5GMediaHUB: Immersive AR, VR and XR Applications

Andrea Castelli
R&D Department
Brainstorm Multimedia
Valencia, Spain
acastelli@brainstorm3d.com

Øyvind Sørdal Klungre
Department of Electronic Systems
NTNU
Trondheim, Norway
oyvind.klungre@ntnu.no

Touradj Ebrahimi
Rayshaper SA
Crans Montana, Switzerland
touradj@rayshapersa.ch

Kambiz Homayounafar
RayShaper SA
Crans Montana, Switzerland
kambiz@rayshapersa.ch

Javier Montesa
R&D Department
Brainstorm Multimedia
Valencia, Spain
jmontesa@brainstorm3d.com

Andrew Perkis
Department of Electronic Systems
NTNU
Trondheim, Norway
andrew.perkis@ntnu.no

***Abstract*—5GMediaHUB aims at building a platform for development, testing and validation of advanced media applications over 5G by providing novel tools for applications developers in the media industry. To validate the platform, different Use Cases have been proposed. In this paper, the first use case “Immersive AR, VR and XR Applications” and its scenarios are described. The Use Case focuses on the transmission of novel multimedia services and content and how this can be optimized to be transmitted over 5G network infrastructures in a way to enhance relevant immersive features. The focus of the Use Case is on 360 immersive video represented at highest available 8K resolution and interactive immersive media using multiple portable devices.**

Keywords— Virtual reality, Interactive immersive media experiences, Panorama video streaming, Interactive Digital Narratives.

I. Introduction

The primary objective of the use case “Immersive AR, VR and XR Applications” within the 5GMediaHUB project [1] is to explore how 5G network infrastructures can deal with new types of interactive and immersive media services. Streaming of visual and multimedia contents have become very importance nowadays, but despite tangible improvement in graphic hardware capabilities, computing resources, rendering technologies and visualization solutions, networks capacity remains a bottleneck when delivering high-end immersive contents to end users. Thanks to the advances in 5G technologies, it is now possible for high-quality, 8K resolution, high dynamic range, and wide color gamut contents, to be both produced remotely and delivered to consumers with immersive and smart devices such as advanced handsets and portable VR headsets. The focus of this paper is on the transmission of high-quality and highly immersive audio-visual contents in real time over 5G in the framework of two specific scenarios:

- *Scenario I:* Live streaming and visualization of immersive panoramic 360 contents with an embedded presenter.
- *Scenario II:* Streaming and visualization of pre-recorded immersive contents using an Interactive Digital Narratives (IDNs) project [2] framework.

These two scenarios share technologies for streaming of immersive and interactive visual contents, related to virtual reality experiences, as depicted in Figure 1. In particular, the first scenario focuses on live VR experiences while the second, in addition, also offers interactive contents that are produced not only to be consumed on adequately selected immersive or conventional displays accessible to consumers but also can offer the latter enhanced experiences through second screens.

