Media Services beyond 2D: View on Requirements and Architecture

This report analyses the potential challenges for the provision of Advanced Media Services using the 5G network. The analysis includes a series of use cases, media services and technology features within the scope of media delivery and the 3GPP system.

Contents

1. Interactive Hub

1.1. Architecture and Deployment

The convergence of cutting-edge technologies—including ultra-high-definition (UHD), 3D, and extended reality (XR) media, high-speed 5G connectivity, and machine vision AI—is transforming the traditional set-top box (STB) and TV service into a dynamic interactive hub. This evolution is shifting user behavior beyond passive TV viewing to immersive at-home interactive entertainment, opening exciting new possibilities for the industry.

One compelling example is the rise of interactive home fitness. This increasingly popular lifestyle choice allows users to follow on-screen workout videos while their movements are captured by a camera and analyzed by on-device vision algorithms. The application then provides real-time feedback and guidance, optimizing the user's exercise effectiveness.

Figure 1. The Concept of Interactive Fitness at Home.

1.1.1. Services & Content

Service providers can deploy their services and contents on the Cloud Engine, and the users need to pre-install the application on his/her device. The contents and services can include:

- UHD interactive media content;
- 3D or immersive media content;
- Cloud/Edge rendering CG content;
- Users can upload and share their achievements about fitness on social media any time.

1.1.2. Distribution & Access

All the media contents will be streamed and distributed to the end users at home which can be accessed through fixed network or 5G FWA.

- The Cloud Platform will provide content and service which includes video streaming;
- There can be a CDN system to accelerate the media content delivery which can help choose the best route to distribute the content to user's home.
- Computing power offloading at Edge node is enabled to cooperate with terminal devices for rendering and display complex CGs.
- User’s experience depends on the operator's guarantee of network QoS.

1.1.3. Terminal & Devices

Users can choose from variety types of terminal devices at home:

- Smart phone or tablet and other smart devices, STB + TV (Large Screen) with cameras;
- XR Headset, glasses and sensors;
- Smart exercise apparatus.

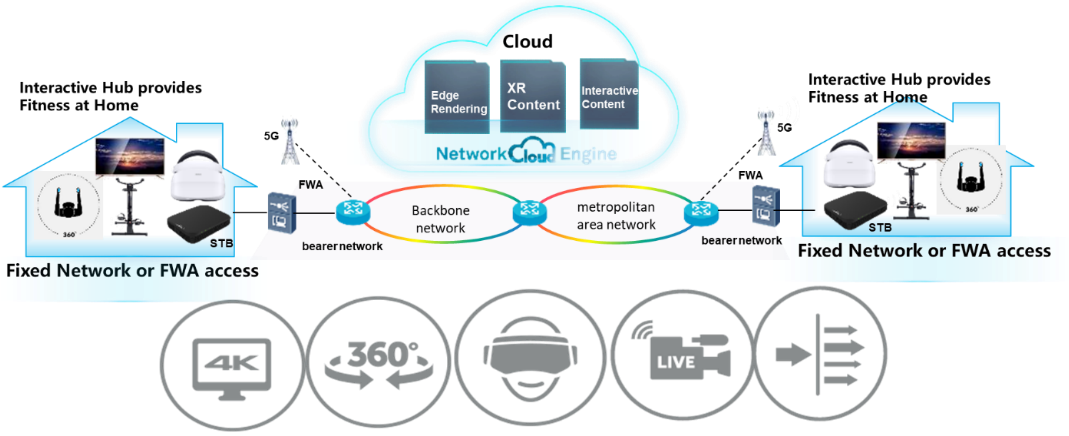


Figure 2. A Typical Deployment of Solution for Interactive Fitness

1.2. Content formats for production and distribution

In this workflow, all fitness content is created by the service provider and ingested into the Video Cloud Platform. From there, it's processed and distributed as streaming video for on-demand access via the Content Delivery Network (CDN). Simultaneously, a separate video stream is captured by the user's camera and uploaded to the Operation Platform. This stream can be shared as a live social media stream with friends and family. The Operation Platform also hosts the fitness assessment algorithm, which analyzes the user's performance in real-time, providing immediate feedback and a fitness score. This score is then used for social features like rankings and sharing with friends.

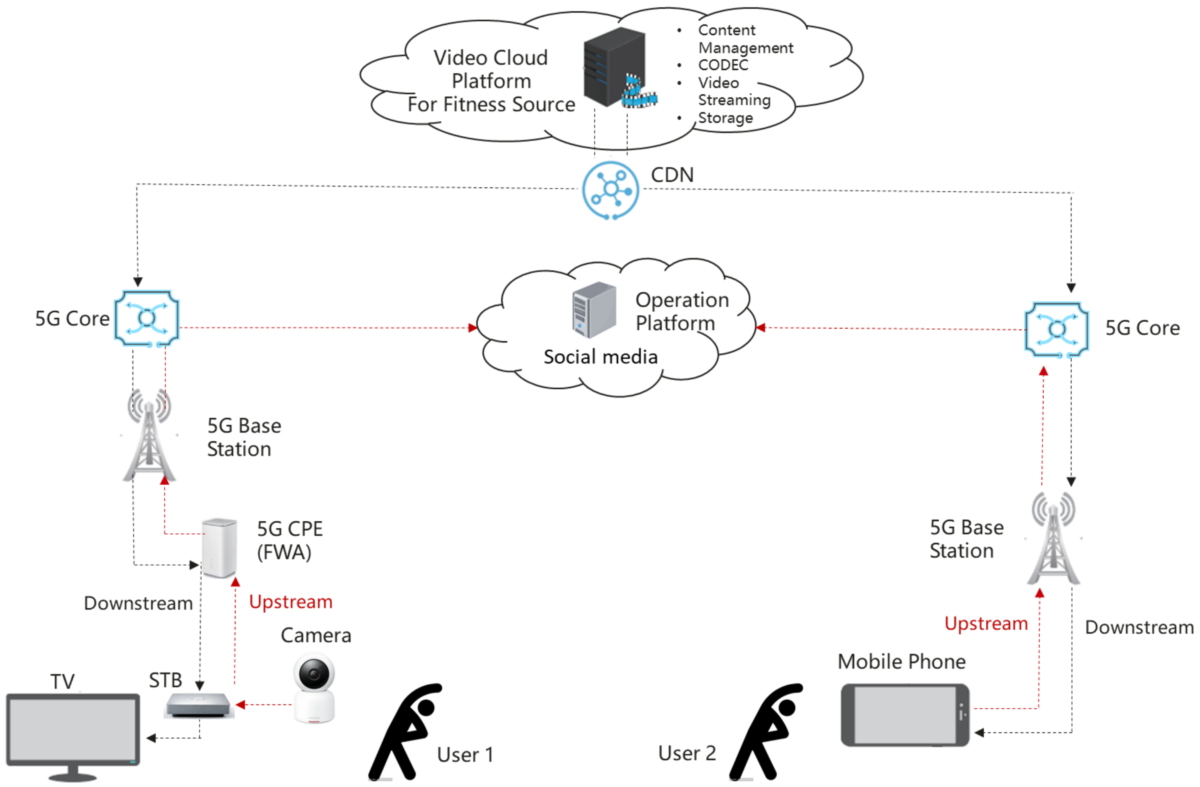


Figure 3. A work flow of Interactive Fitness.

