3GPP TR 26.998 V0.5.0 (2021-02)

Technical Report

3rd Generation Partnership Project;
Technical Specification Group SA WG4;
Support of 5G Glass-type Augmented Reality / Mixed Reality (AR/MR) devices;
(Release 17)

Keywords
AR, MR, 5G

3GPP

Postal address

3GPP support office address
650 Route des Lucioles - Sophia Antipolis
Valbonne - FRANCE
Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Internet
http://www.3gpp.org

Contents

Foreword

This Technical Report has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

- x the first digit:
  - 1 presented to TSG for information;
  - 2 presented to TSG for approval;
  - 3 or greater indicates TSG approved document under change control.
- y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.
- z the third digit is incremented when editorial only changes have been incorporated in the document.

Introduction

Augmented Reality (AR) and Mixed Reality (MR) are considered as new experiences for immersive media services. It is assumed that the form factors of devices for these services are not different from those of typical glasses, which leaves smaller space for various sensors, circuit boards, antennas, cameras, and batteries than in typical smartphones and therefore reduces the media processing and communication capability that can be supported.

1 Scope

The present document collects information on glass-type AR/MR devices in the context of 5G radio and network services. The primary scope of this Technical Report is the documentation of the following aspects:

- Providing formal definitions for the functional structures of AR glasses, including their capabilities and constraints
- Documenting core use cases for AR services over 5G and defining relevant processing functions and reference architectures
- Identifying media exchange formats and profiles relevant to the core use cases
- Identifying necessary content delivery transport protocols and capability exchange mechanisms, as well as suitable 5G system functionalities (including device, edge, and network) and required QoS (including radio access and core network technologies)
- Identifying key performance indicators and quality of experience factors
- Identifying relevant radio and system parameters (required bitrates, latencies, loss rates, range, etc.) to support the identified AR use cases and the required QoE
- Providing a detailed overall power analysis for media AR related processing and communication

2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document in the same Release as the present document.

[1] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".

[x] 3GPP TR 26.928: “Extended Reality (XR) in 5G”

[4.3.0] ETSI GS ISG ARF 003 v1.1.1 (2020-03): “Augmented Reality Framework (ARF) AR framework architecture”

[4.3.a] 3GPP TR 26.928: “Extended Reality (XR) in 5G”

[4.3.b] 3GPP TS 22.261: “Service requirements for the 5G system”

[4.3.c] 3GPP TR 22.873: “Study on evolution of the IP Multimedia Subsystem (IMS) multimedia telephony service”

[4.3.d] 3GPP TS 26.114: “IP Multimedia Subsystem (IMS); Multimedia telephony; Media handling and interaction”

[4.3.e] 3GPP RP-193241: “New SID on XR Evaluations for NR”

[4.3.f] ISO/IEC 23090-2:2019: “Information technology — Coded representation of immersive media — Part 2: Omnidirectional media format”

[4.3.g] ISO/IEC 23090-3:2020 FDIS: “Information technology — Coded representation of immersive media — Part 3: Versatile video coding”

[4.3.h] ISO/IEC 23090-5:2020 FDIS: “Information technology — Coded representation of immersive media — Part 5: Visual Volumetric Video-based Coding (V3C) and Video-based Point Cloud Compression (V-PCC)”

[4.3.i] ISO/IEC 23090-8:2020 FDIS: “Information technology — Coded representation of immersive media — Part 8: Network based media processing”

[A.1] Google Draco: https://google.github.io/draco/

[A.2] T.Ebner, O.Schreer, I. Feldmann, P.Kauff, T.v.Unger, “m42921 HHI Point cloud dataset of boxing trainer”, MPEG 123rd meeting, Ljubljana, Slovenia

[A.3] Scene understanding, https://docs.microsoft.com/en-us/windows/mixed-reality/scene-understanding

[A.4] Serhan Gül, Dimitri Podborski, Jangwoo Son, Gurdeep Singh Bhullar, Thomas Buchholz, Thomas Schierl, Cornelius Hellge, “Cloud Rendering-based Volumetric Video Streaming System for Mixed Reality Services”, Proceedings of the 11th ACM Multimedia Systems Conference (MMSys'20), June 2020

[A.5] Scene lighting: https://docs.microsoft.com/en-us/azure/remote-rendering/overview/features/lights

[A.6] PBR material: https://docs.microsoft.com/en-us/azure/remote-rendering/overview/features/pbr-materials

[A.7] Color Material: https://docs.microsoft.com/en-us/azure/remote-rendering/overview/features/color-materials

[A.8] S. N. B. Gunkel, H. M. Stokking, M. J. Prins, N. van der Stap, F.B.T. Haar, and O.A. Niamut, 2018, June. Virtual Reality Conferencing: Multi-user immersive VR experiences on the web. *In Proceedings of the 9th ACM Multimedia Systems Conference* (pp. 498-501). ACM.

[A.9] Dijkstra-Soudarissanane, Sylvie, et al. "Multi-sensor capture and network processing for virtual reality conferencing." *Proceedings of the 10th ACM Multimedia Systems Conference.* 2019.

[A.10] VRTogether, a media project funded by the European Commission as part of the H2020 program, https://vrtogether.eu/, November 2020.

[A.11] MPEG131 Press Release: Point Cloud Compression – WG11 (MPEG) promotes a Video-based Point Cloud Compression Technology to the FDIS stage: https://multimediacommunication.blogspot.com/2020/07/mpeg131-press-release-point-cloud.html

…

[x] <doctype> <#>[([up to and including]{yyyy[-mm]|V<a[.b[.c]]>}[onwards])]: "<Title>".

3 Definitions, symbols and abbreviations

Delete from the above heading those words which are not applicable.

Clause numbering depends on applicability and should be renumbered accordingly.

3.1 Definitions

For the purposes of the present document, the terms and definitions given in 3GPP TR 21.905 [1] and the following apply. A term defined in the present document takes precedence over the definition of the same term, if any, in 3GPP TR 21.905 [1].

Definition format (Normal)

<defined term>: *<definition>.*

example: text used to clarify abstract rules by applying them literally.

3.2 Symbols

For the purposes of the present document, the following symbols apply:

Symbol format (EW)

<symbol> <Explanation>

3.3 Abbreviations

For the purposes of the present document, the abbreviations given in 3GPP TR 21.905 [1] and the following apply. An abbreviation defined in the present document takes precedence over the definition of the same abbreviation, if any, in 3GPP TR 21.905 [1].

Abbreviation format (EW)

AR	Augmented Reality
MR	Mixed Reality
XR	Extended reality

4 Introduction to Glass-type AR/MR Devices

Editor's Notes:

<Relevant to Objective #1>

1) Provide formal definitions for the functional structures of AR glasses, classified as device types of XR5G-A4 (standalone) and XR5G-A2, A5 (wirelessly tethered) in TR 26.928, including their capabilities and constraints with respect to communication, computing and graphics processing, tracking, sensors, display, and power consumption

NOTE 1: Device type of XR5G-A3 (video see-through HMD) are not the primary scope of this study, but are not excluded per se.

4.1 General

<Provides relevant definitions of core parts for AR glasses, such as vision, SLAM or localization, object recognition>

4.2 Device Functional Structure

4.2.1 Device Functions

AR glasses contain various functions that are used to support a variety of different AR services as highlight by the different use cases in clause 5.

The various functions that are essential for enabling AR glass-related services within an AR device functional structure include:

a) Tracking and sensing

- Inside-out tracking for 6DoF user position
- Eye Tracking

- Hand Tracking
- Sensors

b) Capturing

- Vision camera: capturing (in addition to tracking and sensing) of the user's surroundings for vision related functions
- Media camera: capturing of scenes or objects for media data generation where required

  NOTE: vision and media camera logical functions may be mapped to the same physical camera, or to separate cameras. Camera devices may also be attached to other device hardware (AR glasses or smartphone), or exist as a separate external device.

- Microphones: capturing of audio sources including environmental audio sources as well as users' voice.

c) Basic AR functions

- 2D media encoders: encoders providing compressed versions of camera visual data, microphone audio data and/or other sensor data.
- 2D media decoders: media decoders to decode visual/audio 2D media to be rendered and presented
- Vision engine: engine which performs processing for AR related localisation, mapping, 6DoF pose generation, object detection etc., i.e. SLAM, objecting tracking, and media data objects. The main purpose of the vision engine is to "register" the device, i.e. the different sets of data from real and virtual world are transformed into the single world coordinate system.
- Pose corrector: function for pose correction that helps stabilise AR media when the user. Typically, this is done by asynchronous time warping (ATW) or late stage reprojection (LSR).

d) AR/MR functions

- Immersive media decoders: media decoders to decode compressed immersive media as inputs to the immersive media renderer. Immersive media decoders include both 2D and 3D visual/audio media decoder functionalities.
- Immersive media encoders: encoders providing compressed versions of visual/audio immersive media data.
- Compositor: compositing layers of images at different levels of depth for presentation
- Immersive media renderer: the generation of one (monoscopic displays) or two (stereoscopic displays) eye buffers from the visual content, typically using GPUs. Rendering operations may be different depending on the rendering pipeline of the media, and may include 2D or 3D visual/audio rendering, as well as pose correction functionalities.
- Immersive media reconstruction: process of capturing the shape and appearance of real objects.
- Semantic perception: process of converting signals captured on the AR glass into semantical concept. Typically uses some sort of AI/ML. Examples include object recognition, object classification, etc.

e) Tethering and network interfaces for AR/MR immersive content delivery

- The AR glasses may be tethered through non-5G connectivity (wired, WiFi)
- The AR glasses may be tethered through 5G connectivity
- The AR glasses may be tethered through different flavours for 5G connectivity
- The requirements for such a connectivity are for further study, in particular for different glass-types. (see for example https://docs.microsoft.com/en-us/azure/remote-rendering/reference/network-requirements)

f) Physical Rendering

- Display: Optical see-through displays allow the user to see the real world "directly" (through a set of optical elements though). AR displays add virtual content by adding additional light on top of the light coming in

from the real-world. Some good reads: https://www.linkedin.com/pulse/why-making-good-ar-displays-so-hard-daniel-wagner/

g) AR/MR Application

- An application that makes use of the AR and MR capabilities to provide a user experience.

