.

Use a en-dash not period after the figure number like I did here; apply elsewhere.

7 Functional requirements of 3DIT

The general and high-level requirements of ILE have been addressed in Recommendation H.430.1 [ITU-T H.430.1]. Regarding the 3DIT related scenarios in H.430.3 [ITU-T H.430.3 ~~(V2)~~], the requirements of 3DIT are presented as follows.

The high-level requirements of IIS, in addition to high-level requirements of ILE, include:

– Support for real time interactive capabilities

– Support for synchronous transmission of concurrent streams

– Distribution of massive multimedia data

– Media processing for immersive interactive information

– Network status awareness and QoE scheduling

[Editor' Note: Add more 3DIT specific functional requirements]

8 Functional frameworks of 3DIT

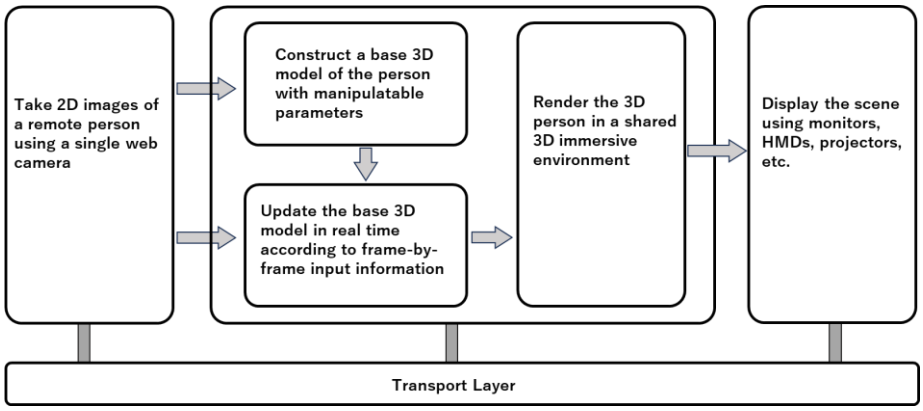

Figure 2. An overview of 3DIT, which constructs a 3D model of each remote participant from camera images, updates the 3D model in real time, and renders the 3D person in a shared 3D environment.

Commented [TSB4]: Editor: keep captions concise. Any explanations should go into the in text where the figure is cross-referenced.

TBD

9 Candidate technologies to realize 3DIT

A more specific framework of 3D model-based immersive telepresence system using a candidate technology (Realistic and EXpressive avataR, REXR) presented in IEEE VR conference 2022 [b-IEEE VR] is depicted in Figure 3.

Each participant in the meeting is sitting at his/her remote location. As in current online meetings, each participant's 2D video is taken by a single web camera and transmitted to other locations.

The system first constructs a base 3D model of each participant. For example, participants turn around in front of a web camera and a set of 2D images are captured from different directions. From these 2D images, the system reconstructs a 3D shape and texture images of the person [b-IEEE VR]. The 3D model contains a number of parameters that allow manipulation of 3D shape, 3D angle of each joint in the body, and facial expressions.

During a meeting, it is important to convey the constantly changing facial expressions and gestures of each participant to other participants in order to establish smooth communication. To achieve this, the system estimates each participant's 3D facial expression (its shape and texture) and 3D body posture (joint angles) online from 2D images. Based on these face and posture information, the base 3D model of the person is updated and the 3D model of the person for each input frame is generated in real time.

The 3D model of each participant is placed in a shared immersive environment to make the participants feel as if they were having a meeting in the same space. The shared immersive environment can be a virtual 3D space or a 3D digital twin environment, as shown in Case 1. Alternatively, as shown in Case 2, a participant could wear an MR (mixed reality) headset to observe

other participants' 3D models as if they were in the real room. In either case, to ensure the meeting is as natural as face-to-face communication, it is necessary to change the posture of 3D models and place them in the appropriate location.

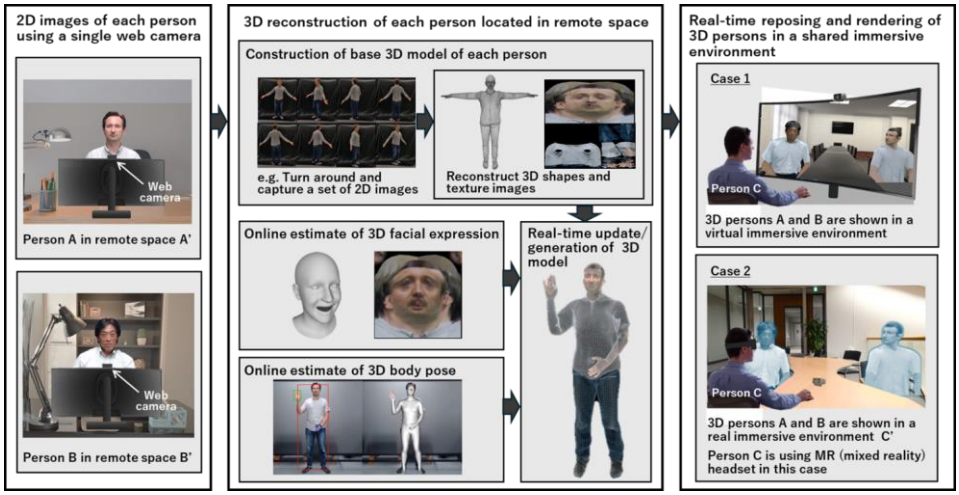

Figure 3. A more specific framework of 3DIT implemented by a candidate technology (REXR from NICT)

TBD.

Annex A

<Annex Title>

(This annex forms an integral part of this Recommendation.)

<Body of annex A>

Appendix I

<Appendix Title>

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

<Body of appendix I>

Bibliography

[IEEEVR] M. Joachimczak, J. Liu, H. Ando: "*Creating 3D Personal Avatars with High Quality Facial Expressions for Telecommunication and Telepresence*", IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) (2022)

_____