**3GPP TSG-SA WG4 Meeting #128 S4-240916**

**Korea, Jeju, 20 – 24 May 2024** revision of S4aV240011

Agenda item: 9.9

Source: Qualcomm Incorporated

**Title:** [FS\_Beyond2D] Scenarios: Messaging

**Document for Discussion and** Agreement

# Introduction

During SA4#127-bis-e several scenarios for the Beyond 2D study have been proposed. We observe that all of the scenarios are quite complex and attempt to address very specific deployment scenarios. Based on document S4-240831 and S4-240915, this document provides a messaging exchange scenario taking into account the high-level proposals in the mentioned two docs.

This document provides some updates based on the comments received during the telco on May 7, 2024.

|  |  |  |  |
| --- | --- | --- | --- |
| **[S4aV240011](https://www.3gpp.org/ftp/TSG_SA/WG4_CODEC/3GPP_SA4_AHOC_MTGs/SA4_VIDEO/Docs/S4aV240011.zip)** | [FS\_Beyond2D] Messaging Scenario | Qualcomm Germany | Thomas Stockhammer |

**Presenter**: Thomas Stockhammer

**Online Discussion**:

* Jiayi: Propose to have 17 as QP as well
  + Thomas: No problem
* Gaelle: Not ok with the text in 4.X
  + Thomas: can we revise, it can revise to VOPS
* Rufael: Is this generalized evaluation or specific to the scenario?
  + Thomas: It is general, but some aspects are specific specific
* Serhan: Should not say “messaging is focus”
  + Thomas: ok
* Serhan: What is market relevance?
  + Thomas: Re-used from study item description
* Gaelle: What about UE coding
  + Thomas: scenario is whatsapp encoding, no latency, but real-time
* Gilles summarizes the comments.

**Decision**:

* Expect a revision based on the comments

[S4aV240011](https://www.3gpp.org/ftp/TSG_SA/WG4_CODEC/3GPP_SA4_AHOC_MTGs/SA4_VIDEO/Docs/S4aV240011.zip) is **noted**.

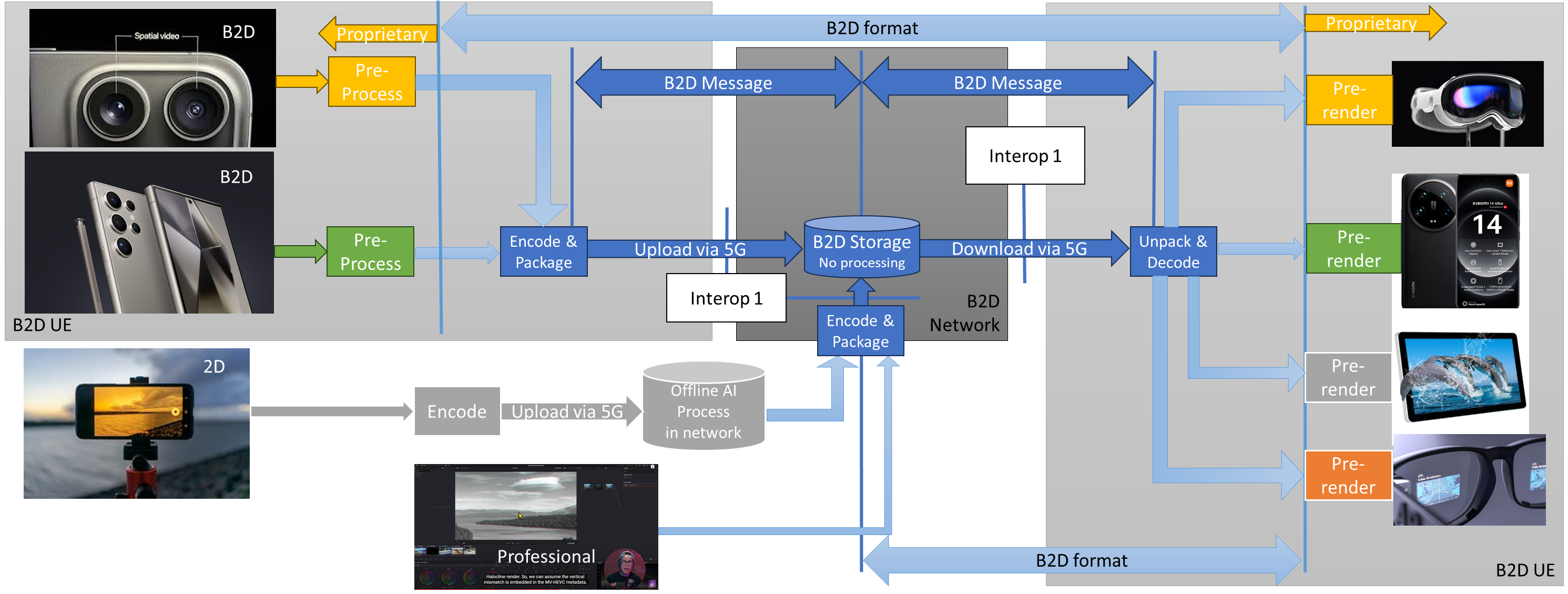
# Scenario Introduction: UE-to-UE message exchange