4.2.2 Generic reference device functional structure device types

4.2.2.1 Overview

In TR 26.928, different AR and VR device types had been introduced in clause 4.8. This clause provides an update and refinement in particular for AR devices. The focus in this clause mostly on functional components and not on physical implementation of the glass/HMD. Also, in the context the device is viewed as a UE, i.e. which functions are included in the UE.

A summary of the different device types is provided in Table 4.1. The table also covers:

- how the devices are connected to get access to information an
- where the 5G Uu modem is expected to be placed
- where the basic AR functions (as specified in 4.2.1) are placed
- where the AR/MR functions (as specified in 4.2.1) are placed
- where the AR/MR application is running
- where the power supply/battery is placed.

In all glass device types, the sensors, cameras and microphones are on the device.

The definition for Split AR/MR in Table 4.1 is as follows:

- Split: the tethered device or external entity (cloud/edge) does some pre-processing (e.g, a pre-rendering of the viewport based on sensor and pose information), and the AR/MR device and/or tethered device performs a rendering considering the latest sensor information (e.g. applying pose correction). Different degrees of split exist, between different devices and entities. Similarly, vision engine functionalities and other AR/MR functions (such as AR/MR media reconstruction, encoding and decoding) can be subject to split computation.

Table 4.1: 5G Augmented Reality device types

Device Type Name	Reference	Tethering	5G Uu Modem	Basic AR Functions	AR/MR Functions	AR/MR Application	Power Supply
5G Standalone AR UE	1: STAR	N/A	Device	Device	Device/Split [1)]	Device	Device
5G EDGe-Dependent AR UE	2: EDGAR	N/A	Device	Device	Split [1)]	Cloud/Edge	Device
5G WireLess Tethered AR UE	3: WLAR	802.11ad, 5G sidelink, etc.	Tethered device (phone/puck)	Device	Split [2)]	Tethered device	Device
5G Wired Tethered AR UE [3)]	4: WTAR	USB-C	Tethered device (phone/puck)	Tethered device	Split [2)]	Tethered device	Tethered device
1) Cloud/Edge 2) Phone/Puck and/or Cloud/Edge 3) Not considered in this document							

The Wired Tethered STAR Glass device type is for reference purposes only and not considered in this document as it is not included as part of the study item objectives.

Generally, the STAR and WLAR device according to Table 4.1 are expected to have similar functionalities from a 5G System perspective.

Based on this, the focus is on three main different device types in the remainder of this document following the rows 1 to 3 in Table 4.1.

4.2.2.2 Type 1: 5G STandalone AR (STAR) UE

Figure 4.2.1 provides a functional structure for Type 1: 5G STandalone AR (STAR) UE.

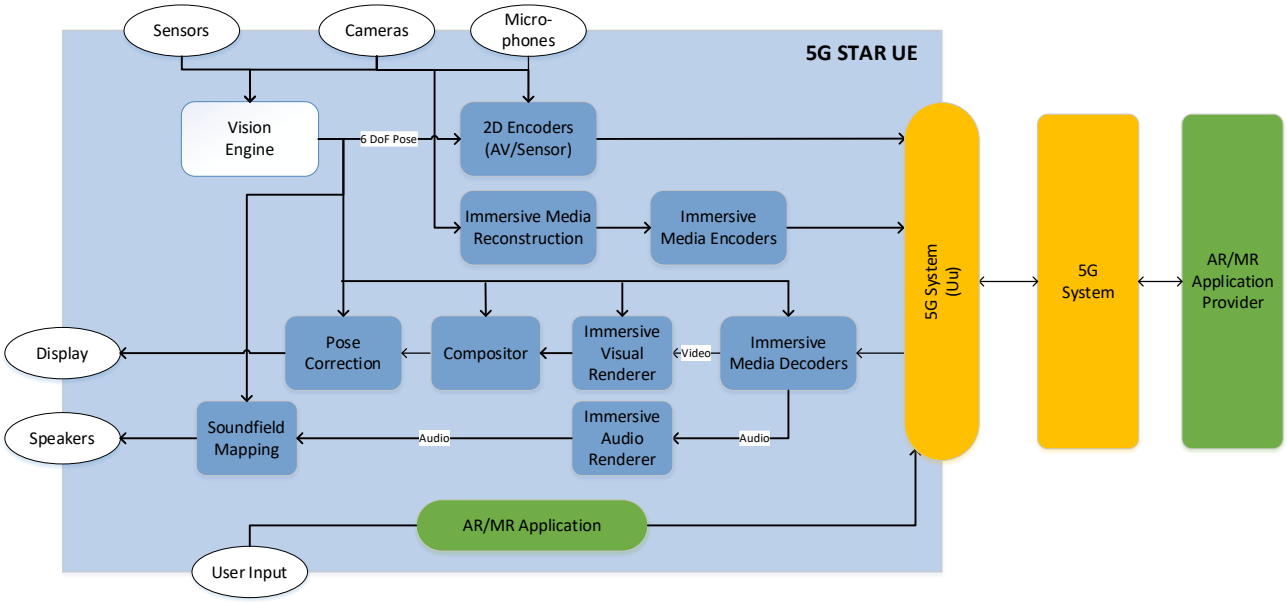


Figure 4.2.1: Functional structure for Type 1: 5G STandalone AR (STAR) UE

Main characteristics of Type 1: 5G STandalone AR (STAR) UE:

- As a standalone device, 5G connectivity is provided through an embedded 5G modem
- User control is local and is obtained from sensors, audio inputs or video inputs
- AR/MR functions are either on the AR/MR device, or split
- The AR/MR application is resident on the device
- Due to the amount of processing required, such devices are likely to require a higher power consumption in comparison to the other device types.
- Functionality is more important than design

4.2.2.3 Type 2: 5G EDGe-Dependent AR (EDGAR) UE

Figure 4.2.2 provides a functional structure for Type 2: 5G EDGe-Dependent AR (EDGAR) UE.

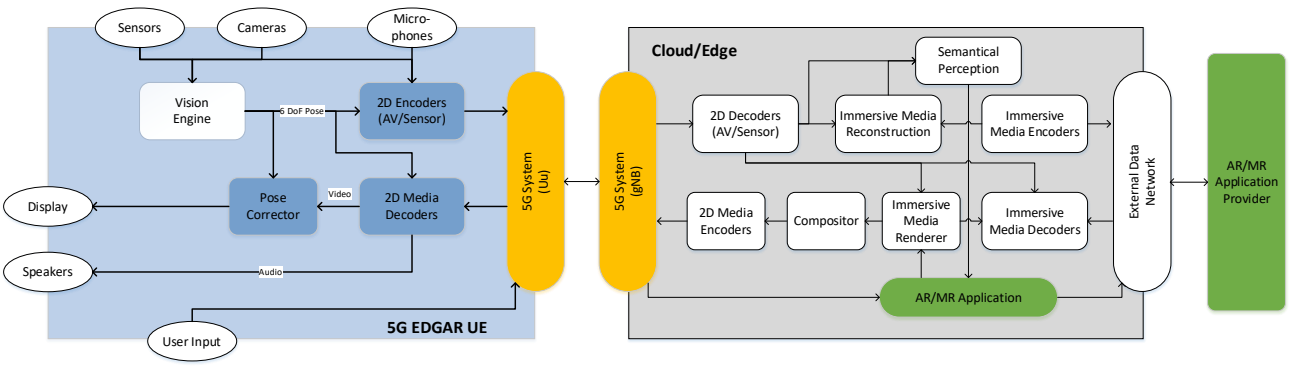


Figure 4.2.2: Functional structure for Type 2

Main characteristics of Type 2: 5G EDGe-Dependent AR (EDGAR) UE:

- As a standalone device, 5G connectivity is provided through an embedded 5G modem
- User control is local and is obtained from sensors, audio inputs or video inputs.
- Media processing is local, the device needs to embed all media codecs required for decoding pre-rendered viewports
- The basic AR Functions are local to the AR/MR device, and the AR/MR functions are on the 5G cloud/edge
- The AR/MR application resides on the cloud/edge.
- Power consumption on such glasses must be low enough to fit the form factors. Heat dissipation is essential.
- Design is typically more important than functionality.