Users can access the interactive fitness service through various devices, such as a set-top box connected to a TV or a mobile phone via 5G or Fixed Wireless Access (FWA). While the downstream traffic for high-quality fitness content requires significant bandwidth to ensure an immersive user experience, the upstream traffic—used primarily for sharing 1080p video on social media—is less demanding and unlikely to create an upload bandwidth bottleneck.

Requirements on Video Streaming for Interactive Fitness

Resolution	1080P	4K
Frame Rate	30FPS	50FPS
Downstream Average Bitrates	≥4Mbps @H.264	≥32Mbps @H.264
	≥2Mbps@H.265	≥16Mbps@H.265
	≥1Mbps@H.266	≥6Mbps@H.266
Download Bandwidth	1.5 to 6Mbps	10 to 46Mbps
Upstream Average Bitrates	4Mbps@1080P 30FPS	4Mbps@1080P 30FPS
Upload Bandwidth	6 Mbps	6 Mbps
Glass to Glass Latency	＜200ms	<200ms

1.3. Analysis of 3GPP systems supporting the use case

Analysis of the interactive hub use case suggests that existing 3GPP network KPIs are sufficient to meet the 5G network requirements for content distribution. These requirements align with the 5G QoS definitions outlined in 3GPP TS 23.501 (System architecture for the 5G System (5GS) Stage 2 (Release 18)). Furthermore, as described in 3GPP TS 23.316 (Wireless and wireline convergence access support for the 5G System (Release 18)), interactive services may benefit from dedicated Packet Data Unit (PDU) sessions, similar to those used for IPTV. These specialized PDU sessions can enable tailored QoS policies, optimizing the user experience for interactive applications.

2. Multiview video services

2.1. Architecture and Deployment

Multiview services enhance the viewing experience, particularly in real or virtual sports applications, by allowing audiences to switch between different camera angles or simultaneously access multiple perspectives.
A typical multiview video service system encompasses several key stages: front-end camera capture, on-site pre-processing, edge cloud media processing, a video cloud platform, CDN distribution, and finally, consumption on mobile phones and set-top boxes.

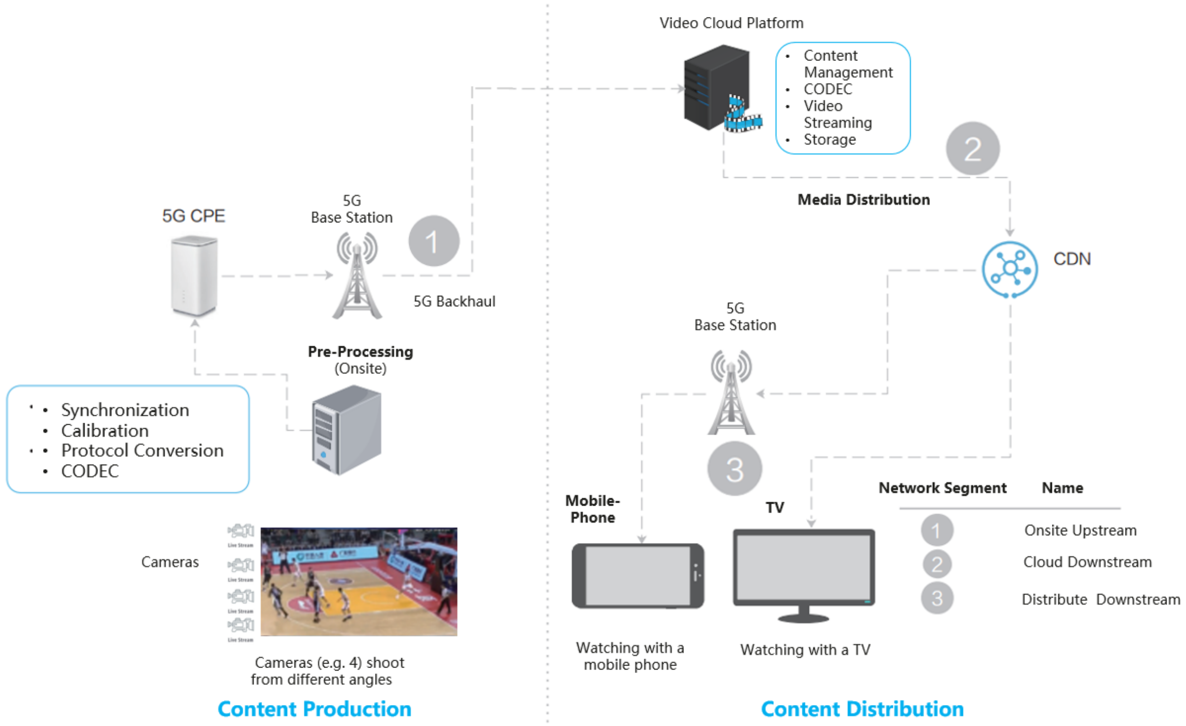


Figure 4. A typical Multiview video service system Architecture and work flow.

2.2. Content formats for production and distribution

2.2.1. Media acquisition

Content capture is realized by means of camera arrays that are able to provide multiple views of the free view-point scene and objects enabling their consumption from different angles and views. The capture formats characteristics used are captured in the table below.

In the case of on-site acquisition system, cameras and equipment (e.g., monitor, encoder and etc.) are usually connected through wired cables which means the requirements in this part are not related to 5G network.

Characteristics of the Multiview video on-site acquisition system

Typical number of cameras (range)	4	
Formats	RTMP For consumer equipment	SDI For professional equipment
Codecs	H.264/H.265/VVC (H.266)	Uncompressed digital video signals

Requirements (Non 5G mobile network)

Synchronization for a common clock to each of camera	<10ms

2.2.2. Media processing

To enable the generation of multi-view scene where viewpoints can be dynamically explored by users, the images from all cameras are stored and processed.

<u>Pre-Processing (on site)</u>

After obtaining video signals captured by multiple cameras on-site, the on-site equipment:

- Performs time synchronization (frame level synchronization) for multiple video contents to reduce the time differences between multiple cameras to millisecond level.
- Encodes the video content before upload to the video cloud platform for further processing.

Requirements (Upload to video could platform carried out through fixed networks or 5G network)

Typical bitrates (range) per camera for encode and upload	1080P 30FPS Video	4K 60FPS Video
H.264 AVC	10Mbps	60Mbps
H.265 HEVC	5Mbps	30Mbps
H.266 VVC	2Mbps	12Mbps
Bandwidth for upstream (4 cameras Average bitrates * 4 / 70%)	12Mbps – 58Mbps	69Mbps – 343Mbps
Synchronization among multi-view & multi-track with deviation	< 120ms	< 120ms

In this part, the bitrates of video will be used for content source production which are higher than the normal bitrates of video for distributions consumption.

2.2.3. Content formats for distribution and consumption

<u>Video Cloud Platform and CDN</u>

Video cloud platform performs the following processes:

- Content storage, management and operation.
- Video transcoding for multiple video slices and caching for different UE requests. An additional low quality video stream (240P) will be transcoded for each view point for pre-view.
- Video streaming with DASH technology to enhance the timeliness and real-time performance for video playback.

CDN performs routing scheduling and delivery by caching popular contents to the network node closest to UE. In the case of 5G wireless access, CDN can only distribute the content to the nearest 5GC/UPF, then it will be delivered to UE by 5G system.