## 2.1 Introduction

In a first scenario and likely the most important one, existing and emerging capturing systems on mobile devices are used to capture video formats that go beyond regular 2D video formats. The captured scene is recorded and shared with another UE, for example another mobile device with a 2D screen, a mobile device with an Autostereoscopic Display, a VR or MR HMD, or it consumed on AR glass. While sharing may be done by different means (upload and download, stream, communication), the initial focus is on message exchange.

The scenario is shown in the below figure. On the upper part, two examples of UE camera setup are shown that allow to generate a B2D message on the device. The input the encoder and packager is the result of a proprietary capturing towards a well-defined B2D format. The B2D message is shated with the network that stores B2D messages in a well-defined format. On the lower end, a regular 2D UE may produce content and upload the content to a network server. The server processes the data and produces again a B2D format that can be encoded and packaged to meet a well-defined B2D formats. In yet another scenario, some professional content is produced and exported in a well-defined B2D format, that is then encoded and packaged and also made available as B2D message.

Any of these messages can then be accessed/downloaded by B2D UEs that can unpack and decode in the included data and provide the B2D formatted content to the proprietary rendering systems.



From the scenario, the B2D format and the encoding and packaging is expected to meet a few criteria, namely:

* The B2D delivery streams/formats can be unpacked and decoded on *market-relevant* UEs
* The B2D formats can provide a B2D experience on *market-relevant* UEs
* The B2D delivery streams/formats can be delivered via a 5G network in the upload and/or download.
* As a benefit, as not all devices supported B2D streams/formats, the B2D B2D streams/formats can be decoded and processed by legacy UEs without B2D capabilities, for example by presenting only a single view.

## Some Core Questions

Based on the introduction above, some follow-up questions require more detailed analysis

* What are B2D source formats of market relevance?
* What beyond 2D experiences can be created on market-relevant devices with B2D formats?
* What is a formalized description of the B2D formats?
* Do we have example sequences for such B2D formats?
* How can we encode and package the source formats?
* What are encoding restrictions to be considered in the evaluation?
* What are the resulting bitrates?
* What are suitable technologies for compression? Multiple HEVC streams, SbS/TaB, MV-HEVC, MIV?
* What are the packaging options and requirements?
* What are good metrics to compare the different technologies?
* Do we have to consider any “backward-compatibility” aspects?

## 2.3 B2D Devices and Formats

While it is understood that there is currently no harmonized set of formats for production and or device playback, the attempt towards improved format compatiblity and is the core driver for global standards. 3GPP should definitely contribute to this as we already do for the VOPS case with stereo MV-HEVC.

Note that devices are expected to render B2D formats based on their implementation, i.e. there is no need for interoperability or requirements beyond the decoding and the provisioning of metadata. The issue of rendering and presenting should not be part of the discussion. However, it can be checked how existing devices can render such formats in a beyond 2D experiences and invite interested companies to do so and bring data.