4.2.2.4 Type 3: 5G WireLess Tethered AR UE

Figure 4.2.3 provides a functional structure for Type 3: 5G WireLess Tethered AR UE.

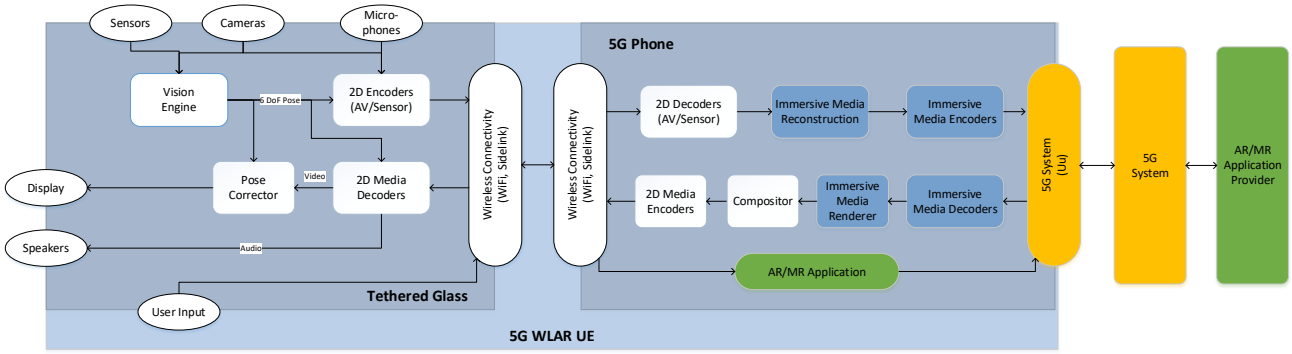


Figure 4.2.3: Functional structure for Type-3: 5G WireLess Tethered AR UE

Main characteristics of Type 3: 5G WireLess Tethered AR UE:

- 5G connectivity is provided through a tethered device which embeds the 5G modem. Wireless tethered connectivity is through WiFi or 5G sidelink. BLE (Bluetooth Low Energy) connectivity may be used for audio.
- User control is mostly provided locally to the AR/MR device; some remote user interactions may be initiated from the tethered device as well.
- AR/MR functions (including SLAM/registration and pose correction) are either in the AR/MR device, or split.
- While media processing (for 2D media) can be done locally to the AR glasses, heavy AR/MR media processing may be done on the AR/MR tethered device or split.
- While such devices are likely to use significantly less processing than Type 1: 5G STAR devices by making use of the processing capabilities of the tethered device, they can still support a lot of local media and AR/MR processing. Such devices are expected to provide 8-10h of battery life while keeping a significantly low weight.

4.2.3 Interfaces

The following interfaces between the different functions of the device functional structure for device type #1 are defined:

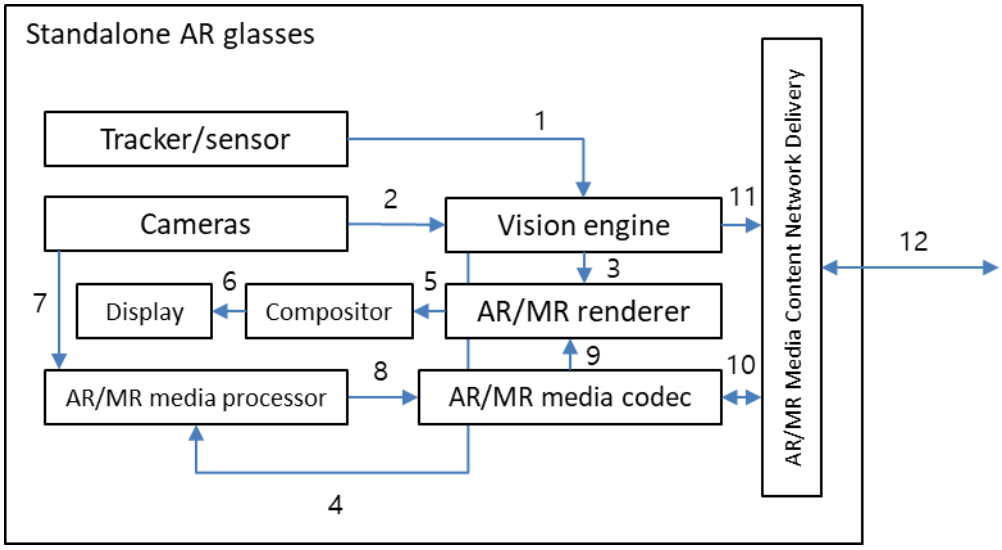


Figure 4.2.4: Interfaces for device type #1 functional structure

1) Tracker/sensor output interface

- Raw outputs from various tracking and sensor device components, namely IMUs (inertial measurement unit). These outputs are typically sent as inputs to the vision engine, which performs operations such as 6DoF pose generation.
- AR/MR devices might contain the following related components:

  a) Accelerometer – used by the system to determine linear acceleration along the X, Y and Z axes and gravity

  b) Gyro – used by the system to determine rotations

  c) Magnetometer – used by the system to estimate absolute orientation

2) Cameras output to Vision engine interface

- Depending on the device hardware, different types of camera outputs may be available as inputs into the vision engine or the camera outputs may be sent to the network. Possible camera signals include:,

  a) Visible light environment tracking cameras – typically gray-scale cameras used by a system for head tracking and map building

  b) Depth cameras (ToF)– these may be short-throw (near-depth) or long-throw (far-depth), depending on the application desired (e.g. hand tracking, or spatial mapping)

  c) Infrared (IR cameras) – used for eye tracking and hand tracking

  d) World facing RGB camera – used for image processing tasks and locating the camera's position in and perspective on the scene

  e) Media camera for capturing persons or objects in the scene for media data generation and consumption

3) Vision engine to AR/MR renderer interface

- The vision engine provides all the information required for the AR/MR renderer to adapt the rendering for a consistent combination of virtual content with the real world [4.3.0]. This information may be the output of vision engine processes such as spatial mapping, scene understanding, and room scan visualization. Vision engine processes are typically SLAM related, differing depending on the specific device and/or platform implementation.
- One typical output of the vision engine is the device/user pose information, which may include estimation properties depending on the service and application.
- Additional output can be 3D objects or representations for media consumption.

Editor's note: in the case of other (tethered) device types, the split of functions between entities are FFS.

4) Vision engine to AR/MR media processor interface

- The vision engine provides all the information required for the AR/MR media processor to perform processes such as 3D modelling. Information from the vision engine as used by the AR/MR renderer may also be sent to and used by the AR/MR media processor for relevant media processing.

5) AR/MR renderer to Compositor interface

- The AR/MR rendering typically outputs a rendered 2D frame (per eye) for a given time instance according to the device/user's current pose in his/her surrounding environment. This rendered 2D frame is sent as an input to the compositor. In the future, it can be predicted that non-2D displays will require a different output from the AR/MR renderer in order to support 3D displays e.g., Light Field 3D display.

6) Compositor to Display interface

- Rendered 2D frames are composited by the compositor before being passed onto the device display. The compositor may perform certain pose correction functions depending on the overall rendering pipeline used.

7) Cameras output to Media processor interface

- RGB and depth cameras are be used to capture RGB/depth images and videos, which can be consumed as regular 2D or 3D images and videos, or may be used as inputs to a media processor for further media processing.

8) Media processor to AR/MR media codec interface

- A media processor performs processes such as 3D modelling, in order to output uncompressed media data into the AR/MR media codec for encoding. One example of the media data at this interface is raw point cloud media data.

9) AR/MR media codec to AR/MR renderer interface

- Compressed AR/MR media content intended for rendering, composition and display is decoded by the AR/MR media codec, and fed into the AR/MR renderer. Media data through this interface may be different depending on the rendering pipeline.

10) AR/MR media codec to Network delivery interface

- Media contents that are captured or generated by the device (from interface 5 and 6) are encoded by the AR/MR media codec before being passed onto the network delivery entity for packetization and delivery over the 5G network. For 2D AR/MR media contents, this is typically a compressed video bitstream that is conformant to the video codec used by the AR/MR media codec.

- Media contents that are received through the network delivery interface over the 5G network are depacketized by the network delivery entity and fed into the AR/MR media codec. The subsequent decoded bitstream is handled through interfaces 7 and 4.

11) Vision engine to Network delivery interface

- Certain vision engine outputs may be sent to a remote processor which exists outside the device. Such data is passed from the vision engine to the network delivery entity for packetization and delivery over the 5G network.

12) Network delivery interface

- Network interface for AR/MR content delivery over the 5G network. In device type #1 for which interfaces are defined in this clause, the 5G modem exists inside the standalone AR glasses device.

4.3 Related Work

4.3.1 Related Work in 3GPP

This clause documents the 3GPP activity related to services using AR/MR device.

- 3GPP TR 26.928 [4.3.a] provides an introduction to XR including AR and a mapping to 5G media centric architectures. It also specified the core use cases for XR and device types.

- 3GPP TS 22.261 [4.3.b] identified use cases and requirements for 5G systems including AR and 3GPP TR 22.873 [4.3.c] is currently developing new scenarios of AR communication for IMS Multimedia Telephony service.