The UE will decide whether to request high bitrate or low bitrate video slices based on the current state of the bandwidth of the 5G network.

Requirements for high bitrate (for download and consumption which is related to 5G network)

	Format	Codecs	Bitrates	Bandwidth	Synchronization among multi-view & multi-track with deviation
1080P 30FPS Video for mobile phone via 5G access	DASH	H.264	4Mbps	6Mbps	<120ms
		H.265	2Mbps	3Mbps	
		H.266	1Mbps	1.5Mbps	
4K 50FPS Video for TV via fixed network access and FWA	DASH	H.264	32Mbps	46Mbps	<120ms
		H.265	16Mbps	25Mbps	
		H.266	6Mbps	10Mbps	
240P 15FPS Video for Pre-view	DASH	H.264	128Kbps	< 1Mbps	<120ms

<u>User equipment requirements</u>

Normally, there are two types of terminals:

- 5G connected mobile terminal can access via 5G network.
- TV/STB can access via fixed broadband network or via 5G CPE FWA.

A 5G connected mobile terminal and TV/STB terminals will integrate the media player SDK for multi-view playback.

The processing of the terminal function:

- Support collaboration with CDN to request the DASH video slices
- Support to show one angle of video as a primary video on a single terminal chosen by audience. The audience does not experience any synchronization issues during the switch.
- No matter which angle of the primary video the user currently chooses to play, pre-view streams from other angles of the video will also be requested through the CDN system for playback. But the bitrates of those pre-view video will be only low-quality streams (total band widths < 1MB), so it will not be the bottle-neck of in the system.

Figure 5. A concept of the Multiview video service

2.3. Analysis of 3GPP systems supporting the use case

According to the previous analysis for the use case of multi-view services, from content production to content distribution, the network requirements related to 5G network at each stage can be met by the existing 3GPP network KPIs. Refer to 5G QoS definition in the Section 5.7.2.1 in 3GPP TS 23.501(System architecture for the 5G System (5GS) Stage 2 (Release 18))

According to the previous analysis for the use case of multi-view, DASH profile shall support multi-view video service. The 3GP-DASH defined in 3GPP TS 26.247 has been reviewed that it can meet the requirements.

3. Stereoscopic 360° multi-camera

Stereoscopic 360 multi-camera is an immersive media experience in which an audience can view and enjoy the action of scenes from various spots within an event like, for example, a music show or sports match. Users can enjoy the experience with a single HMD or with a mobile device as a second screen to complement the main 4K HDR production transmitted in DVB-T2 & 5G Broadcast.

Besides the main 4K HDR TV production, several stereoscopic 360 cameras can be set in different privileged positions in order to allow users to select the camera to watch from their homes. These cameras are transmitted as live-streaming to end users with personal computers, tablets, smartphones and HMD. They receive these streams via high bandwidth 5G gateway to Wi-Fi or 5G mobile data connections.

While TV programs are enjoyed on Smart TVs through DVB-T/T2 or 5G Broadcast, users also have the possibility to enjoy synchronously on an immersive 360 application from their 5G enabled mobile phones, and switch from several immersive cameras in 4K resolution, which offer a good quality experience on small screens.

Most advanced users can enjoy the same program using their 5G enabled HMDs with higher resolution of the immersive videos in stereoscopic 8K 360°.

3.1. Architecture and Deployment

Below there is a graphical example of the production and distribution of the immersive multi-camera video.

In this diagram, a number of N stereoscopic 360 cameras are set in different spots of a TV production studio or a venue of an event. For the graphical example, a musical show is chosen.

Every 360-camera spot shows a special interesting point of view like the stage, the jury, the participants, the audience in-situ or the production control.

The system allows the parallel encoding of the N camera signals for streaming, perfectly synchronised between them, to computers, tablets, smartphones and HMD.

5G devices, like smartphones or future HMD would leverage 5G technologies to optimize the user experience thanks to a very low latency and high bandwidth transmission, as well as multicast distribution possibilities inherent to new 5G standards and DRM for protecting the media contents.

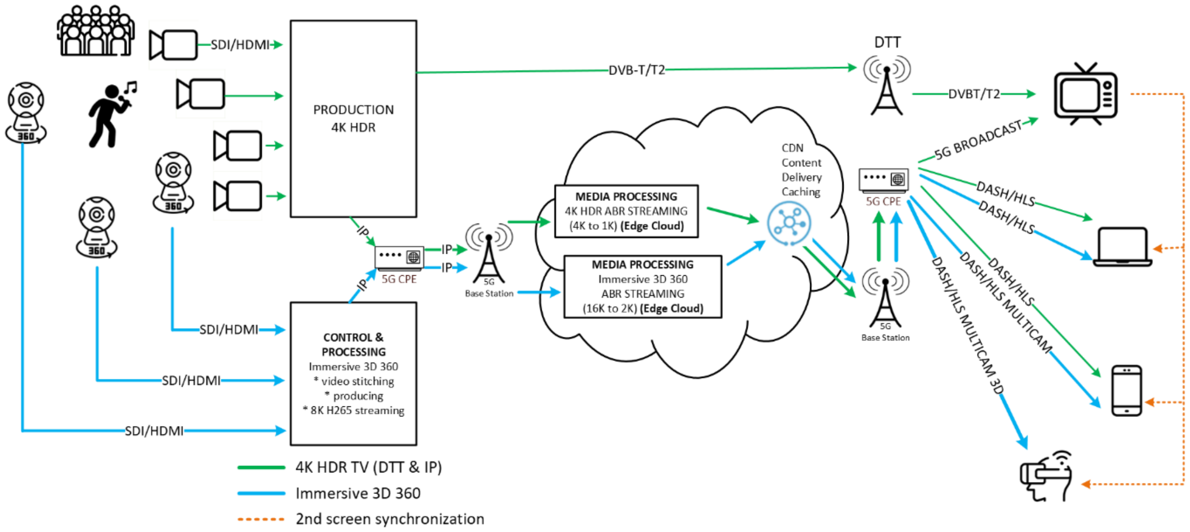


Figure 6. 360 multi-camera event

3.2. Content formats for production and distribution

3.2.1. Media processing

Content capture is realized by means of N 360° stereoscopic cameras installed in privileged positions around the event to allow to capture the action from the first line.

There are several 360° cameras solutions, with different output options like HDMI, SDI or IP Stream. In this document, it is always considered that the output of the camera has already been encoded in a high-quality stream, in order to contribute the resulting stream through IP network to cloud edge computing where it will be transcoded for broadband distribution.

On the one hand, there is the television production sent as a single signal made with 4K HDR quality and broadcast by DVB-T2 and 5G Broadcast as well as streamed to display devices such as computers and mobiles.

On the other hand, the signal from several 360° stereoscopic cameras, is sent with a quality of 8K to the cloud where the ABR signal will be processed with multiple qualities.

While the smartphone experience is offered with a 4K monoscopic 360° definition sufficient for small screens, in parallel 8K 360° stereoscopic high-quality streaming is sent for HMD glasses.

Below is a first table that calculates the approximate bandwidth required for the contribution of all the signals from the study or event to the cloud where they will be processed and prepared for distribution.

A 5G Private Network with a Gateway to fiber optics high bandwidth internet connection is installed in the event studio, so every camera is connected through 5G Private Network to the cloud edge for contribution, while high bandwidth wireless connection is guaranteed by the 5G technology.