In order to understand the B2D formats, here are some data points

* <https://techcrunch.com/2023/12/11/apple-releases-spatial-video-recording-on-iphone-15-pro/>
  + Spatial Video if 1080p at 30fps
* <https://9to5mac.com/2024/01/04/will-the-iphone-16-be-able-to-record-4k-spatial-video/>
  + Spatial Video with 4K may come
* <https://appleinsider.com/articles/24/03/06/capturing-spatial-video-apple-vision-pro-vs-iphone-15-pro>
  + The spatial video captured is in a square 1:1 format at 2200 pixels by 2200 pixels. It is a near-perfect recreation of the passthrough viewed by the user.
* Some screenshots from [video](https://developer.apple.com/videos/play/wwdc2023/10071):

A screenshot of a video recording

Description automatically generated

* <https://developer.apple.com/documentation/avfoundation/avoutputsettingspreset/4172596-mvhevc1440x1440>
* Once capture stereo on iPhone 15 and offline postprocess to acquire accompanying depth (using [ZoeDepth](https://github.com/isl-org/ZoeDepth" \o "https://github.com/isl-org/zoedepth" \t "_blank) or similar).
* Meta Quest can record spatial video: <https://360rumors.com/quest-3-3d-videos/>
  + After recording, you’ll see that the video or photo will be captured in side by side format, with a square aspect ratio. Photos will also be side by side but they are stretched vertically, and need to be edited to fix that.
* <https://deovr.com/blog/84-record-vr-footage-on-the-meta-quest-3>
  + The Quest 3 features two cameras that deliver full-color passthrough, allowing users to record content in 4K (2k per eye), using the Meta Quest Developer HUB.
  + The Quest 3's passthrough cameras record footage that is flat 120-100 (possibly 90) degrees.

Based on all of the observations, the following baseline formats may be considered:

* Stereo views for left and right eye
* Resolution per eye: 1080x1080, 1440x1440, 2048x2048
* Frame rates: 30, 50, 60 fps
* Chroma format: Y’CbCr (RGB?)
* Chroma subsampling: 4:2:0 (4:4:4)
* Bit depth: 10
* Colour space formats: BT.709, BT.2020
* Transfer characteristics BT.709, BT.2100 (HDR)

Beyond the basic information, the following information may be available:

* Static metadata:
  + Relative position between cameras
  + Camera intrinsics
  + Line time (per camera)
  + Examples: https://github.com/MPEGGroup/FileFormatConformance/tree/m62054\_exintrinsics/data/file\_features/under\_consideration
* Disparity/depth map: 10bit, same resolution as source content, monochrome

A few spatial video sequences are provided here: <https://www.zdnet.com/article/download-my-spatial-videos-from-tech-events-and-view-them-on-quest-3-or-vision-pro/>. Also here some information is provided on how to capture spatial video: https://www.zdnet.com/article/how-to-capture-spatial-video-with-the-iphone-15-pro-theres-a-trick/

## 2.4 Potential Criteria for Market relevance of formats

For formats produced on devices:

* Are there mainstream devices that can capture and produce such content?
* Can we capture test sequences with existing devices?
* Can the formats be rendered on market-relevant devices?

## 2.5 From Scenario to Evaluation and Characterization

### 2.1 Background

Based on the above discussion, an evaluation of the 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 technology under evaluation 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.
3. The packaging requirements in order to deliver the data in interoperable manner.

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.