- 3GPP SA4 is working on the documentation of 360-degree video support to MTSI in 3GPP TS 26.114 [4.3.d]. It will provide the recommendations of codec configuration and signalling mechanisms for viewport-dependent media delivery.

- In the context of Release-17, 3GPP RAN work [4.3.e] is ongoing in order to identify a traffic model for XR application and an evaluation methodology to access XR performance.

4.3.2 MPEG

MPEG has developed a suite of standards for immersive media with a project of MPEG-I (ISO/IEC 23090 Coded Representation of Immersive Media). It contains all the media related components, including video, audio, and system for AR/MR as well as 360-degree video.

- Part 1 – Immersive Media Architectures: Provides the structure of MPEG-I, core use cases and scenarios, and definitions of terminologies for immersive media

- Part 2 – Omnidirectional MediA Format (OMAF): Defines a media format that enables omnidirectional media applications (360-degree video) based on ISOBMFF. The first edition of OMAF published in 2019 [4.3.f] supports 3DoF, and the second edition is currently being developed to support 3DoF+ and 6DoF.

- Part 3 – Versatile Video Coding (VVC): Describes the 2D video compression standard, providing the improved compression performance and new functionalities as compared to HEVC. The Final Draft International Standard (FDIS) was published in 2020 [4.3.g].

- Part 4 – Immersive Audio Coding: It provides the compression and rendering technologies to deliver 6DoF immersive audio experience.

- Part 5 – Visual Volumetric Video-based Coding (V3C) and Video-based Point Cloud Compression (V-PCC): It defines the coding technologies for point cloud media data, utilizing the legacy and future 2D video coding standards. The first edition of FDIS was published in 2020 [4.3.h].

- Part 6 – Immersive Media Metrics: It specifies a list of media metric and a measurement framework to evalutate the immersive media quality and experience.

- Part 7 – Immersive Media Metadata: It defienes common immersive media metadata to be referenced to various other standards.

- Part 8 – Network-Based Media Processing (NBMP): It defines a media framework to support media processing for immersive media which can be performed in the network entities. It also specifies the composition of network-based media processing services and provides the common interfaces. The FDIS was published in 2020. [4.3.i]

- Part 9 – Geometry-based Point Cloud Compression (G-PCC): It defines the coding technologies for point cloud media data, using techniques that traverse directly the 3D space in order to create the predictors for compression.

- Part 10 – Carriage of Visual Volumetric Video-based Coding Data: It specifies the storage format for V3C and V-PCC coded data. It also supports flexible extraction of component streams at delivery and/or decoding time.

- Part 11 – Implementation Guidelines for Network-based Media Processing
- Part 12 – Immersive Video: It provides coding technology of multiple texture and depth views representing immersive video for 6DoF.
- Part 13 – Video Decoding Interface for Immersive Media: It provides the interface and operation of video engines to support flexible use of media decoder.
- Part 14 – Scene Description for MPEG Media: It describes the spatial-temporal relationship among individual media objects to be integrated.
- Part 15 – Conformance Testing for Versatile Video Coding
- Part 16 – Reference Software for Versatile Video Coding
- Part 17 – Reference Software and Conformance for Omnidirectional MediA Format
- Part 18 – Carriage of Geometry-based Point Cloud Compression Data
- Part 19 – Reference Software for V-PCC
- Part 20 – Conformance for V-PCC
- Part 21 – Reference Software for G-PCC
- Part 22 – Conformance for G-PCC

4.3.3 ETSI Industry Specification Group

ETSI Industry Specification Group AR Framework (ISG ARF) has developed a framework for AR components and systems. The work has been derived from a collection of use cases and survey from the industries and it provides a general architecture of AR devices from the functional elements. In March 2020, ETSI ISG ARF published the Group Specification [4.3.0] and it is available in
https://www.etsi.org/deliver/etsi_gs/ARF/001_099/003/01.01.01_60/gs_ARF003v010101p.pdf

5 Core Use Cases

5.1 Introduction

This clause documents the core use cases and scenarios for AR/MR devices, which can be served to identify requirements, functional structure, related media format, and protocols for the 5G systems. Parts of the use cases are derived from XR use cases in TR26.928[x] based on the relevance to AR/MR device type. In addition, the other use cases and scenarios are collected in Annex A of this document.

Table 5.1 provides a list of all the collected use cases.

Table 5.1. List of use cases for AR/MR services

No	Use Case	Reference
1	3D Image Messaging	Annex A.2 in [x]
2	AR Sharing	Annex A.3 in [x]
3	Real-time 3D Communication	Annex A.8 in [x]
4	AR guided assistant at remote location (industrial services)	Annex A.9 in [x]
5	Police Critical Mission with AR	Annex A.10 in [x]
6	Online shopping from a catalogue – downloading	Annex A.11 in [x]
7	Real-time communication with the shop assistant	Annex A.12 in [x]
8	360-degree conference meeting	Annex A.13 in [x]
9	XR Meeting	Annex A.16 in [x]
10	Convention / Poster Session	Annex A.17 in [x]
11	AR animated avatar calls	Annex A.18 in [x]
12	AR avatar multi-party calls	Annex A.19 in [x]
13	Front-facing camera video multi-party calls	Annex A.20 in [x]
14	AR Streaming with Localization Registry	Annex A.21 in [x]
15	5G Shared Spatial Data	Annex A.24 in [x]
16	AR remote cooperation	Annex A.2
17	AR remote advertising	Annex A.3
18	Streaming of volumetric video for glass-type MR devices	Annex A.4
19	AR Conferencing	Annex A.5
20	AR IoT	Annex A.6

The use cases are grouped into several categories based on the similar requirements for media flow and device functional structure.

5.2 Summary of Core Use Cases

<Provides a summary in a form of Table>

6 Mapping to 5G System Architecture

6.1 General

< Editor Note : to be mapped onto one or more relevant architectures below

a) 5GMS downlink (extensions ongoing in EMSA, building on SA2/SA6 for adding edge)

b) 5GMS uplink (extensions ongoing in EMSA, building on SA2/SA6 for adding edge)

c) MTSI/Conversational

d) Interactive Immersive Services

>

< Note 2: Relevant to Network types

- Stand Alone AR UE (Type 1, 3, 4) = STAR
  - Rendering on device possible
  - Edge may be used for certain rendering
- Edge-Dependent AR UE (Type 2) =EDGAR
  - Rendering needs to happen on edge
  - Edge is needed for rendering
  - We should identify
    - Assumption on what is established
    - Any functionalities that need to be used for AR type of use cases when you use the edge/cl

6.2 Immersive media downlink streaming

6.2.1 Introduction

This clause introduces the case where immersive AR/MR media is streamed to a 5G AR UE using basic functionalities as defined in 5G Media Streaming for downlink (5GMSd).

6.2.2 Relevant use cases

<Thomas, Samsung>

6.2.3 Architectures

6.2.3.1 STAR-based

Figure 6.2.1 provides a basic extension of 5G Media Streaming download for immersive media using a STAR UE.

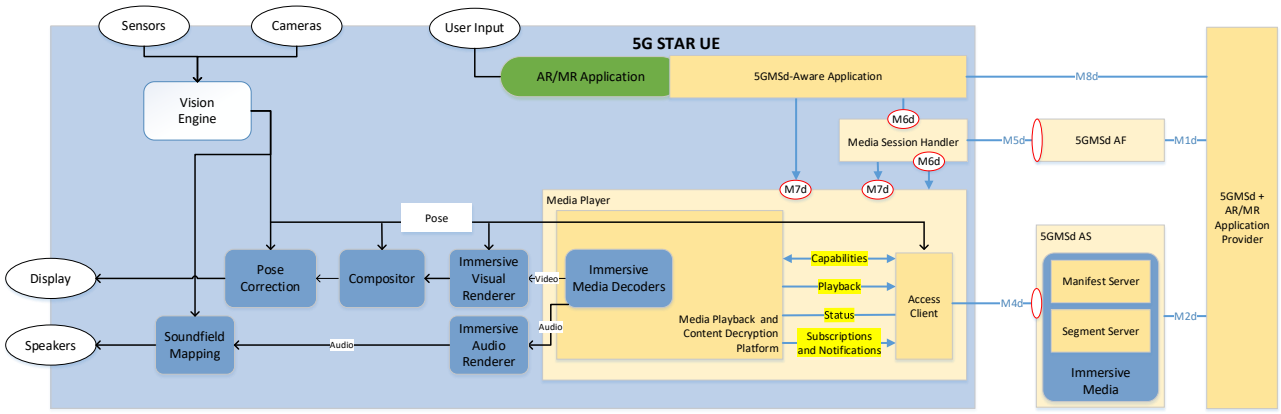


Figure 6.2.1: STAR-based 5GMS Download Architecture

6.2.3.2 EDGAR-based

Figure 6.2.2 provides a basic extension of 5G Media Streaming download for immersive media using an EDGAR UE.