Requirements of uplink on the stereoscopic 360 multi-camera use case

	Stream resolution	Compression codec	Stream BW
TV Production	4K HDR 50fps	H.265 (HEVC)	64Mbps (minimum level considered for contribution with visual lossless compression)
stereoscopic 360 camera	8K equirectangular 50fps	H.265 (HEVC)	512Mbps
TOTAL	(Assuming N=4)		2112Mbps

3.2.2. Media distribution

Cloud Edge Computing will process and transcode all required streams to be pushed to a CDN from where multi-quality streams will be distributed through IP networks.

Using the same methodology as DASH or HLS for multi-quality streaming, every camera is transcoded to multiple resolutions and bandwidth, so end users will always receive the best stream depending on bandwidth and resolution possibilities of their devices.

Next table represents the bandwidth load that the MEDIA PROCESSING module must deliver to the CDN for the distribution of all streams in ABR.

Requirements of down-link on the stereoscopic 360 multi-camera use case

	Stream resolution	Compression codec	Stream BW
Stereoscopic 360 camera (a 8K stereoscopic 360° equirectangular video, has a resolution of 8196x8196, which doubles the size of a standard 8K video)	8K 3D equirectangular 50fps High quality 8K 3D equirectangular 50fps Normal quality 4K 3D equirectangular 50fps High quality 4K 3D equirectangular 50fps Normal quality 2K 3D equirectangular 50fps High quality 2K 3D equirectangular 50fps Normal quality	H.265 (HEVC)	96Mbps 48Mbps 24Mbps 12Mbps 6Mbps 3Mbps Total: 189Mbps
TOTAL	(Assuming N=4)		756Mbps

All multi-quality streams need to be transmitted from the Media Server to the CDN for the distribution, from where every single end user will only receive one stream at a time using DASH/HLS standard. The User Interface in the device player will allow end user to switch between 360° cameras.

Latency for compression, contribution, recompression, distribution and decompression should be lower than the time for DVB-T2 or 5G Broadcast transmission to TV so buffering in the end-user device player will allow to adjust the delay and synchronize both streams with the aim to warranty the best immersive experience.

3.3. Analysis of 3GPP systems supporting the use case

According to the previous analysis for the use case of Stereoscopic 360º multi-camera, from content production to content distribution, the network requirements related to 5G network at each stage can be met by the existing 3GPP network KPIs. Refer to 5G QoS definition in the Section 5.7.2.1 in 3GPP TS 23.501(System architecture for the 5G System (5GS) Stage 2 (Release 18)).

Only in the scenario of multi-cameras for uploading 8K equirectangular 50fps videos, the requirement of upstream bandwidth is more than 2Gbps, which can be carried by a 5G Private Network with a Gateway to fiber optics high bandwidth internet connection. Users usually do not use 5G networks in this extremely high-bandwidth uplink scenario. Instead, they can flexibly choose to upload to the fixed network through cables or optical fibers,

4. Glasses-Free 3D video

4.1. Architecture and Deployment

There are two major glass-free 3D video display experiences:

- Single-view 3D display experience. This is usually applied at mobile phone or PAD and allows single user to watch a 3D video in a proper viewpoint. With the eye tracking system, the user still can watch the 3D video when changing the viewpoint within an applicable range.
- Multi-view 3D display experience: This is usually applied at TV set or outdoor billboard and allows multiple users (e.g., 50) watch a 3D video from multiple view point simultaneously.

Glasses-free 3D is any method of displaying stereoscopic images (adding binocular perception of 3D depth) without the use of special headgear, glasses, something that affects vision, or anything for eyes on the part of the viewer. With the rapid development of glass-free 3D display technology, 3D video is expected to be widely deployed in the next few years.

There are various glasses-free 3D display technologies, e.g., parallax barrier, lenticular lens, light field displays, etc., to offer better image quality with high resolution. Eye tracking and multiple views are two of the approaches that have been utilized to accommodate the motion of the viewer.

There are many technologies to produce and deliver 3D video. As shown in Figure 7, a typical Glasses-Free 3D video system includes front-end shooting, video cloud platform, CDN distribution, and consumption by mobile phone and set-top box terminals.

A set of reference points are identified, where information about commonly used codecs, protocols, and formats are elaborated further in the following subsections.

There are two types of terminals:

- Mobile phone can access via 5G network.
- TV/STB can access via fixed broadband network or via 5G CPE FWA.

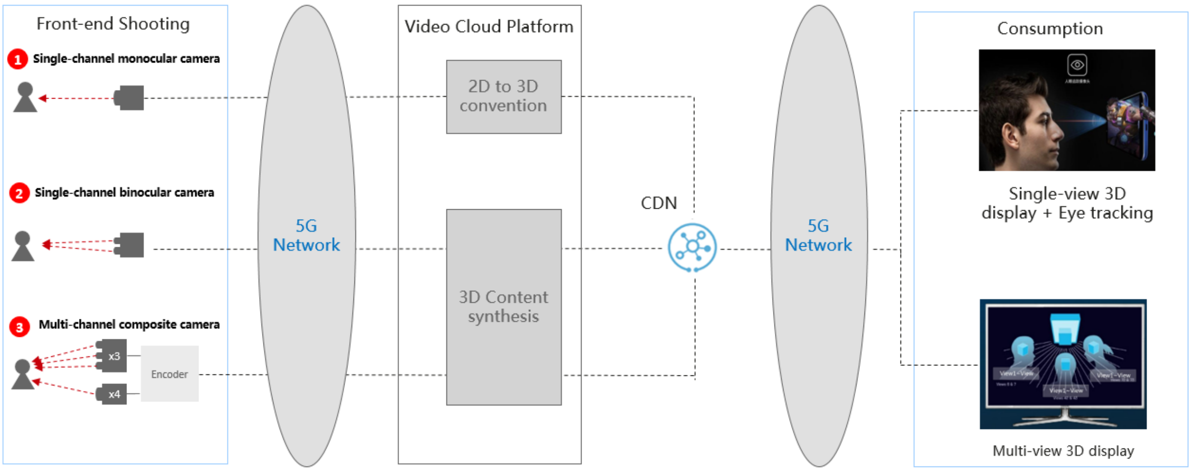


Figure 7. A typical glass-free 3D video system architecture and work flow

4.2. Content formats for production and distribution

4.2.1. Media acquisition

Typical, Content capture is realized by single-channel monocular camera, single-channel binocular camera or multi-channel composite camera.

Requirements of uplink on the Glasses-free 3D video acquisition scenario

Typical bitrates (range) per camera for encode and upload	1080P 30FPS Video	4K 60FPS Video
H.264 AVC	10Mbps	60Mbps
H.265 HEVC	5Mbps	30Mbps
H.266 VVC	2Mbps	12Mbps
Bandwidth for upstream for single-channel monocular camera (1 camera Average bitrates / 70%)	3 Mbps - 15Mbps	18Mbps – 90Mbps
Bandwidth for upstream for single-channel binocular camera (2 cameras Average bitrates * 2 / 70%)	6Mbps - 30Mbps	36Mbps –180Mbps
Bandwidth for upstream for multi-channel composite camera (7 cameras Average bitrates * 7 / 70%)	21 Mbps - 105Mbps	128Mbps – 630Mbps

4.2.2. Media processing

For the video captured by single-channel monocular camera, the 2D contents are converted to 3D contents. For the video captured by single-channel binocular camera or multi-channel composite camera, the contents are synthesised to 3D contents. For multi-view 3D display, the contents are converted to multi-view 3D contents which include the image. Depth-based conversion is most widely used technology in automatic conversion.