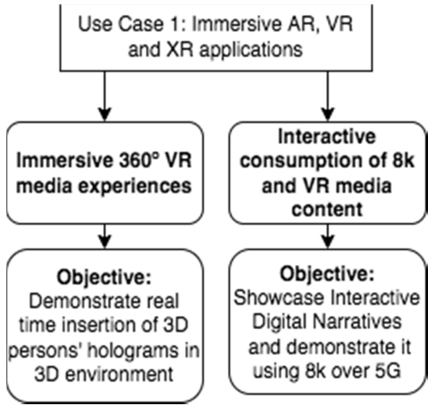

Fig. 1. Use Case Scenarios Diagram

The use case proposed in this paper allows the 5GMediaHUB consortium to validate the experimentation facilities in the project by testing their suitability in the context of immersive and interactive technologies. One of the specific objectives in such a testing is to deploy immersive applications within the 5GMediaHUB environment, to assess the offered Quality of Experience (QoE) [3] in the context of immersion [4], interaction and gamification of new media. The two mentioned scenarios have been designed to achieve the highest QoE by relying on transmission of highest quality video while offering the lowest latency in different usage contexts. The first context consists of transmitting immersive video to virtual reality headsets, with the aim of providing experiences where the end user will be immersed, in educational, business, or commercial paradigms. The second use case consists of participation of users in an interactive narrative, by transmitting an interactive 8K content through smart devices, in addition to the possibility of also making use of either tablets or virtual reality headsets for direct visualization of immersive scenes. Both scenarios require the use of 5G networks given the high demand for bandwidth to transmit high-resolution video in real time with low latency to enable responsive systems with dynamic interactivity that are essential in offering richer user experiences.

II. Materials and Methods

A. Scenario I – Immersive 360º VR Media Experience

Immersive 360º VR is emerging as a very promising enabling technology for the consumer market, as it allows users equipped with VR headsets to immerse themselves into a virtual environment where they can explore the latter in three degrees of freedom, also known as 360. Using specialized sensors at the VR headset, the video follows the user's head movements to present in real-time a 360 stereoscopic view of the scene projected in the viewing window of the headset's display. This use case will leverage Brainstorm's InfinitySet rendering engine, which allows for live 3D insertion of persons, captured over a green screen, in virtual environments. The environment is also rendered in real time by the same engine. As InfinitySet is a CPU and GPU intensive application, a cloud-based platform provides adequate CPU and GPU resources to the engine when the workload increases. The specific scenario under consideration in this use case consists of inserting remotely located teachers into a VR classroom and have them explain a certain concept to remote participants that will feel Presence and Immersion in the virtual classroom environment as if they were students taking a lecture in a physical classroom with a physically present teacher. Although such a compelling application is already feasible today with wired connectivity to the InfinitySet engine, the use case will further leverage on 5G technologies to allow untethered connection, which will significantly improve immersion and facilitates an unobtrusive 360 view without imposing any limitations on users' mobility because of the existence of bothersome cables in tethered configurations. This will be handled via the Immersive Media NetApp (a virtualized version of Brainstorm InfinitySet) within an eMBB (enhanced Mobile Broadband) slice, which will greatly simplify the implementation of 360 VR applications, as it will handle the mapping of the 360 VR sphere to the current users' field of view. Various techniques to reduce the bandwidth requirements will be tested, such as spatially partitioning the 360 VR sphere in tiles based on head tracking data and only streaming those that are part of users' viewing window and leveraging different video tiling and transcoding techniques to implement spatial mapping.

The proposed system setup for this scenario will allow, once in place, to connect remote teachers and remote students in a virtual room where students will be immersed via Head Mounted Displays (HMDs) in the scenario where the teacher provides explanations of a concept, and interactive graphics are used to reconstruct the scene. This scenario raises the possibility of a new type of VR headset that may emerge in the near future, thanks to the continuous evolution of such devices.

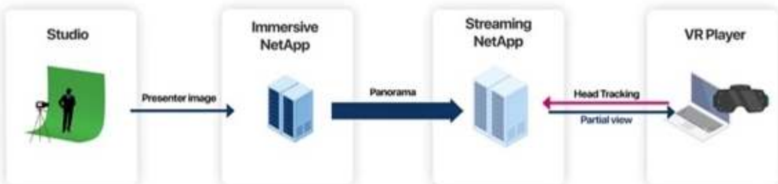

Fig. 2. Scenario 1 workflow

The latter will make use of a 5G modem to receive a video transmission with high-quality content, but instead of being generated in the nearby graphics station, it will be rendered in another location (e.g., in the cloud), and sent to the headset directly, so that the advantages of having a wireless device and a high-end graphics station come together, providing the possibility of displaying high-quality virtual reality content remotely in portable VR headsets. To make this possible, it is necessary to take advantage of unique features in 5G networks such as their greater bandwidth for the transmission of content and their lower latency to allow interactivity and adequate connectivity with content servers. The current technology and current VR headsets do not support such a direct 5G connectivity and are not capable of decoding high-resolution content in real time.