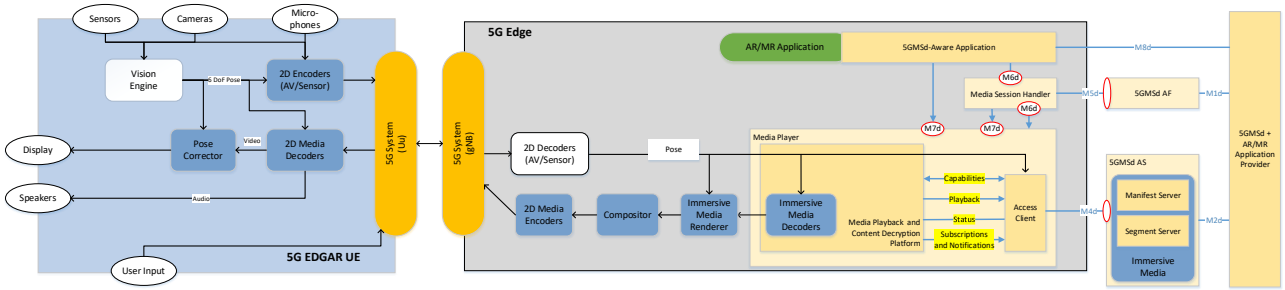


Figure 6.2.2: EDGAR-based 5GMS Download Architecture

6.2.4 Procedures and call flows

6.2.5 Content formats and codecs

6.2.6 KPIs and relevant quality of experience parameters

6.2.7 Potentially required QoS

6.2.8 Standardization areas

6.3 5G interactive immersive services

<Yago, Eric, Samsung>

6.3.1 Introduction

This clause introduces the case where interactive immersive service. In this case uplink information is pose.

6.3.2 Relevant use cases

6.3.3 Architectures

6.3.3.1 STAR-based

Figure 6.3.1 provides a basic extension of 5G Media Streaming download for immersive media using a STAR UE.

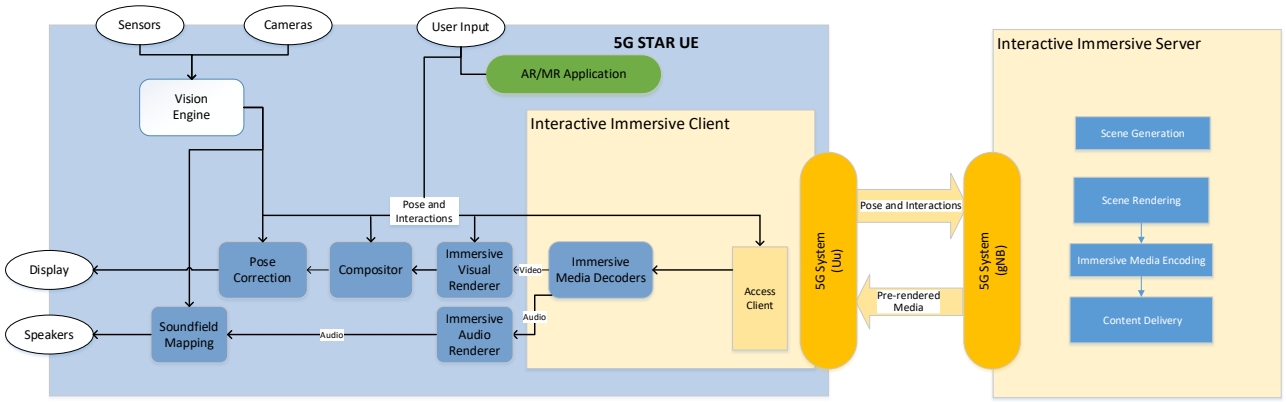


Figure 6.3.1: STAR-based 5G interactive immersive service architecture

6.3.3.2 EDGAR-based

Figure 6.3.2 provides a basic extension of 5G Media Streaming download for immersive media using an EDGAR UE.

Figure 6.3.2: EDGAR-based 5G interactive immersive service architecture

6.3.4 Procedures and call flows

6.3.5 Content formats and codecs

6.3.6 KPIs and relevant quality of experience parameters

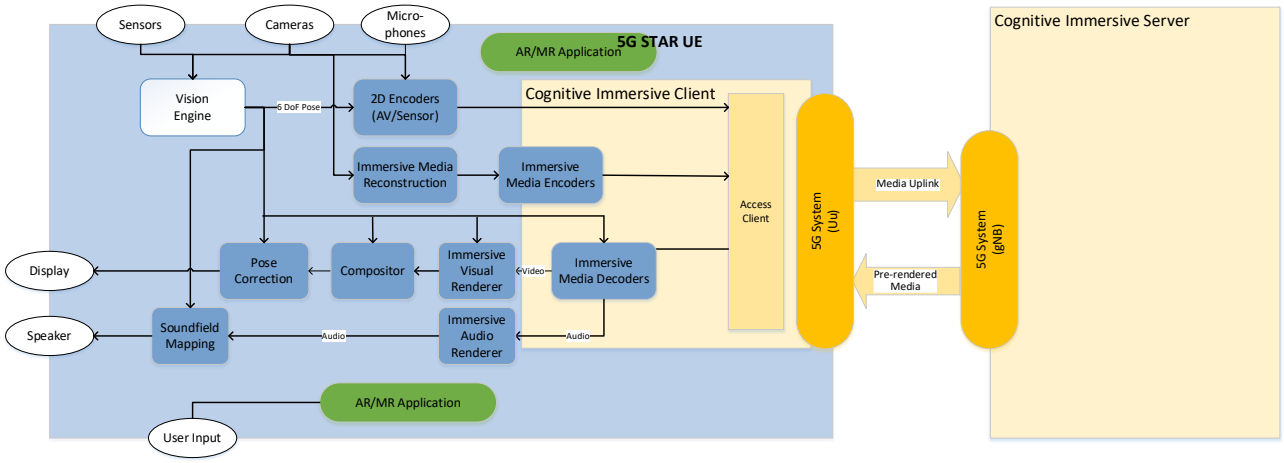
6.3.7 Potentially required QoS

6.3.8 Standardization areas

6.4 5G cognitive immersive service

<Thomas, Yago, ML, Samsung>

<Cognitive means that I am sending environment data>



6.5 AR two-party calls

<Ali - Ericsson, Samsung>

<Baseline UE-UE>

6.6 AR conferencing

<Ali - Ericsson, Samsung>

<Baseline MRF/MCU>

6.7 Summary

7 Devices Form-factor related Issues

<Provides the practical issues of glass-type AR/MR devices including overall power analysis and connectivity>

8 Potential Normative Work

<Identifies any missing or insufficient elements for AR glasses and potentially proposes the candidate solutions>

9 Conclusions

Annex A:
Collection of Glass-type AR/MR Use Cases

A.1 Introduction and Template

<Documents a common template for use cases proposal>

A.2 Use Case 16: AR remote cooperation

Use Case Name
AR remote cooperation
Description
As described in Annex A.9 of 3GPP TR26.928[x], a remote expert makes AR actions (e.g. overlaying graphics and drawing of instructions) to the received local video streams. This use case highlights that both parties can share their own video streams and overlay 2D/3D objects on top of these video streams compared with the scenario from TR 26.928. For example, a car technician contacts the technical support department of the car components manufacture by phone when he has some difficulty in repairing a consumers' car. The technical support department can arrange an engineer to help him remotely via real-time communication supporting AR. The car technician makes a video call with the remote engineer, uses his camera to capture the damaged parts of the car and shares them with the remote engineer in-call. And he marks possible points of failure by drawing instructions on the top of these video contents in order that the remote engineer can see the marks and make a detailed discussion. Also, they have respectively FOVs on their sides to check the failure. Likewise, the remote engineer can also overlay graphics and animated objects based on these shared video contents to adjust or correct the technician's operations. Furthermore, if the maintenance procedures are complex, the remote engineer can show the maintenance procedures step by step which are captured in real-time to the local technician. Therefore, the local technician can follow the operations. Finally, they find out the problems and fix them. It looks like that the remote engineer is beside the technician, discusses and solves the problems together. In the extension to this use case, it the remote engineer enables front-facing and back-facing cameras at the same time, the car technician can see a small video stream, which is captured by the front-facing camera of the remote engineer to achieve more attentive experiences.
Categorization
Type: AR, MR **Degrees of Freedom: 3DoF+, 6DoF** **Delivery: Interactive, Conversational** **Device: XR5G-P1, XR5G-A2, XR5G-A3, XR5G-A4, XR5G-A5, others**
Preconditions
<provides conditions that are necessary to run the use case, for example support for functionalities on the end device or network> Both parties on the device with the following features - Support for conversational audio and video

- Collect and delivery of AR actions and viewer information
- Enabling of the front-facing and back-facing cameras at the same time

The network with the following features

- Rendering of overlying AR actions and viewer information
- Rendering of virtual and real superposition of different video contents

Requirements and QoS/QoE Considerations

<provides a summary on potential requirements as well as considerations on KPIs/QoE as well as QoS requirements>

QoS:

- conversational QoS requirements
- sufficient bandwidth to delivery compressed 2D/3D objects

QoE:

- Synchronized rendering of overlay AR actions and pose information
- Synchronized rendering of audio and video
- Fast and accurate positioning information

Feasibility and Industry Practices

<How could the use case be implemented based on technologies available today or expected to be available in a foreseeable timeline, at most within 3 years?

- What are the technology challenges to make this use case happen?
- Do you have any implementation information?
  - Demos
  - Proof of concept
  - Existing services
  - References
- Could a reduced experience of the use case be implemented in an earlier timeframe or is it even available today?