4.2.3. Content formats for distribution and consumption

Video Cloud Platform and CDN

Video cloud platform performs:

- Content storage, management and operation.
- Video transcoding for multiple video slices and caching for different UE requests.
- Video streaming with DASH technology to enhance the timeliness and real-time performance for video playback.

CDN performs routing scheduling and delivery by caching hot contents to the network node closest to UE. In the case of 5G wireless access, CDN can only distribute the content to the nearest 5GC/UPF, then it will be delivered to UE by 5G system.

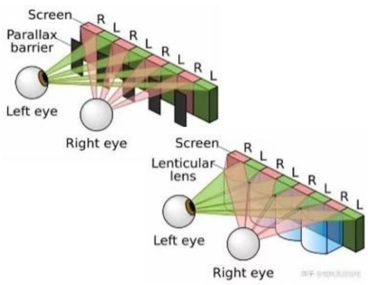
The Media player of UE will decide whether to request high bitrate or low bitrate video slices based on the current state of the bandwidth of the 5G network

Requirements of down-link on the Glasses-free single-view display 3D video distribution scenario

	1080P 30FPS Video	4K 60FPS Video
H.264 AVC	10Mbps	60Mbps
H.265 HEVC	5Mbps	30Mbps
H.266 VVC	2Mbps	12Mbps
Bandwidth for single stream	3Mbps - 15Mbps	18Mbps – 90Mbps
Bandwidth for a 3D displayer cost	6Mbps - 30Mbps (4K displayer needed)	36Mpbs – 180Mbps (8K displayer needed)

See the concept of autostereoscopy in Figure 8, half of the pixels are for left eye and the other half for right eye. Therefor a single-view 3D display device can only provide half-resolution of its hardware capability of experience to a user, e.g.;

- A 4K (3840 × 2160) monitor can request two 2K (1920 x 1080) video streams and display them.
- An 8K (7680 × 4320) monitor can request two 4K (3840 x 2160) video streams.



Autostereoscopy

Figure 8. The concept of Autostereocopy

Requirements of down-link on the Glasses-free multi-view display 3D video distribution scenario

Bitrate per single view	Number of viewpoints	Multi-view layout for image	Content resolution	Bottom Screen Resolution	Fps	Bandwidth cost
1080P (1920*1080)	4	2*2	3840*2160	4K	60	18Mbps – 90Mbps
2K (2560*1440)	9	3*3	7680*4320	8K	60	36Mbps -180Mbps
4K (3840*2160)	9	3*3	11520*6480	12K	60	108Mbps -540Mbps

See the concept of multi-view point 3D display in Figure 8, a screen will be divided according to the number of viewpoints, and its pixels will be evenly distributed to each region. Each. Specific pixels are specifically designed to serve users at specific view-point, e.g.:

- A 4K screen is divided as 2*2 regions to serve the users at 4 different view-point, each user can feel the experience of HD (1080P), and the total bandwidth cost will be similar as a 4K source video.
- A 12K screen is divided as 3*3 regions to serve the users at 9 different view-point, each user can feel the experience of UHD (4K), and the total bandwidth cost will be similar as a 12K source video.

Figure 9. The concept of Multi-view 3D display

<u>User equipment requirements</u>

Normally, there are two types of terminals:

- 5G mobile terminals/mobile phones can access via 5G network.
- TV/STB can access via fixed broadband network or via 5G CPE FWA.

Mobile phone and TV/STB terminals will integrate the media player SDK for 3D video playback.

The processing of the terminal function:

- Support collaboration with CDN to request the DASH video slices
- Support to show 3D video without wearing inconvenient 3D glasses.

Eyes tracking and display processing will be handled on device, and there are no additional requirements on network.

4.2.4. Analysis of 3GPP systems supporting the use case

According to the previous analysis for the use case of glass free 3D services, from content production to content distribution, the network requirements related to 5G network at each stage can be met by the existing 3GPP network KPIs. Refer to 5G QoS definition in the Section 5.7.2.1 in 3GPP TS 23.501(System architecture for the 5G System (5GS) Stage 2 (Release 18))

According to the previous analysis for the use case of glass free 3D services, DASH profile shall support multi-view video service. The 3GP-DASH defined in 3GPP TS 26.247 has been reviewed that it can meet the requirements.

5. Free-Viewpoint Video

5.1. Architecture and Deployment

As shown in Figure 10, a typical FVV system (with video stitching) includes front-end shooting, on-site pre-processing, edge cloud media processing, video cloud platform, CDN distribution, and consumption by mobile phone and set-top box terminals.

A set of reference points are identified, where information about commonly used codecs, protocols, and formats are elaborated further in the following subsections.

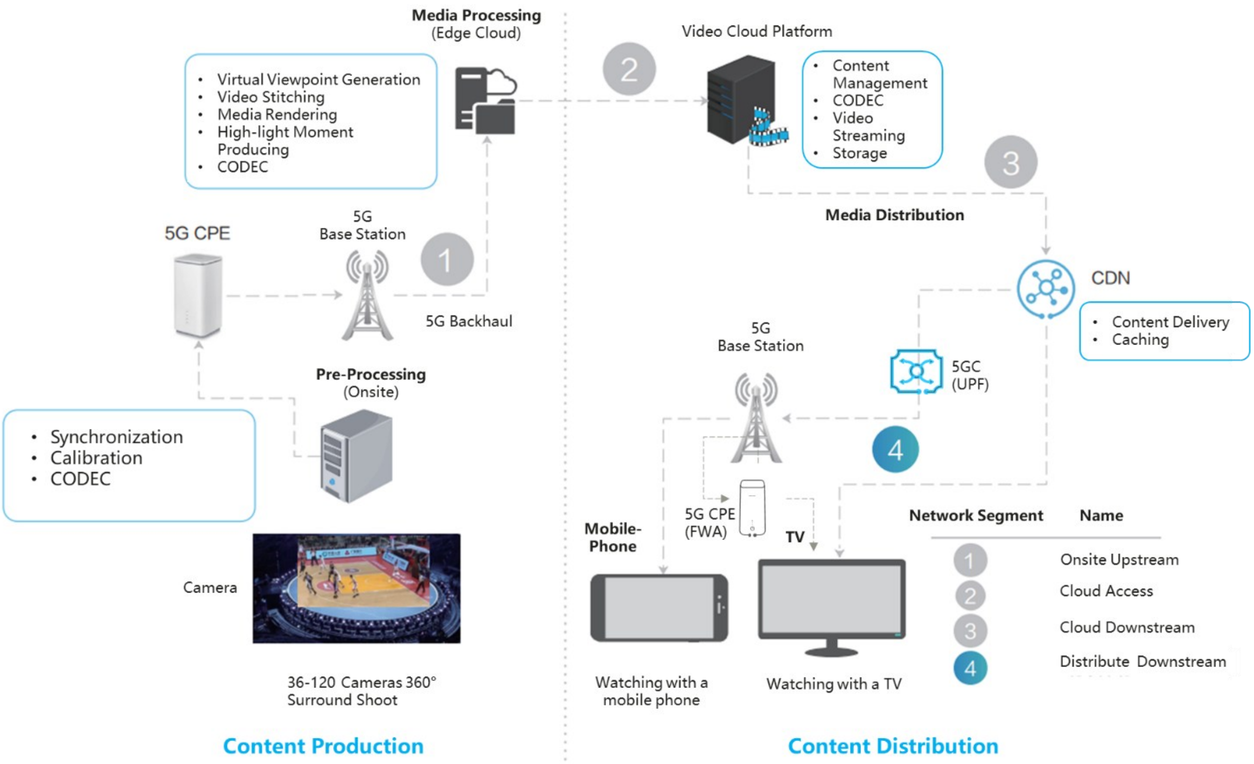


Figure 10. A typical FVV system Architecture and work flow

5.2. Content formats for production and distribution

5.2.1. Media acquisition

Content capture is realized by means of camera arrays that are able to provide multiple views of the free view-point scene and objects enabling their consumption from different angles and views. The capture formats characteristics used are captured in the table below.