From a high-level perspective, this scenario workflow is based on two main systems; first the Immersive Media NetApp based on the InfinitySet graphics engine is responsible to host and to animate the scene, including interactive graphics, receiving teacher video stream, inserting it in the scene, and finally providing a final rendered content to be sent to the Streaming NetApp, in form of a 360 degrees sphere. The Streaming NetApp in turn will be responsible for decoding of the received stream and sending only the specific part of it to every student (i.e., the respective view port) depending on which part of the scene each student is looking at during the lecture. The head orientation information will be captured by the headset via the head tracking feature, sent to the laptop and transmitted to the Streaming NetApp as shown in Figure 2.

In addition to the previously mentioned NetApps, two dedicated software modules have been implemented; the teacher software module will be able to capture and to deliver a live video stream of the teacher, and also to trigger specific events related to the graphic environment, to modify and to evolve it during students' virtual experiences. These events triggered by the teacher will reach the graphic station and will allow control of the scenario's behavior, displaying an image or a video or bringing a diagram or a chart to the virtual scene.

The student's software module will connect to the Streaming NetApp, receives, and decodes the video stream and renders and delivers it to the VR headset. Moreover, it will capture the head tracking information and will send it to the Streaming NetApp to properly manage the video tiling method and deliver the correct viewing window to each user.

The scenario I is centered around a real-time high resolution and high-quality video transmission, and for that reason the requirements are mainly centered on high throughput of the underlying network and its low latency, because the need of a fast response from the video server to clients. These requirements are verified trough measurements of KPIs listed in Table 1.

TABLE I. SCENARIO I KPIS

KPI	Baseline value (achievable with 4G)	Expected value
Motion to photon Latency	> 50ms	< 10ms
Rendered Video Quality	< 4 MOS	> 4 MOS
Missing display area	> 25%	< 10 %
Application latency	50ms	< 10ms
Frame error rate	<0.1%	< 0.1%
Video jitter	0.5%	0.1%
Video resolution	HD	4k-30fps

B. *Scenario II - Interactive Consumption of 8K and VR media content*

This scenario demonstrates immersive media by using an IDN framework to implement the service on 5G networks and highlights the benefit from high bandwidth and low latency features of the latter. In IDNs, the user interacts with the story as it unfolds. A story will be implemented using transmedia storytelling, with content presented to the user on a first screen device. Meanwhile, user interaction happens on a second screen device.

The story has a working title *They Came – Nova.* It takes place within a science fiction universe. They Came is created by Norwegian media company Kapoow and recorded by RayShaper SA. A linear narrative is shown on a first-screen device, as an encoded video stream with several marked frames. A marked frame triggers an action on the second screen, where the user interacts with the narrative via an application. This marked frame is called a Keyframe Trigger. As a second screen device, the user might either use a 5G capable device (e.g., a 5G smartphone or a tablet), or an HMD for VR experiences. Three Keyframe Triggers are foreseen in the video stream.

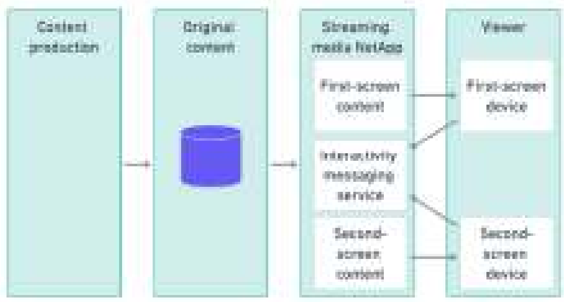

Fig. 3. Use Case Scenarios Diagram

TABLE II. SCENARIO II KPIS

KPI	Baseline value (achievable with 4G)	Expected value
MTP Latency	>50	<10ms
Rendered Video Quality	>4MOS	>4MOS
Application latency	>50ms	<10ms
Stall probability	<0.1%	<0.1%
Frame Error Rate	<0.1%	<0.1%
Video Jitter	0.5%	0.1%
Video bitrate	100Mbps	1000Mbps
Frame Rate	30fps	>30fps
Video Resolution	-	8k-30fps