Diagram

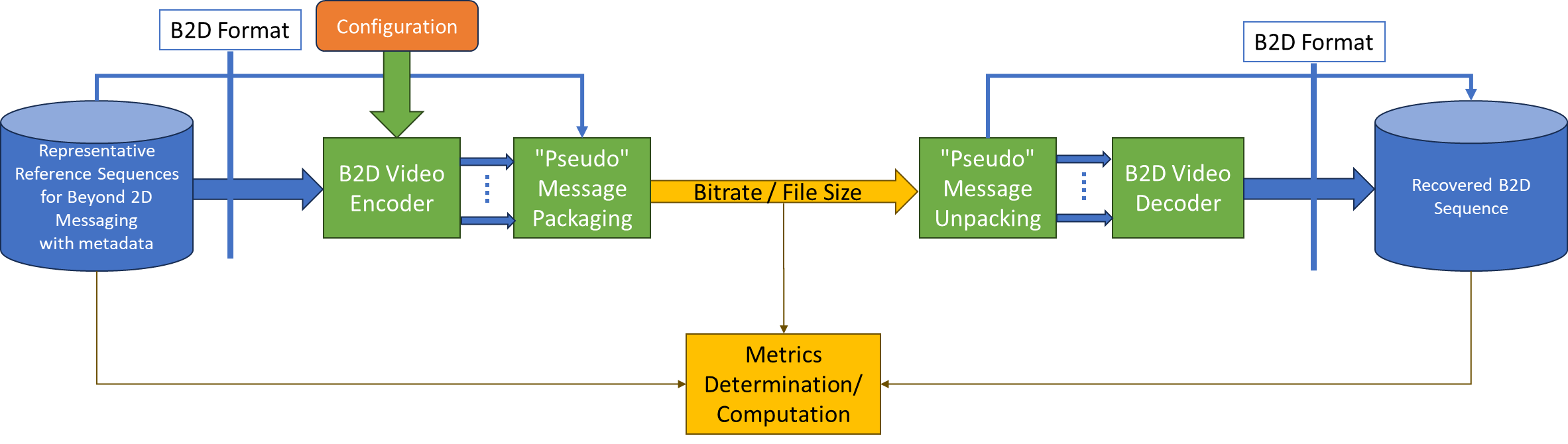
Description automatically generated

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

A diagram of a diagram

Description automatically generated

### 2.2 Evaluation Framework got B2D Messaging

A specific evalation framework for the B2D messaging is presented in the following. Representative Reference sequences are collected and stored in a well defined B2D format. For a video encoder, the a configuration is provided that matches the application constraints. The resulting video streams are “pseudo”-packaged in order to determine the file size/bitrate. The data is then unpackaged, and a B2D video decoder is used to reconstruct data in the B2D format again. The data is stored. The original sequence and the recovered sequence are used determine metrics. The sequences may also be inspected subjectively. 

## 2.6 Codec Technologies

In order to provide B2D experiences, the following candidate technologies aligned with existing 3GPP codecs may be considered for the messaging scenario.

* Multiple HEVC Streams (+ depth + alpha + metadata)
* SbS and TaB formats (+ depth)
* Stereo MV-HEVC (+ depth)
* 3D MV-HEVC (+depth)
* MIV + HEVC

More details on configurations of different codecs are needed.

## 2.7 Reference Sequences

Reference sequences options:

* Produce with iPhone15 max pro è volunteers?
* Produce with Meta Quest 3 è volunteers?
* Use existing data sets from MPEG/JVET è volunteers?
* Produce exported data sets from professional tools è volunteers?

Proposed annotation

* Follow TR 26.955
  + 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>
  + Extensions needed in json
    - YUV of left and right eye
    - Pre-view: Packaged Left & Right Eye & Depth
    - Thumbnail: Packaged Left & Right Eye & Depth
    - Depth sequence
    - Additional Metadata

Proposed storage location:

* <https://dash-large-files.akamaized.net/WAVE/3GPP/Beyond2D/ReferenceSequences/>

After communication with AG05 chair, it is clear that reference and test sequences are scarce, and also provide an indication for market relevance of the technology. In particular it should be avoided that test and reference sequences are used, that are already used in the the development of codecs. He also encouraged 3GPP as a potential customer of MPEG technologies in this domain should produce relevant reference sequences and use those in “independent tests”.

## 2.8 Metrics

In general, the metrics used in TR 26.955 can be re-used, weighting two different views.