In the case of on-site acquisition system, cameras and equipment (e.g., monitor, encoder and etc.) are usually connected through wired cables which means the requirements in this part are not related to 5G network.

Characteristics of the Free-viewpoint video on-site acquisition system

Typical number of cameras (range)	36 - 120	
Coverage of Stage by cameras	180° - 360°	
Formats	RTMP For consumer equipment	SDI For professional equipment
Codecs	H.264/H.265/VVC (H.266)	Uncompressed digital video signals

Synchronization requirements on the Free-viewpoint video service

Synchronization for a common clock to each of camera	<10ms

5.2.2. Media processing

To enable the generation of a free view-point scene where viewpoints can be dynamically explored by users, the images from all cameras are stored and processed.

<u>Pre-Processing (on site)</u>

After obtaining video signals captured by multiple cameras on-site, the on-site equipment:

- Performs time synchronization (frame level synchronization) for multiple video contents to reduce the time differences between multiple cameras to millisecond level.
- Uses AI algorithm to automatically adjust the centre point of multi-channel image focusing for accurate calibration cameras for focus and central axis position. This is to enable users to view from multiple perspectives via smooth switching between different angles, and the screen remains stable along the axis during rotation.
- Encodes the video content before upload to the edge cloud for further processing.

Requirements on uplink to edge cloud processing of the Free-viewpoint video service

Typical bitrates (range) per camera for encode and upload	1080P 30FPS Video	4K 60FPS Video
H.264 AVC	10Mbps	60Mbps
H.265 HEVC	5Mbps	30Mbps
H.266 VVC	2Mbps	12Mbps
Bandwidth for upstream (36 cameras Average bitrates * 36 / 70%)	100Mbps - 500Mbps	600Mbps – 3Gbps

<u>Media Processing (Edge cloud)</u>

Edge cloud computing with AI algorithms to process video contents from multiple cameras:

- AI algorithm will simulate multiple virtual view-points between each camera to make the rotation of the video smoother.
- Video stitching algorithm will combine all the video contents from different cameras to generate a whole FVV
- Video editing generates the high-light moment content.
- Encode the generated video content for streaming to inject into the video cloud platform.

Requirements on uplink to CDN platform of the Free-viewpoint video service

Typical bitrates of generated video for rotation viewing	1080P 30FPS Video for mobile phone access through 5G	4K 50 FPS Video for TV access through network or FWA
H.264 AVC	24Mbps (4Mbps for Static viewing + 5Mbps for Rotation viewing in low-bandwidth + 15Mbps for Rotation viewing in high-bandwidth)	112Mbps (32Mbps for Static viewing + 80Mbps for F viewing)
H.265 HEVC	12Mbps (2Mbps for Static viewing + 2.5Mbps for Rotation viewing in low-bandwidth + 7.5Mbps for Rotation viewing in high-bandwidth)	56Mbps (16Mbps for Static viewing + 40Mbps for F viewing)
H.266 VVC	5Mbps (1Mbps for Static viewing + 1.5Mbps for Rotation viewing in low-bandwidth + 2.5Mbps for Rotation viewing in high-bandwidth)	22Mbps (6Mbps for Static viewing + 16Mbps for R viewing)

5.2.3. Content formats for distribution and consumption

Video Cloud Platform and CDN

Video cloud platform performs:

- Content storage, management and operation.
- Video transcoding for multiple video slices and caching for different UE requests.
- Video streaming with DASH technology to enhance the timeliness and real-time performance for video playback.

CDN performs routing scheduling and delivery by caching hot contents to the network node closest to UE.

In the case of 5G wireless access, CDN can only distribute the content to the nearest 5GC/UPF, then it will be delivered to UE by 5G system.

The UE will decide whether to request high bitrate or low bitrate video slices based on the current state of the bandwidth of the 5G network

Requirements on down-link of the Free-viewpoint video service distribution

	Format	Codecs	Bitrates	Bandwidth (Range)	Interaction latency
1080P 30FPS Video for mobile phone via 5G access	RTMP DASH	H.264	9Mbps (4+5Mbps in low-bandwidth)	13Mbps -27Mbps	<300ms
			19Mbps (4+15Mbps in high-bandwidth)		
		H.265	4.5Mbps (2+2.5Mbps in low-bandwidth)	6.5Mbps – 14Mbps	
			9.5Mbps (2+7.5Mbps in high-bandwidth)		
		H.266	2.5Mbps (1+1.5Mbps in low-bandwidth)	3.6Mbps – 5Mbps	
			3.5Mbps (1+2.5Mbps in high-bandwidth)		
4K 50FPS Video for TV via fixed network access and FWA	RTMP DASH	H.264	112Mbps (32+80Mbps)	160Mbps	<300ms
		H.265	56Mbps (16+40Mbps)	80Mbps	
		H.266	22Mbps (6+16Mbps)	32Mbps	

Generally speaking, in FVV scenarios, in order to reduce glass-to-glass latency to improve users' experience by lowering video encoding and decoding latency that fast codec mode is used, which may result in a higher video bitrate in distribution.

User equipment requirements

Normally, there are two types of terminals:

- Mobile phone can access via 5G network.
- TV/STB can access via fixed broadband network or via 5G CPE FWA.

Mobile phone and TV/STB terminals will integrate the media player SDK for FVV playback.

The processing of the terminal function:

- Support collaboration with CDN to request the DASH video slices of different view-points for the interaction with users such as switching left rotation, right rotation, zoom-in and zoom-out actions.
- Support caching and orchestration all the video slices to provide real-time FVV playback smoothly within low latency;

5.3. Analysis of 3GPP systems supporting the use case

According to the previous analysis for the use case of FVV, from content production to content distribution, the network requirements related to 5G network at each stage can be met by the existing 3GPP network KPIs. Refer to 5G QoS definition in the Section 5.7.2.1 in 3GPP TS 23.501(System architecture for the 5G System (5GS) Stage 2 (Release 18))

However, the issue is that with increasing demand for bandwidth of advanced media services, high traffic costs are the key limiting factor for the development of new businesses, especially for cross-network export traffic, for example from fixed network to mobile network, as well as across networks between different MNOs, which is a big gap that needs to be solved.

In the scenarios of accessing by mobile devices and FWA, different users will request different content slices. Although these content slices are cached in the CDN of fixed network which locates at the N6 interface from 5G system architecture perspective , and all the traffic accessed by UEs has to go through N6 interface, but due to the cost of CDN egress bandwidth at N6 interface, when the traffic of all users go to the CDN node via N6 interface there will be a significant increase in high cost export bandwidth traffic from the fixed network CDN to the mobile core network UPF which shown in Figure 11, finally the high cost of export bandwidth will hinder the deployment of FVV service with high quality.