>

Enhancements in media processing for multiple video streams both from different parties and/or the same party together with all kinds of AR actions may be performed in the network (e.g. by a media gateway) and in order to enable richer real-time experiences. Accordingly, the extensive hardware capabilities (e.g. multi-GPU) are required.

Potential Standardization Status and Needs

<identifies potential standardization needs>

- MTSI regular audio and video call between both parties
- Standardized format for AR actions (e.g. static and/or dynamic 2D/3D objects) and posture information
- Delivery protocols for AR actions and posture information
- Rendering of more than one video stream

A.3 Use Case 17: AR remote advertising

Use Case Name
AR remote advertising
Description
Compared with the use cases described in Annex A.8 and A.12 of 3GPP TR 26.928[x], this use case emphasizes that the shared video contents between two parties of a session are from a third party. Furthermore, the shared video contents may be 3D model objects, 360 degree and even free-viewpoint in order to help people have more interactive and immersive experiences. For example, a real estate salesman initiaes an audio call to a client by his smartphone to advertise houses remotely. The real estate salesman can request some video contents which are restructured 3D objects for houses to be sold/rent in advance from the third content provider and then switch a video call. The real estate salesman and the client can receive the video contents from the third content provider simultaneously. The real estate salesman can introduce via audio while he rotates the model. At the same time, the client can hear the introduction and see the rotational model via the touch-screen of his smartphone. And vice versa, the client is able to ask what he cares via audio while he is marking in different colours on the shared model, the client can hear the questions and see the colourful marks in real-time. In an extension to the use case, if the video content is free-viewpoint and embed some 3D objects representing furniture, the client wearing an AR-glass is able to see layouts and furnishings of the virtual houses which can rendered following his posture. It seems that the client is just inside the advertised and virtual house, and is able to walk around different rooms (e.g., dining room and living room). Furthermore, the client can draw a 3D object for a small couch back and forth in the living room using his hand. In addition, the real estate salesman can insert his 3D animated model in the virtual house and it can move following the view scope of the client as if he is just beside the client and introduces the house to the client. In another extension to the use case, the client can invite his friend to see the virtual house together. They can see it from their respectively viewpoint. They can also walk around the virtual house when wearing an AR-glass and communicate with each other via audio.
Categorization
Type: AR, MR **Degrees of Freedom: 6DoF** **Delivery: Interactive, Conversational, Download, Streaming** **Device: XR5G-P1, XR5G-A2, XR5G-A3, XR5G-A4, XR5G-A5, others**
Preconditions
<provides conditions that are necessary to run the use case, for example support for functionalities on the end device or network> the devices with the following features - Support for conversational audio and video - Support for receiving the video contents from the third party - Collecting of AR actions (e.g. rotation and mark) and posture information - Support of depth location technologies (e.g. SLAM for AR-glasses) the network with the following features - Rendering of overlying AR actions and posture information - Delivery of AR actions and posture information

- Support for establishing a connection the third party

Requirements and QoS/QoE Considerations

<provides a summary on potential requirements as well as considerations on KPIs/QoE as well as QoS requirements>

QoS:

- conversational QoS requirements
- sufficient bandwidth to delivery compressed 2D/3D objects
- Accurate user positioning information

QoE:

- Synchronized rendering of overlay AR actions and posture information
- Synchronized rendering of audio and video
- High-quality depth video captured from both parties

Feasibility and Industry Practices

<How could the use case be implemented based on technologies available today or expected to be available in a foreseeable timeline, at most within 3 years?

- What are the technology challenges to make this use case happen?
- Do you have any implementation information?
  - Demos
  - Proof of concept
  - Existing services
  - References
- Could a reduced experience of the use case be implemented in an earlier timeframe or is it even available today?

>

Enhancements in media processing for the media contents from the third party together with all kinds of AR actions and 2D/3D model objects may be performed in the network (e.g. by a media gateway) and in order to enable richer interactive and immersive experiences.

Potential Standardization Status and Needs

<identifies potential standardization needs>

- Delivery protocol of the shared media contents from the third party
- Standardized format and delivery protocols of AR actions and 2D/3D objects
- Standardized format and delivery protocols of posture information
- More than one communication channels can be setup

A.4 Use Case 18: Streaming of volumetric video for glass-type MR devices

Use Case Description: Streaming volumetric video for glass-type MR devices

Bob and Patrick are gym instructors and run a gym 'VolFit'. 'VolFit' provides their clients with a mixed-reality application to choose and select different workout routines on a 5G-enabled OHMD. The workout routines are available as high-quality photorealistic volumetric videos of the different gym instructors performing the routines. Bob and Patrick book a professional capture studio for a high-quality photorealistic volumetric capture of the different workout routines for their clients. Bob and Patrick perform the workout routines in the studio capture area. The studio captures Bob and Patrick volumetrically.

Alice is a member of 'VolFit' gym. Alice owns a 5G-enabled glass-type OHMD device. The 'VolFit' MR application is installed on her OHMD. The OHMD has an untethered connection to a 5G network.

Alice wears her OHMD device. The MR application collects and maps spatial information of Alice's surrounding from the set of sensors available on the OHMD. The OHMD can further process the spatial mapping information to provide a semantic description of the Alice's surrounding.

Alice wants to learn a workout routine from her instructors, Bob and Patrick. The photorealistic volumetric videos of Alice's instructors are streamed to the MR application installed on her OHMD. The MR application allows Alice to position the volumetric representations of Bob and Patrick on real-world surfaces in her surroundings. Alice can move around with 6DoF, and view the volumetric videos from different angles. The volumetric representations are occluded by real-world objects in the XR view of Alice; when Alice move to a location where the volumetric objects are positioned behind real-world objects or vice-versa. During the workout session, Alice gets the illusion that Bob and Patrick are *physically present* in her surroundings, to teach her the workout routine effectively.

The MR application allows Alice to play, pause and rewind the volumetric videos. The functions can be triggered for example by hand-gestures, a dedicated controller connected to the OHMD, etc.

Categorization

Type: MR (XR5G-A1, XR5G-A2, XR5G-A4, XR5G-A5)

Degrees of Freedom: 6DoF

Delivery: Streaming, Split-rendering

Device: OHMD with/without a controller

Preconditions

- The application uses existing hardware capabilities on the device, including A/V decoders, rendering functionalities as well as sensors. Inside-out tracking is available.
- Spatial mapping to provide a detailed representation of real-world surfaces around the device
- Media is captured properly (refer to clause 4.6.7. TR 26.928). The quality of the capture depends on different factors:

    1. Point-cloud based workflows
        - Studio setup i.e. camera lenses, distance of the captured object from the camera(s),

stage lights

- Filtering/Denoising algorithms

2. Mesh-based workflows

- Mesh reconstruction algorithms (e.g. Poisson surface reconstruction)
- Geometric resolution of the object i.e. poly counts
- Texture resolution e.g. 4K, 8K, etc.

- Media is accessible on a server
- Connectivity to the network is provided

Requirements and QoS/QoE Considerations

- QoS:
  - bitrates and latencies that are sufficient to stream a high-quality volumetric content within the immersive limits
  - bitrate for a single compressed volumetric video (mesh compression using tools such as Google Draco [A.1] and texture compression using video encoding tools such as H.264), for example, "Boxing trainer" sequence [A.2] further processed to generate a 3D mesh sequence with 65,000 triangles; 25fps, Texture: 2048x2048 pixels; 25fps:
    - Data rate of 47.3Mbps, which constitutes of following:
      - Mesh sequence: 37 Mbps (using Google Draco [A.1])
      - Texture sequence: approximately 10 Mbps (encoding using H.264)
      - Audio: 133 kbps (AAC)
  - access link bitrate estimates in case of split-rendering delivery methods (multiple objects):
    - approximately 30% higher bitrate than typical video streaming due to ultra-low delay coding structure (e.g. IPPP)
    - left and right view (packed stereo frame)
    - bitrates of a compressed stereo video depend on rendered objects resolution
    - bitrates approximately 1Mbps (small objects)-35 Mbps (objects covering majority of the rendered viewport)
- Required QoE-related aspects:
  - volumetric video captured roughly in the range of ~1-10 million points per frame (this is dependent on capturing workflows as well as the level of details in captured object e.g. clothes' textures)
  - high geometric resolution of the volumetric object's geometry to achieve accurate realistic simulations of rendering equations
  - frame rate at least 30 FPS and above
  - high-quality content rendering according to the user's viewpoint
  - real-time rendering of multiple high-quality volumetric objects
  - fast reaction to user's head and body movements
  - fast reaction to hand-gestures, or a connected controller, etc
  - real-time content decoding

- accurate spatial mapping
- accurate tracking

- Desired QoE-related aspects:
  - accurate scene lighting [Note: PBR feasibility]

Feasibility

Volumetric content production:

- Volucap studios: https://volucap.de/
- Mixed Reality studio: https://www.microsoft.com/en-us/mixed-reality/capture-studios
- Metastage: https://metastage.com/

Device Features:

- Spatial mapping
- Tracking
- Scene understanding [A.3]
- A/V decode resources

Selected Devices/XR Platforms supporting this:

- Microsoft HoloLens: https://www.microsoft.com/en-us/hololens
- Nreal Light glasses: https://www.nreal.ai/
- Magic Leap 1: https://www.magicleap.com/en-us/magic-leap-1

Current solutions:

For a real-time mobile on-device mesh-based system, an acceptable-quality experience for 30 FPS can be achieved using at least 30,000-60,000 poly count for a volumetric object's geometry with at least 4K texture resolution. On-device rendering of multiple complex 3D models is limited by graphics capabilities of the device.