Other metrics that may be considered:

* IV-PSNR: <https://gitlab.com/mpeg-i-visual/ivpsnr> and here: <https://ieeexplore.ieee.org/document/9785987>
* Others?
* The conversation with the AG05 chair revealed that classical pixel-domain metrics in the above setup can be used and provide meaningful results. He encouraged to define a “mapping” to a 2D plane and on this plane the metric is applied. He will provide more information.

## 2.9 Encoding Constraints

See TR 26.955 for Messaging applications.

Messaging (aligned with consideration in TR 26.955, Scenario 4): A UE generates B2D content in real-time and shares the media content with another UE 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 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
* Expected criteria from above

The AG05 chair also indicated that encoding constraints should not be over optimized (for example using different “distortion”-levels for different eyes), but straightforward settings help.

## 2.10 Interoperability and Packaging

Encapsulation into ISO BMFF including all metadata.

# Proposal

Based on the discussion it is proposed to:

* Adopt the scenario
* Create the relevant information in the PD to track the status
  + A text proposal will be provided if agreed in principle
* Add the proposed input in clause 4 to the draft TR
* Collect information related to test sequences, metrics and so on
  + Initiate a repo here: <https://dash-large-files.akamaized.net/WAVE/3GPP/Beyond2D> collecting equivalent information as for TR 26.955

# Proposed Input to TR 26.9xx (based on TR 26.955)

Based on the above considerations, this documents the potential additions to the TR using a structure as defined in TR 26.955.

## ===== CHANGE =====

# 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".

[26140] 3GPP TS 26.140: "Multimedia Messaging Service (MMS); Media formats and codecs".

[26143] 3GPP TS 26.143: "Messaging Media profiles".

[26966] 3GPP TR 26.966: " Evaluation of new HEVC coding tools".

## ===== CHANGE =====

## ===== CHANGE =====

# 5 Evaluation and Characterization Framework

## 5.1 Overview

Generally, the test and characterization framework as documented in TR 26.955, clause 5 also applies to this document. This clause only documents differences and extensions that are needed for beyond 2D Evaluation and characterization framework.

The overview of the evaluation framework for the B2D messaging is presented in Figure 5.1-1. Representative reference sequences are collected and stored in a well defined B2D format. For a video encoder, the a configuration is provided that matches the application constraints. The resulting video streams are “pseudo”-packaged in order to determine the file size/bitrate. The data is then unpackaged, and a B2D video decoder is used to reconstruct data in the B2D format again. The data is stored. The original sequence and the recovered sequence are used determine metrics. The sequences may also be inspected subjectively.

A diagram of a computer

Description automatically generated

Figure 5.1-1 B2D Evaluation framework

## 5.2 Reference Sequences

This document provides reference sequences that are used to generate anchors and are also made available in order to generate test bitstreams for other codecs. Reference sequences are selected to be representative for a scenario.

Reference sequences are described in Annex C of this document along with their properties and their licenses. A format for raw reference sequences based on a JSON schema is defined in clause B.2.

Annex D describes how to upload new proposed reference sequences and how to download the reference sequences.

## 5.3 Reference Software Tools

Editor’s Note: For further study

## 5.4 Metrics

Editor’s Note: For further study

The metrics in clause 5.5 of TR 26.955 also apply for this report.

In addition, the following is defined Editor’s Note: For further study

For metrics reporting, the following csv scheme is defined: tbd

## 5.5 Encoding Constraints

The encoding constraint definition in clause 5.6 of TR 26.955 also apply for this report.

In addition, the following is defined:

* *Equal Quality Views*: equal quality views refers to the encoding such that each view when decoded has the same quality target, typically applying the same QP.

Editor’s Note: More details are to be defined

## ===== CHANGE =====

## 6.X Scenario X: Beyond 2D Messaging and Social Sharing

### 6.X.1 Motivation

In TR 26.955, clause 6.5 introduces messaging and social scenario. It addressed existing standardized and proprietary messaging services. Messaging applications, especially on the upstream, continue to become a critical part of the mobile experience, replacing old style text messaging, and increasingly are video-based. Four of the top 20 applications on the upstream are messaging apps. In another activity, operators are promoting enhanced messaging services based on Rich Communication Services (RCS).