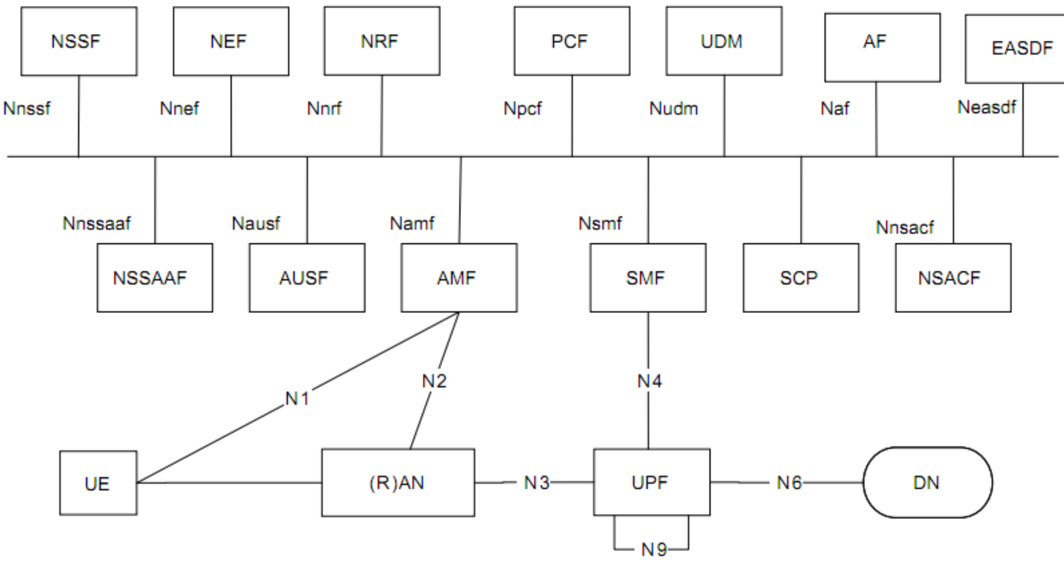


Figure 11. Non-Roaming 5G System Architecture

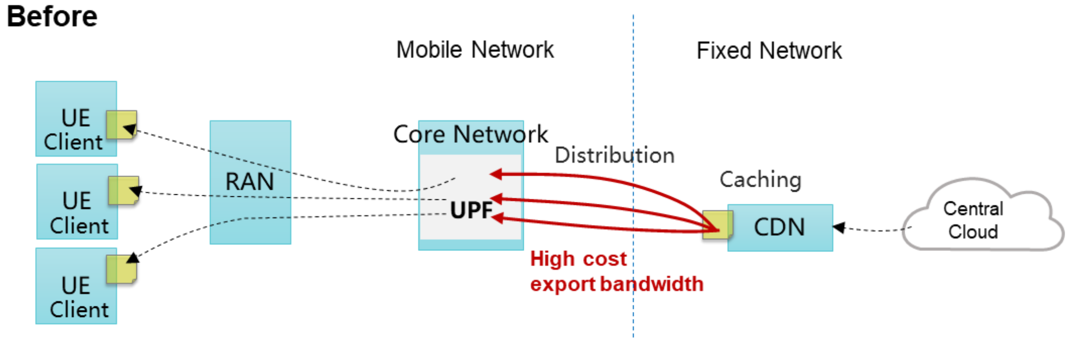


Figure 12 High cost associated with cross-network traffics

The existing network architecture can be optimized by adding a content caching and distribution capabilities to the core network of UPF, then UEs can access the content from the nearest node collocated at UPF, it can significantly reduce the export traffic between different networks, which will help solve this problem. See the proposed in-network delivery architecture in Figure 12.

Meanwhile, this solution also helps to provide users with a lower latency experience.

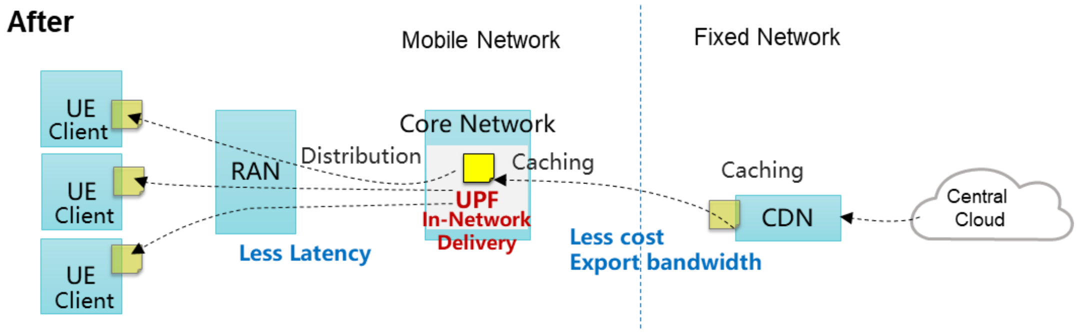


Figure 13. Less latency and less cost associated with cross-network traffic with in-network delivery

In 3GPP, TS 22.261 defines similar requirements on the in-network caching and delivery, see the description in Section 6.6.

It is suggested that further study should be considered to address the gap in the future work.

6. Watch Together

6.1. Architecture and Deployment

There are several OTT APPs supporting watch together experience, e.g., BT Sport watch Together, Tiktok VR Live. The users of different OTT service providers are like several isolated islands. Users have to install different application clients, and only users of the same specific OTT service providers can invite and share content with each other. Currently there's no way for users to invite and share watch together services across difference OTT service providers.

This clause describes a use case by integrating existing video service network and MNOs' IMS network. The IPTV network and IMS network of operators have a large enough number of existing users who can find and invite their friends to have a wonderful watch together time across different operators easily with just a phone number. Even more, due to the unified standards and architecture of 3GPP network, sharing among users across different MNOs is also technically feasible.

All the participants of watch together service are the subscribers of a TV service provider. One participant can share the TV programme (e.g., linear TV programme or on demand TV programme) with other participants and they can chat with each other when they are watching the programme as if they are sitting in front of the same TV. Synchronized playout is key to these experiences.

As shown in Figure 14, a typical watch together system includes video cloud platform, CDN distribution, IMS and consumption by mobile phone and set-top box terminals.

Logically speaking, it can be divided into two parts in Figure 14:

- IMS where the data channel services;
- Video service system.

A set of reference points are identified, where information about commonly used codecs, protocols, and formats are elaborated further in the following subsections.

There are two types of terminals:

- Mobile phone can access via 5G network.
- TV/STB can access via fixed broadband network or via 5G CPE FWA.

Mobile phone and TV/STB terminals will integrate the application supporting watch together.