In addition, advanced rendering techniques for lighting, reflection and etc, are subject to complex rendering equations which may result in inconsistent frame rate and increased power consumption. Therefore, it is challenging to achieve a real-time volumetric streaming for multiple high-quality 3D models with current networks and on-device hardware resources. Some existing solutions use remote rendering for streaming volumetric video:

- Azure remote rendering, https://azure.microsoft.com/en-us/services/remote-rendering/

, allows to render a huge and complex 3D model with millions of polygons remotely in cloud and stream in real-time to a MR device such as HoloLens 2. An intuitive demonstration of the Azure remote rendering of 3D model with approximately 18 million polygons on HoloLens 2 is publicly available at: https://www.youtube.com/watch?v=XR1iaCcZPrU

- Mesh-based multiple high-quality volumetric video streaming using remote rendering [A.4]. More information is available at: https://www.hhi.fraunhofer.de/5GXR

- Nvidia CloudXR: https://developer.nvidia.com/nvidia-cloudxr-sdk

Scene Lighting:

The light sources included in the scene have a significant impact on the rendering results. The light sources share some common properties such as;

- Colour: the colour of the light

- Intensity: the brightness of the light

Under Azure remote rendering, Scene lighting [A.5] provides the functionality to add different light types to a scene. Only the objects in the scene with PBR material type [A.6] are affected by light sources. Simpler material types such Color material [A.7] don't receive any kind of lighting.

Potential Standardization Status and Needs

The following aspects may require standardization work:

- Storage and access formats
- Network conditions that fulfill the QoS and QoE Requirements
- Relevant rendering APIs
- Scene composition and description
- Architecture design

A.5 Use Case 19: AR Conferencing

Use Case Description: AR Conferencing

This clause describes an AR conferencing use-case that allows participants in a 3D volumetric representation, e.g. point clouds or meshes, in order to provide an immersive conferencing experience.

3.2.1 AR Conferencing (1:1)

Bob and Alice want to make an AR conferencing call. Both are wearing AR glasses. Bob is located in Stockholm while Alice is located in Aachen. One or more cameras are placed in each location and are filming Bob and Alice, respectively. Bob can see a 3D volumetric representation of Alice on his AR headset and Alice can see a 3D volumetric representation of Bob on her AR headset. Bob and Alice can enjoy a truly immersive audio-visual experience.

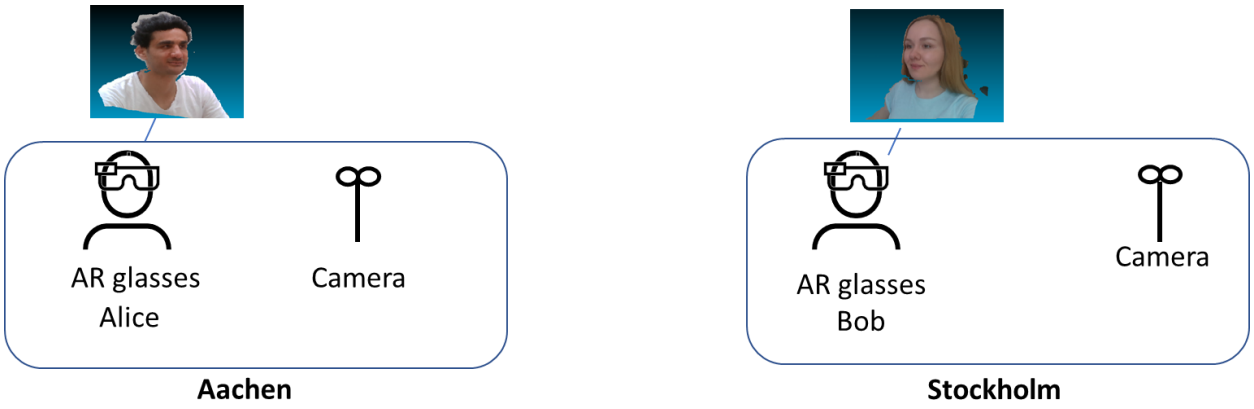


Figure 5.X-1: AR Conferencing (1:1)

3.2.1 AR Conferencing (1:many)

Bob and Alice are invited to an escalation meeting. Bob is able to physically attend the meeting, whereas Alice is virtually joining the meeting. Alice can be seen by Bob and other participants as a 3D volumetric representation on their AR glasses. Bob and other participants can interact with the 3D volumetric representations (e.g. rotate, zoom-in, resize). Alice can see and interact with Bob and other participants. Alice may use a laptop, phone, AR or VR device to visualize participants in the office. All participants can enjoy a truly immersive audio-visual experience.

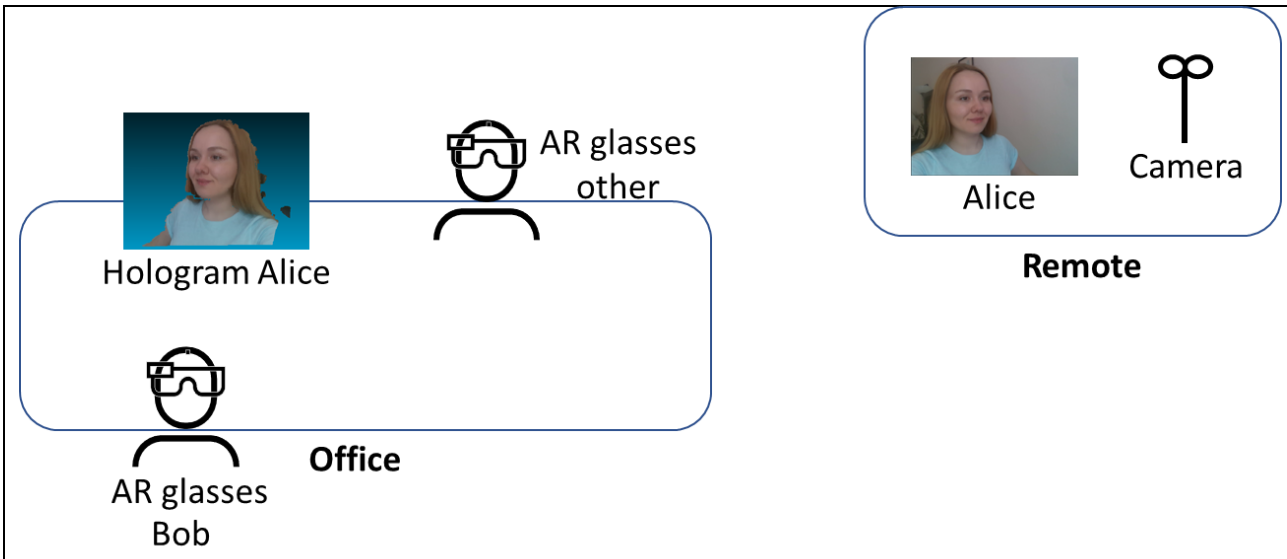


Figure 5.X-1: AR Conferencing (1:many)

Categorization

Type: AR

Degrees of Freedom: 3DoF+ or 6DoF

Delivery: Conversational

Device: AR glasses

Preconditions

- The participants are located in a room that is equipped with cameras that allow the capturing of participants including depth information. The movements can be captured by other means (e.g. AR glasses or phone camera).

- The participants are wearing AR glasses that allow the 3D volumetric representation of other participants.

Requirements and QoS/QoE Considerations

The network shall support the delivery of 3D volumetric streams for real-time conversational services:

- Support of different volumetric user representation formats.
- bitrates and latencies that are sufficient to stream volumetric user representations under conversational real-time constrains.

Feasibility

The bandwidth and latency requirements for AR conferencing using 3D volumetric representations present a challenge to mobile networks. The complexity of the 3D volumetric representations is challenging for the endpoints and introduces additional delay for processing and rendering functions. Intermediate edge or cloud components are needed.

In the following some indicative values of potential solution and transmission format for different types of user representation:

- A point cloud stream has raw bandwidth requirement of up to 2 Gbps. The transmission bandwidth is expected to be lower after encoding and optimization.
- Preliminary data from MPEG V-PCC codec evaluation indicates compression ratios ”in the range of 100:1 to 300:1”[A.11]. For dynamic sequences of 1M points per frame this could result into an encoding bitrate of “8

Mbps with good perceptual quality" [A.11]. For conversational services, we expect lower compression ratios.

- 2D/RGB+Depth: >2.7Mbps (1 camera @ 30fps with total resolution of 1080x960 [A.8]), >5.4Mbps (2 Camera @ 30fps with total resolution of 1080x1,920 [A.9]).
- 3D Mesh: ~30 Mbps @ 20-25 FPS (with a voxel grid resolution of 64x128x64 and 12-15k vertices) [A.10].
- Preliminary data from 3D GPCC show that bitrates in the range of 5-50 Mbps @ 30 fps with varying octree depth and varying JPEG QP are expected [A.10].