As summarized in TR 26.955, sharing and uploading content as part of social sharing is predominantly pictures and videos that uploaded directly into a cloud and uploaded to one or many social networks, and then discussed (or shared again) over messaging networks. The relevant quality-of-experience factors include the quality of shared content, the time it takes to upload, the costs associated with the upload and also the processing and battery consumption requirements on the device to prepare the content for upload.

In the context of beyond 2D video experiences, existing and emerging capturing systems on mobile devices are used to capture video formats that go beyond regular 2D video formats. The captured scene is recorded and shared with another UE, for example another mobile device with a 2D screen, a mobile device with an Autostereoscopic Display, a VR or MR HMD, or it consumed on AR glass. While sharing may be done by different means (upload and download, stream, communication), the initial focus is on message exchange.

The scenario is shown in the below figure.

Editor’s Note: Diagram will be aligned with 1064 diagram

On the upper part, two examples of UE camera setup are shown that allow to generate a B2D message on the device. The input the encoder and packager is the result of a proprietary capturing towards a well-defined B2D format. The B2D message is shated with the network that stores B2D messages in a well-defined format. On the lower end, a regular 2D UE may produce content and upload the content to a network server. The server processes the data and produces again a B2D format that can be encoded and packaged to meet a well-defined B2D formats. In yet another scenario, some professional content is produced and exported in a well-defined B2D format, that is then encoded and packaged and also made available as B2D message.

Any of these messages can then be accessed/downloaded by B2D UEs that can unpack and decode in the included data and provide the B2D formatted content to the proprietary rendering systems.

A computer screen shot of a diagram

Description automatically generated

From the scenario, the B2D format and the encoding and packaging is expected to meet a few criteria, namely:

* The B2D format can be generated with market-relevant UEs
* The B2D messages can be encoded and packaged with market-relevant UEs
* The B2D messages can be unpacked and decoded on market-relevant UEs
* The B2D formats can provide a B2D experience on market-relevant UEs
* The B2D messages can be delivered via a 5G network in the upload and/or download.

Editor’s Note: On the typical and market-relevant formats, some information from clause 2.3 may be used.

### 6.X.2 Description of the Anticipated Application

In the context of 3GPP services, the Media Messaging Profiles and the architecture and the capabilities in TS 26.143 provides the appropriate context.

Editor’s Note: Architecture and capabilities from TS 26.143 may be added here

The considered scenario is the UE-based generation and consumption of messages and encapsulating into ISO/BMFF container format. Important aspects that are expected to be considered when evaluating a codec in the context of this:

- Quality and Coding Efficiency:

- The ability to compress a B2D video sequence targeting the maximum file size and maintaining high quality.

- The ability to compress 2 or more B2D streams in real-time to minimize storage requirements and sending latency.

- Considered settings for encoding:

- Regular random access at least every 2 seconds, preferably more often

- No specific encoding latency constraints are applicable

- Encoding in this scenario is typically done as

- Real-time encoding for social sharing

- UE-based Encoding

### 6.X.3 B2D Messaging Video Formats

Table 6.X.3-1 provides an overview of typical beyond 2D source signal properties for UE-to-UE messaging. This information is used to select proper test sequences.

Editor’s Note: This is a first draft, details needs to be confirmed.

Table 6.5.3-1 Beyond 2D Source Format Properties for UE-to-UE messaging

|  |  |
| --- | --- |
| Source format properties | Social Sharing |
| Number of views | 2 |
| Spatial resolution for each view | 2160 x 2160, 1440 x 1440, 1080 x 1080 |
| Chroma format | Y’CbCr |
| Chroma subsampling | 4:2:0 |
| Picture aspect ratio | 1:1 |
| Frame rates | 30, 50, 60 Hz |
| Bit depth | 10 |
| Colour space formats | BT.709, BT.2020 |
| Transfer characteristics | BT.709, BT.2100 (HDR) |
| **Optional Depth/Disparity** | |
| Spatial Resolution | Same resolution as view |
| Chroma format | Y’CbCr |
| Chroma subsampling | 4:0:0 |
| Picture aspect ratio | Same as views |
| Frame rates | Same as views |
| Bit depth | 10 |
| Format | **tbd** |

### 6.x.4 Encoding and Decoding Constraints

Table 6.x.4-1 provides an overview of encoding and decoding constraints for H.265/HEVC for UE-to-UE messaging scenario. This information supports the definition of detailed anchor conditions.

