**3GPP TSG-SA WG4 Meeting #128 S4-241158**

**Korea, Jeju, 20 – 24 May 2024** merge of S4-240880 and S4-240915

Agenda item: 9.9

Source: Qualcomm Incorporated, Huawei, Philips

Title: [FS\_Beyond2D] Reference Model and Evaluation Framework

Document for Discussion and Agreement

# Introduction

This document combines the discussion in S4-240915 and S4-240880 to create a B2D reference model and evaluation framework. The reference model and evaluation framework are relevant to be agreed upfront in order in the scenario presentation to identify what would be needed to fully describe the scenario in the evaluation framework.

Note that the focus of the reference model and evaluation framework tries to achieve two aspects:

* It is independent of any specific codec, and codecs should not even be discussed
* It is generic and applies for any scenario to the extent possible.

With this approach, the framework can hopefully be used also for future scenarios.

# Reference Model for Beyond 2D Video

### 4.Y Overview

In contrast to well-established 2D-based video formats and work flows, for beyond 2D video a variety of emerging formats and reference workflows are under discussion. This aspect makes it more difficult to harmonize specific interop points and formats, also taking into account new developments in the industry and in research. In addition, without systematic and explicit identification of format interop points, beyond 2D scenarios or workflows may look overly complex.

However, basing beyond 2D workflows and scenarios on 2D reference workflows and formats, as for example evaluated in TR 26.955 [26955] and extending existing workflows seems to be promising way forward. However, when comparing for example to TR 26.955 [26955] for 2D formats or even omnidirectional video formats as defined in TS 26.118 [26118], additional aspects may need to be considered for beyond 2D video. To help the situation, a generic reference model for beyond 2D video content is is introducted in this sub-clause. This systematic and accurate identification of interoperability points and subcomponents for Beyond 2D video with a high level of abstraction covers the majority of use cases and scenarios.



Figure 4.Y-1 Beyond 2D Video end-to-end reference model

Figure 4.Y.1-1 illustrates a generic beyond 2D Video end-to-end reference model. Examplary, it considers three methods of creation of source content. The first apply a naturalistic way to capture sources indicated in reference point (1) and includes for example variants of UE-based cameras and sensor (1a) or a multi-camera production setup (1b). The third option is based on authoring using computer graphics interfacing technologies or other media production technologies. These may sometimes be combined and, possibly with slight variations, these options cover the majority of media production cases.

The capture of content using cameras, for example light fields using sensors is generalized including traditional *passive sensors*, cameras, camera arrays, or plenoptic cameras. For simplicity we also include *active sensors* LiDaR, Time of Flight in this category. These active sensors also transmit a signal before capturing the reflections. Depending on the setup, the collected data may be quite different depending on the capturing system, expressed in reference point (2) with variants (2a) and (2b).

In the general case, some processing based on the captured data would happen to generate a a well defined B2DV format, possibly referred to as *sensed data converter*. This step is mainly about converting the multiple digital image formats plus metadata to a well-defined beyond 2D representation or format, referenced with reference identifer (3).

For typical 5G workflows, a compressed digital representation of the B2DV is needed for efficient transmission. The *B2DV Encoder* executes a processing step that will result in the compressed Beyond 2D video bitstream that includes a digitally compressed lossy version of the B2DV format. and optional metadata, referred to as reference point (4). The B2DV bitstream is typically delivered through content delivery protocols and 5G radio systems, not shown in Figure 4.2.1-1.

The *B2DV decoder* decodes the B2DV video bitstream and recovers a B2DV format, presented in reference point (5). The recovered signal is forwared to the rendering and display system. In some cases, one viewport of the B2DV format may for example be displayed directly in a 2D Presentation System. In a 3D Presentation System, interacting with the rendering component may allow to generate different views on the content. In an immersive presentation system, pose information may be used to render the views of the content. The format generated by the renderer for the presentation system, indicated in reference point (6), is implementation specific as shown in Figure 4.2.1-1.

Generally, beyond 2D video performance measurement should typically be between interop points (3) and (5) based on the B2DV formats. The last block in the diagram includes the user interactions. Some B2DV scenarios may involve some types of user interactions, such as changing the viewpoint or other interactions. These are captured in the reference diagram for completeness.

# Considered Scenarios

For the B2D study and based on the reference model, we consider the following scenarios as good candidates to be further detailed.

Messaging (aligned with consideration in TR 26.955, Scenario 4): A UE generates B2D video content in real-time and shares the media content with one or several UEs with B2D capabilities. The scenario reflects what is presented in TS 26.143. The following high-level constraints apply

* B2D content can be captured on existing or emerging devices
* Real-time B2D video encoding on market-relevant device is possible
* Encoding latency constraints relaxed
* Uploading of the formats through 5G network is possible
* Media can be packaged into a messaging format

Streaming (aligned with consideration in TR 26.955, Scenario 1 and 2): A content provider offers content to an MNO in a well-defined contribution format and the MNO transcode the B2D video to stream this to several or many UEs. In an alternative the encoding may also be done externally of the MNO network and the encoded content is ingest into the 3GPP network. The workflow is aligned with 5G Media Streaming as defined in TS 26.501 and TS26.512. The following high-level constraints apply

* Cloud encoding, real time encoding for live, and possibly offline for VOD content
* Encoding latency constraints relaxed
* Media can be packaged into a streaming formats, in particular CMAF/DASH/HLS
* Random Access and switching bitrates is needed (IDR frames/RAP)
* Content can originate from different sources, for example may be user generated, may be produced by game engines, or may be professionally produced.
* Well defined source formats exist that can be described in manifests

UE-to-UE Communication (No matching scenario on TR 26.955): A UE to UE communication between two B2D UEs to permit improved user experiences. The workflow is aligned with Real-time Communication as defined in TS 26.506, TS 26.113 and TS 26.114.