Potential Standardization Status and Needs

The following aspects may require standardization work:

- Standardized formats for 3D volumetric representation of participants on AR glasses.
- Cloud APIs for processing and rendering of 3D volumetric streams.
- Conversational methods for call initiation.
- Spatial audio formats and associated metadata.
- Metadata for Spatial characteristics of the AR environment (e.g. positioning of users).

A.6 Use Case 20: AR IoT control

Use Case Description: AR IoT control

Many IoT devices are present in the home and several of them such as smart light bulbs, smart curtains, air conditioning systems, heaters or multimedia devices are present in multiple numbers in different rooms within the home. While some IoT devices are at a fixed position and rarely move (light bulbs, heaters, curtains…) others are portable and nomadic by nature (portable speakers, vacuum cleaner robot, wearable devices…).

There are today many protocols that can be used for IoT (Wifi, Bluetooth, Zigbee, io-homecontrol, Z-wave but also LTE, NB-IoT and now Sidelink).and in the future there will be more and more IoT devices with 5G connectivity. While it is likely all IoT devices within a single home are interconnected through at least one (or more) gateway, D2D communications may also be used when available as it is likely to be more battery efficient for AR glasses.

As the user walk through his home, his AR glasses regularly scan the indoor environment to track the user's position in the home. While there could be several users wearing AR glasses in the same room, the environment reconstruction may also be using volumetric information from other IoT devices in the room (for instance security cameras with depth sensors). In terms of data exchange, upload data to 5G network may be video, depth-maps, sparse point clouds and also sensor information such as gyroscope or accelerometer.

Additionally, thanks to AR glasses' scanning, the home IoT system keeps track of IoT devices positions. Actual identification of an IoT device may also be done in the AR glasses themselves

Typically, the IoT home system runs on the edge network and information sent back to the AR glasses includes metadata information about IoT devices and environment, simple textual overlays for UIs. More advanced UIs would also probably make use of elements such as video or 3D objects

And finally, for D2D communications with 5G enabled IoT devices, control and status information is also exchanged between IoT devices and the AR glasses.

The use case addresses several scenarios:

- The user controls a specific IoT device by just looking at it through the glasses and operates it via an AR

displayed user interface and user controls such as touch or voice controls available on the glasses.

- Since home IoT system can know in real-time the position of the user in the home as well as which IoT devices are present in the same room as the user, the user can control IoT device with simple voice control without even targeting a specific IoT device. For instance, the user can say: "switch off the light" to operate the main light of the room he is currently in.

Categorization

Type: AR

Degrees of Freedom: 6DoF

Delivery: Interactive, Split, device-to-device

Device: AR Glasses

Preconditions

- IoT Home Application is installed on the AR glasses or phone connected to AR glasses
- The application uses existing HW capabilities on the device, rendering functionalities as well as sensors (audio, video, Lidar). Inside-out Tracking is available thanks to AR glasses sensors and may also be enhanced thanks to similar sensing capabilities on IoT devices.
- Connectivity to the network is provided on the glasses or through the connected phone.
- Wayfinding and SLAM is provided to locate user and map the environment and may be provided with split-processing (tethered device or 5G edge network).
- AR and AI functionalities are provided for example for Image & Object Recognition in order to track IoT devices positions.
- Connectivity to IoT devices may be device-to-device (Bluetooth, Sidelink, …), device to local IoT gateway (WiFi) or device-to-edge (5GNR).

Requirements and QoS/QoE Considerations

5G's low-latency high-bandwidth capabilities are used to provide real-time operation of IoT devices and high-speed upload of sensor information (video, Lidar point clouds).

Continuous connectivity is required for the sharing of local information to keep tracking user's position and position of IoT devices in the home so that home IoT system can maintain a real-time map of the home and its user.

The underlying AR maps should be accurate and should be up to date.

In order to optimize energy consumption on the glasses, split processing is favoured and efficient compression technology is required for exchanges with the home IoT system.

Device-to-device communication may also be used with some IoT devices.

Feasibility

- Google Visual Positioning Service: https://www.roadtovr.com/googles-visual-positioning-service-announced-tango-ar-platform/
  - XR clients continuously send sensing data to a cloud service. The service constructs a detailed and timely map from client contributions and provides the map back to clients. Example is Google's Visual Positioning Service
- Drivenet Maps – Open Data real-time road Maps for Autonomous Driving from 3D LIDAR point clouds: https://sdi4apps.eu/2016/03/drivenet-maps-open-data-real-time-road-maps-for-autonomous-driving-from-3d-lidar-point-clouds/
  - An XR HMD receives a detailed reconstruction of a space, potentially captured by a device(s) with superior sensing and processing capabilities. An example of navigation is given in the MPEG-I use case document for

point cloud compression (w16331, section 2.6)

- Xiaomi MIOT Ecosystem – Around the MiHome application, users can control all MIOT devices from many brands : https://xiaomi-mi.com/ecosystem/
  - MIOT devices can be controlled locally or through remote locations thanks to cloud servers. Some MIOT devices such as security cameras or vacuum cleaner robots have the capability to track objects and capture the environment (Audio, Video, Lidar) and make it available to the IoT home system.

Potential Standardization Status and Needs

The following aspects may require standardization work:

- Data representations for AR map of the home
- Collected sensor data to be uploaded to edge or tethered device (with dedicated compression solutions)
- Scalable streaming and storage formats for AR maps
- Content delivery protocols to access AR maps and content items
- Content delivery protocols to send AR UI elements when IoT application runs on tethered device or edge.
- Network conditions that fulfil the QoS and QoE Requirements
- Device-to-device communications.

Annex <X>: Change history

Change history							
Date	**Meeting**	**TDoc**	**CR**	**Rev**	**Cat**	**Subject/Comment**	**New version**
2020-08	SA4#110					Initial draft	0.0.3
2020-11	SA4#111	S4-201496				Agreement during SA4#111e: Use cases added	0.1.0
2021-01	postSA4#111 VIDEO Adhoc	Document withdrawn by mistake				SA4#111e late agreements: S4-201410; S4-201497; S4-201508	0.2.0
2021-02	SA4#112	S4-210113				SA4#111e late agreements: S4-201410; S4-201497; S4-201508 Clause 4.2, 4.3, Annex A.5 updated (Agreement from SA4#111e) Clause 4.2 further updated (Agreement from telcos prior to SA4#112e)	0.3.0
2021-02	SA4#112	S4-210213				Editorial updates on the Change history. Basis for integration of SA4#112 agreements	0.3.1
2021-02	SA4#112	S4-210215				SA4#112 agreements: S4-210214; - Editorial correction in Annex A	0.4.0
2021-02	SA4#112	S4-210267				SA4#112 further agreements: S4-210221; S4-210224; S4-210269. - AR IoT use case - Clause 4.2.2: updates on device type name and functional structure - Structure of Clause 6	0.5.0

Change history of this template:

2001-07	*Copyright date changed to 2001; space character added before TTC in copyright notification; space character before first reference deleted.*	*1.3.3*
2002-01	*Copyright date changed to 2002.*	*1.3.4*
2002-07	*Extra Releases added to title area.*	*1.3.5*
2002-12	*"TM" added to 3GPP logo*	*1.3.6*
2003-02	*Copyright date changed to 2003.*	*1.3.7*
2003-12	*Copyright date changed to 2004. Chinese OP changed from CWTS to CCSA*	*14.0*
2004-04	*North American OP changed from T1 to ATIS*	*1.5.0*
2005-11	*Stock text of clause 3 includes reference to 21.905.*	*1.6.0*
2005-11	*Caters for new TSG structure. Minor corrections.*	*1.6.1*
2006-01	*Revision marks removed.*	*1.6.2*
2008-11	*LTE logo line added, © date changed to 2008, guidance on keywords modified; acknowledgement of trade marks; sundry editorial corrections and cosmetic improvements*	*1.7.0*
2010-02	*3GPP logo changed for cleaner version, with tag line; LTE-Advanced logo line added; © date changed to 2010; editorial change to cover page footnote text; trade marks acknowledgement text modified; additional Releases added on cover page; proforma copyright release text block modified*	*1.8.0*
2010-02	*Smaller 3GPP logo file used.*	*1.8.1*
2010-07	*Guidance note concerning use of LTE-Advanced logo added.*	*1.8.2*
2011-04-01	*Guidance of use of logos on cover page modified; copyright year modified.*	*1.8.3*
2013-05-15	*1. Changed File Properties to MCC macro default 2. Removed R99, added Rel-12/13 3. Modified Copyright year 4. Guidance on annex X Change history*	*1.8.4*
2014-10-27	*Updated Release selection on cover. In clause 3, added "3GPP" to TR 21.905.*	*1.8.5*
2015-01-06	*New Organizational Partner TSDSI added to copyright block. Old Releases removed.*	*1.9.0*
2015-12-03	*Provision for LTE Advanced Pro logo Update copyright year to 2016*	*1.10.0*
2016-03-08	*Standarization of the layout of the Change History table in the last annex.(Unreleased)*	*1.11.0*
2016-06-15	*Minor adjustment to Change History table heading*	*1.11.1*
2017-03-13	*Adds option for 5G logo on cover*	*1.12.0*
2017-05-03	*Smaller 5G logo to reduce file size*	*1.12.1*