Editor’s Note: This is a starting point and additional considerations need to apply.

Table 6.5.4-1 Encoding and Decoding Configurations

|  |  |  |
| --- | --- | --- |
| Encoding and Decoding Constraints | General | H.265/HEVC |
| Relevant Codec and Codec Profile/Levels | Profile suitable for messaging content, no specific requirements.  Levels to meet the above formats | H.265/HEVC Main 10 Profile  Level 4.1, 5.1 |
| Random access frequency | 1 second and 10 seconds | 1 and 10 seconds |
| Bit rates and quality configuration | Capped-VBR (social sharing) and VBR (messaging)  Fixed QP | B = {2.5, 5, 7.5,10} Mbps  Capped-VBR (social sharing) and VBR (messaging)  Fixed QP |
| Bit rate parameters (CBR, VBR, CAE, HRD parameters) | Covering a range of relevant bitrates and qualities | No latency requirements beyond RAP so picture reordering allowed |
| Latency requirements and specific encoding settings | No latency requirements | No specific requirements |
| Encoding complexity context | real-time encoding (social sharing), offline encoding (messaging) on mobile device, single path | see General |
| Required decoding capabilities | Profile suitable for messaging content, no specific requirements.  Levels to meet the above formats | Up to x streams of H.265/HEVC Main 10 Profile  Level 4.1, 5.1 |

### 6.5.5 Performance Metrics

The following metrics are expected to be reported <tbd>

### 6.X.6 Interoperability Considerations

Messaging applications require that the content is included in a packaging and file format.

Editor’s Note: More details on backward-compatibility being a benefit

### 6.X.7 Reference Sequences

Table 6.X.7-1 provides the selected reference sequences for this scenario. Keys are identified to refer to the sequences in the context of the scenario. The sequences are named and a reference to the details of the sequence is provided. A justification is provided, why this sequence is selected.

<Editor’s Note: We should have 1080 sequences and 2K sequences, and we should have sequences with depth available>

Table 6.X.7-1 Reference Sequences for UE-to-UE Sharing

|  |  |  |  |
| --- | --- | --- | --- |
| Key | Name | Reference | Justification/Comment |
| SX-R01 |  | Annex C.5.2 |  |
| SX-R02 |  | Annex C.5.3 |  |
| SX-R03 |  | Annex C.5.4 |  |
| SX-R04 |  | Annex C.5.5 |  |

### 6.X.8 Anchor Definition

#### 6.X.8.1 Overview

This clause provides details on how to generate the anchors for the social sharing and messaging scenario.

The Social sharing and messaging scenario relies on relaxed delay encoding modes with limited encoding complexity.

#### 6.X.8.3 Anchors

Editor’s Note: This serves as a template. Details are to checked.

##### 6.X.8.3.1 Overview

Table 6.X.8.3.1-1 provides an overview of the anchor tuples. Keys are identified to refer to the anchors in the context of the scenario.

The details are also provided here: https://dash-large-files.akamaized.net/WAVE/3GPP/Beyond2D/Bitstreams/Scenario-X-Sharing/265/streams.csv.