In the scenario II architecture (Figure 3), all original contents of both the first screen device and the second screen device are stored on a server. The first screen content is pulled on demand and delivered to the preferred first screen (TV, laptop, tablet) by the IDN application, and viewing sessions are synchronised through the second screen app (either by login or device pairing through for instance QR-code verification). When a trigger frame is reached in the main story, IDN application will ping the Streaming NetApp that in turn pings the second screen app to perform the desired action. After an action is performed, the app pings the server, so that the main story can continue, until another trigger is reached. As the scenario is centred on video transmission, the related KPIs are mostly referring to video transmission and quality over 5G networks.

III. DISCUSSION

Our objective here is to define a use case for the early stages of the 5G system's rollout, in which users can enjoy data rates of up to 100 Mbps in the target area while at the same time benefit from services such as virtual reality and ultra-high definition (UHD) video. We anticipate that 5G subscribers will seek immersive multimedia experiences such as 8K UHD video, virtual reality, and real-time mobile gaming.

Currently, YouTube's 4K UHD video streaming service at 60 frames per second requires at least 53 Mbps data rates. These requirements are likely to increase manyfold with 8K UHD and future evolutions of video streaming. Additionally, specifications can support user-experienced data rates of up to 10 Mbps and peak data rates of up to 1Gbps for static users. These specifications support latencies of around 50ms. The proposed key performance indicators for our 5G use case are 100 Mbps as a baseline data rate, up to 20 Gbps at the maximum, and latencies less than 10ms. Thus, the user-experienced data rate should be increased by tenfold, while the peak data rate should be increased by 20 times over current 4G levels. To achieve real-time experiences in immersive 5G services, latencies should be reduced by at least 5 times.

How much compression would we need? First, let's consider the case of raw, uncompressed 8K video to get an order of magnitude estimation of what kind of (massive) data rates we would be dealing with in VR and UHD video (Table III).

TABLE III. 4K AND 8K BITRATES WITHOUT COMPRESSION

FPS	4K bitrate [bps]	8K bitrate at 16 bpp [bps]	8K bitrate at 32 bpp [bps]
24	3,185,049,600	12,740,198,400	25,480,396,800
30	3,981,321,000	15,925,248,000	31,850,496,000
50	6,635,520,000	26,542,080,000	53,084,160,000
60	7,962,624,000	31,850,496,000	63,700,992,000
120	15,925,248,000	63,700,992,000	127,401,984,000

The raw bit rate for 8K at 24 fps and YUV 4:2:0 chroma sampling would be about 12.74 Gbps, because7680 x 4320 x 24 fps x (8+4+4) = 12,740,198,400 bps. At 4K, we would have about 3.18 Gbps as3840 x 2160 x 24 fps x (8+4+4) = 3,185,049,600 bps. Yet, allowing for UHD TV frame rates of 120, 60, 50, 30 fps we would have bitrates shown in Table III. We would be dealing with a max raw bitrate of 127.4 Gbps. This is as high as we can expect. So, the need for noticing that the max bitrate for 8K video is about 63.7 Gbps at 16 bpp, we can also surmise that at 32 bpp, compression is crystal clear. But how much compression can we get with available and emerging codecs? H.264/AVC is widely available and a good starting point for analysis because newer codecs often report performance in relation to H.264/AVC. Back in 2011, Adobe published an H.264 Primer [1] as a guideline for estimating the acceptable bitrate for a video. Kush Amerasinghe, the author of the latter, suggested the following formula (known as Kuch Gauge) to estimate the bit rate necessary to code a video using H.264:

$$R = C.h.w.f.m_r \qquad (1)$$

With R representing the video bitrate in bps, h the picture height in pixels, w the picture width in pixels, f the framerate in fps and m_r the motion rank, a number with 1 representing low motion, 2 medium motion, and 4 high motion presence in the video. The degree to which the image data varies between frames is referred to as motion. Finally, C in the equation (1) corresponds to a constant that is estimated to be 0.07. As an example, let's take a high-motion (e.g., basketball game) HDR video with 4K@30fps:4K video bitrate = 4320 x 2160 x 30 x 4 x 0.07 = 69.67 Mbps. This is quite close to YouTube's recommended bitrate range of 44-56 Mbps for 4K HDR at 30 fps.

Kush Gauge works as an estimation of constant bitrate (CBR). For the case of variable bitrate (VBR), a target rate of approximately 75% of the estimate can be used as the target rate, and a maximum rate of approximately 150% of the estimate should be used as the maximum rate. This VBR disparity is highly dependent on the nature of the video and the ability of 5G network to absorb bit rate fluctuations. Table IV shows 8K bitrates after compression, with 7680 pixels height, 4320 pixels width, motion rank 4 and constant 0.07.

TABLE IV. 8K VIDEO BITRATE AFTER COMPRESSION WITH H.264/AVC

FPS	CBR [bps]	VBR Target [bps]	VBR Max [bps]
24	222,953,472	167,215,104	334,430,208
30	278,691,840	209,018,880	418,037,760
50	464,486,400	348,364,800	696,729,600
60	557,383,680	418,037,760	836,075,520
120	1,114,767,360	836,075,520	1,672,151,040

In H.265/HEVC codec, it would be reasonable to expect 50% improvement over H.264/AVC. As Table V shows, the highest bitrates we can expect are between 418 Mbps to 836 Mbps in the case of 8K at 120 fps with VBR.

TABLE V. 8K VIDEO BITRATE AFTER COMPRESSION WITH H.265/HEVC

FPS	CBR [bps]	VBR Target [bps]	VBR Max [bps]
24	111,476,736	83,607,552	167,215,104
30	139,345,920	104,509,440	209,018,880
50	232,243,200	174,182,400	348,364,800
60	278,691,840	209,018,880	418,037,760
120	557,383,680	418,037,760	836,075,520

IV. Conclusion

The use case presented in this paper and its technical requirements will be validated using the 5GMediaHUB experimentation platform. The Experimentation Tools will be used to measure the VR video QoE KPIs, ensuring that the bandwidth is sufficient, and the delay requirements are met. The physical resources utilization and the above reported KPIs will be monitored and measured via cross domain orchestrator, to ensure that high resource requirements do not saturate the server. In principle the most important scenario requirements are bandwidth between the Immersive Media NetApp and Streaming NetApp and very low latency between the Streaming NetApp and the final user device.

Acknowledgment

The presented work was supported by 5GMediaHUB project funded from the European Union's Horizon 2020 research and innovation program under the grant agreement number 101016714

References

[1] 5GMediaHUB project - https://www.5gmediahub.eu/

[2] H. Koenitz, G. Ferri, M. Haahr, D. Sezen, T. I. Sezen). Interactive Digital Narrative - History, Theory and Practice. Routledge (2015)

[3] Qualinet White Paper on Definitions of Quality of Experience (2012). European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), Patrick Le Callet, Sebastian Möller and Andrew Perkis, eds., Lausanne, Switzerland, Version 1.2, March 2013.

[4] Perkis, A., Timmerer, C., et al., "QUALINET White Paper on Definitions of Immersive Media Experience (IMEx)", European Network on Quality of Experience in Multimedia Systems and Services, 14th QUALINET meeting (online), May 25, 2020. Online: https://arxiv.org/abs/2007.07032

[5] K. Amerasinghe, H.264 for the rest of us, Adobe, Creative Commons