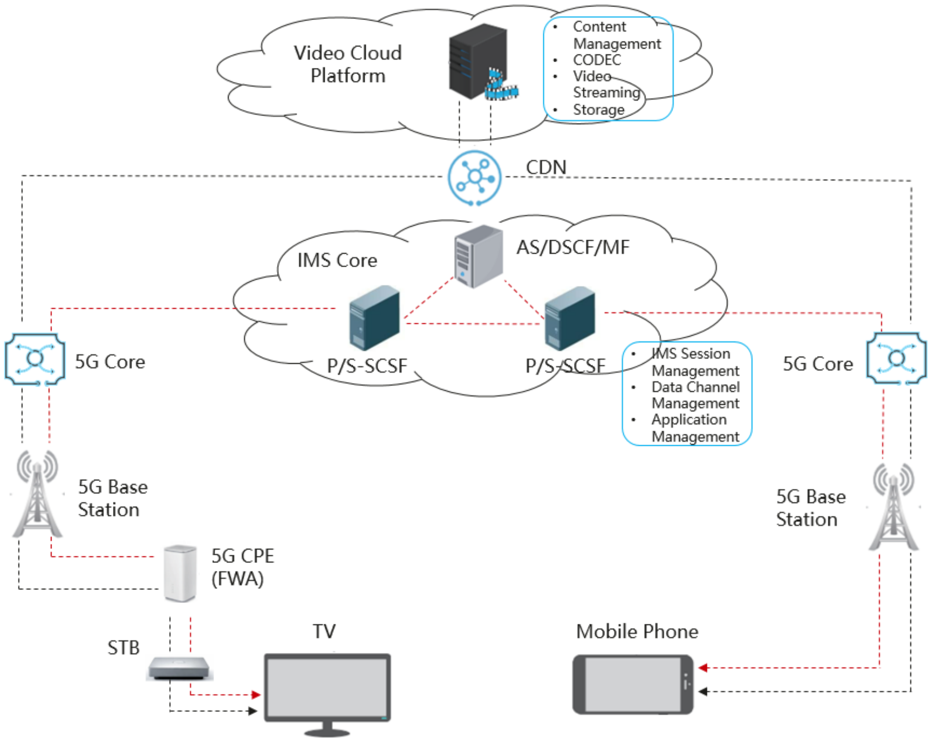


Figure 14. A typical watch together system architecture and work flow

6.2. Content formats for production and distribution

6.2.1. IMS session establishment and screen sharing

A specific application supporting watch together can be integrate into the terminals through pre-installation or dynamic download, e.g., it can be downloaded via the Bootstrap Data Channel set up Signalling Procedure as defined in annex AC.7.1 of 3GPP TS 23.228.

After launching the application and inviting the other audiences to participate the session performed by one audience acting as a chair, an IMS session among all the audiences will be established. During the session establishment, an application specific data channel is set up via the Person-to-Application (P2A) Application Data Channel Setup procedure as defined in annex AC.7.2 of 3GPP TS 23.228.

The video of everybody's face captured by their cameras will be shared among all the audiences via the established application specific data channel, compression and encoding. Details of the video of camera sharing depend on the implementation of the application. Followings are typical bitrates to meet the requirement of the video of camera sharing. The terminals can negotiate the bitrates based on the current state of the bandwidth of the 5G network.

Requirements of uplink on watch together camera sharing scenario

Bitrates based typical codecs	720P 25FPS Video	1080P 30FPS Video
H.264 AVC	2Mbps	4Mbps
H.265 HEVC	1Mbps	2Mbps
H.266 VVC	0.5Mbps	1Mbps

6.2.2.Video distribution and pay-out synchronization

Once the audiences acting the chair selects a programme and start to play, the URL of the segment to be played will be shared with the other audiences via the established application specific data channel. All the terminals retrieve the media segments from the CDN according to the URL and start to play. Any action (e.g., pause, seek, continue, etc.) performed by the audience acting the chair will also be delivered to the terminals of the other audiences during the experience of watch together. Details of pay-out synchronization depend on the implementation of the application which is important to the experience among all audiences.

In the case of 5G wireless access, CDN can only distribute the content to the nearest 5GC/UPF, then it will be delivered to terminal by 5G system.

The terminal can decide the requested bitrates based on the current state of the bandwidth of the 5G network

Requirements of down-link on watch together content distribution and play-out scenario

Typical resolution for Content Video	Format	Codecs	Bitrates	Bandwidth	Synchronization Latency
1080P 30FPS Video	DASH	H.264	4Mbps	6Mbps	<100ms
		H.265	2Mbps	3Mbps	
		H.266	1Mbps	1.5Mbps	
4K 50FPS Video	DASH	H.264	32Mbps	46Mbps	<100ms
		H.265	16Mbps	25Mbps	
		H.266	6Mbps	10Mbps	

6.3. Analysis of 3GPP systems supporting the use case

According to the previous analysis for the use case of watch together, the synchronization of play-out can be delivered via the established data channel for the application supporting watch together. The requirement of the application specific data channel can be supported by existing 3GPP as defined in annex AC of 3GPP TS 23.228.

According to the previous analysis for the use case of watch together, during content distribution, the network requirements related to 5G network at each stage can be met by the existing 3GPP network KPIs. Refer to 5G QoS definition in the Section 5.7.2.1 in 3GPP TS 23.501(System architecture for the 5G System (5GS) Stage 2 (Release 18))

7. Conclusions and Recommendations

After analysing the solution and workflow of 6 typical use cases of Advanced Media Services in this report which contributed by industry partners, we found that in terms of network KPIs, the existing 5G system can already meet most of the mainstream scenarios. But in a few cases the media traffic may bring high bandwidth requirements and reduce the mobile network capacity, e.g., multi-view or free-viewpoint videos with multiple cameras contents switching, there are quite a number of audiences who may access to same contents at the same time. In this case an efficient way of media delivery in mobile network in needed:

Please see the range of network KPIs range in the following table:.

Requirements summary of uplink and down-link bandwidth for this report

	Bandwidth Uplink	Bandwidth Down-link
Interactive Hub	6 Mbps	1.5 to 46 Mbps
Multi-view Video Service	15 to 86Mbps	1.5 to 46 Mbps
Stereoscopic 360°multi-camera	64Mbps	756Mbps
Glasses-free 3D Video	3 to 90Mbps	3 to 540Mbps
Free-Viewpoint Video	15 to 86Mbps	3.6 to 160Mbps
Watch Together	0.5 to 4Mbps	1.5 to 46Mbps

In addition to network KPIs, we found that with the development of advanced media services, there are still some potential gaps in network architecture aspects have been identified, as follows:

- In terms of network QoS for E2E latency and jitter. The current 3GPP system lacks an end-to-end QoS guarantee mechanism for high-value advanced media, E.g., In the use case of latency sensitive interactive fitness scenario which access through FWA, it expects that the 5G network can avoid the unpredictable latency and jitter at N6.

- In terms of media content distribution efficiency, for media scenarios accessed through 5G, both the media distribution within 5G system and media distribution at N6 interface are faced the distribution efficiency issues, E.g., in the scenario of free view-point video services, there are quite a number of audiences who may access to the same contents at the same time which will require high distribution cost at 5G system and at N6 interface, and it is expected that the network architecture can be improved to provide an efficient way of media content distribution.

Published by 5G-MAG | January 2025

www.5g-mag.com

5G Media Action Group (5G-MAG) Association
17A L'Ancienne-Route
1218 Grand-Saconnex (Switzerland)

info@5g-mag.com • www.5g-mag.com

This is a report produced by the 5G-MAG Workgroup UC (Use Cases, Requirements and Opportunities).

Version of the report: v.1.0

Date of publication:: 13th May 2025

This 5G-MAG Report can be downloaded from www.5g-mag.com/reports

Feedback from the industry is welcome through https://github.com/5G-MAG/Requests-for-Feedback