Table 6.5.8.3.1-1 Anchor Tuple generation with H.265/HEVC for Social sharing and messaging

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Key | Clause | Reference Sequence | Reference Encoder | Config | Variations | Anchor Key |
| SX-A01-265 | 6.X.8.3.3 | SX-R01 | HM16.22 | SX-HM-01 | QP = [17, 22,27,32,37] | SX-A01-265-<QP> |
| SX-A02-265 | 6.X.8.3.3 | SX-R02 | HM16.22 | SX-HM-01 | QP = [17, 22,27,32,37] | SX-A02-265-<QP> |
| SX-A03-265 | 6.X.8.3.4 | SX-R03 | HM16.22 | SX-HM-01 | QP = [17, 22,27,32,37] | SX-A03-265-<QP> |
| SX-A04-265 | 6.X.8.3.4 | SX-R04 | HM16.22 | SX-HM-01 | QP = [17, 22,27,32,37] | SX-A04-265-<QP> |
| SX-A05-265 | 6.X.8.3.5 | SX-R01 | HM16.22 | SX-HM-02 | QP = [17, 22,27,32,37] | SX-A05-265-<QP> |
| SX-A06-265 | 6.X.8.3.5 | SX-R02 | HM16.22 | SX-HM-02 | QP = [17, 22,27,32,37] | SX-A06-265-<QP> |
| SX-A07-265 | 6.X.8.3.6 | SX-R03 | HM16.22 | SX-HM-02 | QP = [17, 22,27,32,37] | SX-A07-265-<QP> |
| SX-A08-265 | 6.X.8.3.6 | SX-R04 | HM16.22 | SX-HM-02 | QP = [17, 22,27,32,37] | SX-A08-265-<QP> |

##### 6.X.8.3.2 Common Parameters

Editor’s Note: This is a starting point, details are to be checked.

To generate the anchor bitstreams, HM16.22 is used.

Each source sequence is encoded with:

- QP: [17, 22, 27, 32, 37]

- InternalBitDepth is 10 # codec operating bit-depth where all sequences (including 8 bit sequences) are coded with an internal bitdeph of 10 in accordance with [44] and metrics are calculated in 10 bits.

- SEIDecodedPictureHash=1

As the SEIDecodedPictureHash is set to 1, the effective file size (EFS) needs to take into account the removal of this SEI message when computing the bitrate metric as defined in clause 5.5.2.

##### 6.X.8.3.3 SX-HM-01: no random access

Each source sequence is encoded with the following changes:

- The common parameters as defined in clause 6.X.8.3.2

- IntraPeriod with no fix interval

- GOPSize is equal to 8. Each B picture refers to up to 4 preceding pictures in display order within the GOP

- IntraQPOffset is -1. B picture QP offsets are IntraQPOffset is -1. B picture QP offsets are adjusted based on the base QP and on the QPmod, QPoffset, QPOffsetModelScale and QPOffsetModelOff.

The settings are defined in the attached configuration file sX-hm-01.cfg.

##### 6.X.8.3.4 SX-HM-02: Intra

Each source sequence is encoded with the following changes:

- The common parameters as defined in clause 6.6.8.3.3

- IntraPeriod = power of 2 value that is greater than or equal to the frame rate (fps), such that near 1 second is achieved: 32 for 30fps sequences and 64 for 60fps sequences

- DecodingRefreshType: (2) IDR

- IntraQPOffset and B pictures QPoffsets are set equal to 0

- Each B picture refers to immediately preceding pictures in display order.

The settings are defined in the attached configuration file sX-hm-02.cfg.

### 6.X.9 Anchor Results

## ===== CHANGE =====

Annex B: Data Formats and Metrics

## ===== CHANGE =====

Annex C: Reference Sequences

# C.1 Introduction

This annex provides a summary of candidate reference sequences that where discussed to be potentially suitable for one or multiple of the scenarios introduced in clause 6 of this Technical Report. For each candidate reference sequence, at least the following information is provided.

- A summary of the sequence characteristics

- A screenshot of the sequence

- Source sequence properties

- Information where the source sequence is hosted

- Copyright and license information

The content is provided in JSON files here: [https://dash-large-files.akamaized.net/WAVE/3GPP/Beyond2D/ReferenceSequence](https://dash-large-files.akamaized.net/WAVE/3GPP/5GVideo/Beyond2D). The format of the reference sequences follows the proposed format in Annex B.2.

The sequences are summarized here: https://dash-large-files.akamaized.net/WAVE/3GPP/Beyond2D/ReferenceSequences/sequences.csv.