* Typically content is addressing a communication scenario, for example heads.
* Real-time encoding on typical mobile device is possible.
* Latency constraints stringent to meet communication latency requirements
* Uplink and downlink provide sufficient bitrates over 5G network
* Media can be packaged into a communication formats, in particular RTP
* Expected criteria from above

Other types of communication (No matching scenario on TR 26.955): TBD

Online Gaming/Split Rendering (aligned with consideration in TR 26.955, Scenario 5): A B2D UE uses a cloud rendering/game engine to produce B2D formats to provide improved experiences. In case the user interacts with the content (body pose and game interaction), the encoding and delivery requirements are such that you need to meet the latency requirements. This workflow is aligned with TR 26.565. The following high-level constraints apply:

* Content generated in game engines
* encoding in cloud/edge servers
* Latency constraints very stringent
* Media can be packaged into a Split Rendering formats
* Expected criteria from above

The scenarios may be further refined and detailed by addressing details, for example on the B2DV formats to be evaluated.

# Evaluation and Characterization Framework – Basic Principles

Based on the above discussion, an evaluation of a scenario requires the definition of a evaluation and characterization framework, aligned with the framework in TR 26.955.

An evaluation framework allows to identify at least the following aspects for a technology under evaluation:

1. One or multiple meaningful quality metrics of the scenario and the technology under evaluation for different configurations to determine adequate quality thresholds under typical application constraints. The evaluation configuration needs to take into account restrictions in terms of encoding complexity, latency, and/or other functional requirements, such as random access.
2. The network requirements to delivery such content, primarily the resulting required bitrates can be determined.
3. The packaging requirements in order to deliver the data in interoperable manner can be determined.

Once such an evaluation framework is in place, the framework may also be used for

1. To determine the quality/network parameters for existing 3GPP technologies – referred to as anchors
2. To determine the quality/network parameters for new technologies - referred to as technologies under test
3. Typically, for each of the above not a single configuration is tested, but a tuple (for example to obtain quality rate curves)
4. To compare anchors with technologies under test using the results of the tuples.

The evaluation framework is documented in the figure below and follows the principles as defined in TR 26.955.



Based on all of the results for anchors and tests, technologies may be compared in a characterization framework as shown below



# Evaluation Framework for B2D Delivery

Based on the discussion in clause 2, 3 and 4, an adjusted common evaluation framework for Beyond 2D is presented following Figure 5.1-1.

For evaluating a scenario, it is expected that the following information is present:

* Representative *Reference Sequences* that reflect the scenario are collected and stored in a well defined B2DV format.
* A video encoding system takes the the B2DV reference sequences and produces one or more bitstreams together with metadata which is packaged into a delivery format.
	+ The actual packaging may not be done, but a "pseudo"-packing is applied that allows to interface with the video decoder of the system, but also allows to evaluate a bitrate determination.
	+ Packaging may for example be just to assume interleaving of samples from different video streams, adding relevant metadata that is not included in the bitstream, etc.
	+ A well-defined format should be defined to estimate the resulting bitrate.
* The data is then unpackaged such that it can be made available to the B2D video decoder in its appropriate format. The B2D video decoder input format is out of scope for the evalution. The B2D video decoder reconstructs at the output the a signal in B2DV format again. The sequences may also be inspected subjectively on a 2D plane or in a device that supports B2D rendering.
* By full-reference evaluation and by using the B2DV packed streams, metrics are generated.



**Figure 5.1-1 High-Level evaluation framework**

# Proposed Next Steps

The following next steps are proposed:

* Create a repository at 5G-MAG to host scripts and evaluation data for Beyond2D video 🡺 Qualcomm will check with 5G-MAG to do so
* Define a common annotation format for B2DV reference sequences based on the json format defined in TS 26.955 extending to address this 🡺 Philips is committed to do so by extending
	+ Schema: https://dash-large-files.akamaized.net/WAVE/3GPP/5GVideo/ReferenceSequences/raw-schema.json
	+ Example: <https://dash-large-files.akamaized.net/WAVE/3GPP/5GVideo/ReferenceSequences/TextMixTransitions-FullHD-10bit/TextMixTransitions-FullHD-10bit.json>
	+ Foreseen Extensions needed in json
		- YUV for each view and component
		- Pre-view: Packaged View
		- Thumbnail: Packaged View
		- Depth (optional)
		- Additional Metadata

NOTE: different information may be needed depending on the format considered

* Define a common environment for reference sequences
	+ <https://dash-large-files.akamaized.net/WAVE/3GPP/Beyond2D/ReferenceSequences/>
* Define a common annotation format for B2DV packaged data 🡺 Qualcomm is committed to do so
* Define a common metrics computation environment based on the TR 26.955 frame using HDRTools as well as the experiences collected in MPEG WG04 and AG05 🡺 Qualcomm and Philips are committed to work on this

NOTE: different information may be needed depending on the format considered

* Develop a common metrics computation based on the definition.

Then for each scenario, address the following

* Collect test sequences that properly reflect the typical content and sequence parameters available in the scenario.
* Select for each scenario a set of reference sequences, as well as a set of metrics to be used.
* Identify for each scenario if an anchor in 3GPP exists and if anchor data can be produced.

# Proposal

Based on the discussion it is proposed to:

* Agree the reference model with the description in clause 2
* Agree the considerations in clause 3 on the scenarios as starting point
* Agree the basic principles of the evaluation and characterisation in clause 4
* Agree on the principle evaluation framework in clause 5
* Agree on the proposed next steps in clause 6
* Add the relevant information to TR 26.956
* Add the relevant information to the Permanent document