**3GPP TSG-SA4 Meeting #129-e *S4-241652***

**Online, , 19th Aug 2024 - 23rd Aug 2024 revision of S4-241476**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *CR-Form-v12.3* | | | | | | | | |
| **CHANGE REQUEST** | | | | | | | | |
|  | | | | | | | | |
|  | **26.841** | **CR** | **pseudo** | **rev** | **1** | **Current version:** | **0.2.0** |  |
|  | | | | | | | | |
| *For* [***HE******LP***](http://www.3gpp.org/3G_Specs/CRs.htm#_blank)*on using this form: comprehensive instructions can be found at* [*http://www.3gpp.org/Change-Requests*](http://www.3gpp.org/Change-Requests)*.* | | | | | | | | |
|  | | | | | | | | |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Proposed change affects:*** | UICC apps |  | ME | **X** | Radio Access Network |  | Core Network | **X** |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | | | | |
| ***Title:*** | [FS\_MeMe] 3D Video Messaging Format | | | | | | | | | |
|  |  | | | | | | | | | |
| ***Source to WG:*** | Qualcomm Germany | | | | | | | | | |
| ***Source to TSG:*** |  | | | | | | | | | |
|  |  | | | | | | | | | |
| ***Work item code:*** | FS\_MeMe | | | | |  | ***Date:*** | | | 2024-08-12 |
|  |  | | | |  | |  | | |  |
| ***Category:*** | **B** |  | | | | | ***Release:*** | | | Rel-19 |
|  | *Use one of the following categories:* ***F*** *(correction)* ***A*** *(mirror corresponding to a change in an earlier release)* ***B*** *(addition of feature),* ***C*** *(functional modification of feature)* ***D*** *(editorial modification)*  Detailed explanations of the above categories can be found in 3GPP [TR 21.900](http://www.3gpp.org/ftp/Specs/html-info/21900.htm). | | | | | | | | *Use one of the following releases: Rel-8 (Release 8) Rel-9 (Release 9) Rel-10 (Release 10) Rel-11 (Release 11) … Rel-17 (Release 17) Rel-18 (Release 18) Rel-19 (Release 19)  Rel-20 (Release 20)* | |
|  |  | | | | | | | | | |
| ***Reason for change:*** | | The study item asks for advanced media formats  E) **Additional media experiences**: In 3GPP, advanced media experiences are added to different services, in particular immersive media, new image formats, interactive content, etc. The proper integration of new media experiences should be studied. | | | | | | | | |
|  | |  | | | | | | | | |
| ***Summary of change:*** | | This CR studies the integration of 3D video messages into 3GPP Media Messaging. | | | | | | | | |
|  | |  | | | | | | | | |
| ***Consequences if not approved:*** | |  | | | | | | | | |
|  | |  | | | | | | | | |
| ***Clauses affected:*** | | 2, 5.5 | | | | | | | | |
|  | |  | | | | | | | | |
|  | | **Y** | **N** |  | | | |  | | |
| ***Other specs*** | |  | **X** | Other core specifications | | | | TS/TR ... CR ... | | |
| ***affected:*** | |  | **X** | Test specifications | | | | TS/TR ... CR ... | | |
| ***(show related CRs)*** | |  | **X** | O&M Specifications | | | | TS/TR ... CR ... | | |
|  | |  | | | | | | | | |
| ***Other comments:*** | | There are some dependendies on ongoing work in VOPS and FS\_Beyond2D, as well as in MPEG Systems. | | | | | | | | |
|  | |  | | | | | | | | |
| ***This CR's revision history:*** | |  | | | | | | | | |

## ===== 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".

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

[26857] 3GPP TR 26.857: "5G Media Service Enablers".

[RFC2045] IETF RFC 2045, "Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies", November 1996

[RFC2046] IETF RFC 2046, "Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types"

[RFC2387] IETF RFC 2387, "The MIME Multipart/Related Content-type"

[RFC6381] IETF RFC 6381, "The 'Codecs' and 'Profiles' Parameters for "Bucket" Media Types"

[14496-12] ISO/IEC 14496-12: "Information technology - Coding of audio-visual objects -Part 12: ISO base media file format".

[IETF-MIMI] IETF Draft draft-ietf-mimi-content-03: "More Instant Messaging Interoperability (MIMI) message content", April 2024.

[26265] 3GPP TS 26.265: "Video Capabilities and Operating Points".

[26956] 3GPP TR 26.956: "Evaluation and Characterization of Beyond 2D Video Formats and Codecs".

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

[AMVHEVC] Apple HEVC Stereo Video - Interoperability Profile (Beta), Version 0.9, June 21, 2023, <https://developer.apple.com/av-foundation/HEVC-Stereo-Video-Profile.pdf>

[SPATIAL] Mike Swanson, "Spatial Video", March 7 2024, https://blog.mikeswanson.com/spatial-video/

[COUNTOUR] Video Contour Map Payload, Version 0.9, June 21, 2023, <https://developer.apple.com/av-foundation/Video-Contour-Map-Metadata.pdf>

[26118] 3GPP TS 26.118: "Virtual Reality (VR) profiles for streaming applications".

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

## 5.5 Key Topic #5: Additional media experiences

### 5.5.1 Description

In 3GPP, advanced media experiences are added to different services, in particular immersive media, new image formats, interactive content, etc. The proper integration of new media experiences should be studied.

A specific format of interest is the ability to share 3D video with audio. Existing and emerging capturing systems on mobile devices are used to capture video formats that go beyond regular 2D video formats (B2D). 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 focus in the context of this document is message exchange.

The scenario is shown in figure 5.5.1. On the upper part, two examples of UE camera setup are shown that allow to generate a Beyond (B2D) message on the device. The input to the encoder and packager is the result of a proprietary capturing system that converts the capture to a well-defined B2D format. The B2D message is shared 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 conform to 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 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

Figure 5.5.1: Beyond 2D Messaging

Note that core video formats are expected to be defined in TS 26.265 [26265] and detailed discussions on the properties and benefits of such representation formats are provided in 3GPP TR 26.956 [26956].

While it is understood that there is currently no harmonized set of formats for production and device playback, the attempt towards improved format compatiblity and is the core driver for global standards. A core format for global 2D message formats is stereoscopic video. Stereoscopic video presents one image to the user’s left eye and another image (typically correlated) to the user’s right eye to produce the stereopsis effect, defined as "the perception of depth produced by the reception in the brain of visual stimuli from both eyes in combination; binocular vision." [AMVHEVC].

Different type of Beyond 2D video using projections to left and right eye may be used [SPATIAL]:

- rectangular, traditional 3D video

- spherically-projected 3D video similar as defined in 3GPP TS 26.118 [26118].

- either of the two may be extended with additional depth data, also referred to as video contour maps [CONTOUR].

3GPP decided based on the study documented in 3GPP TR 26.966 [26966] to support stereo MV-HEVC and the work in currently ongoing to add this format to 3GPP TS 26.265 [26265]. In addition, 3GPP has adopted also audio codecs that can augment 3D video formats including IVAS and MPEG-H audio, already included in 3GPP TS 26.143 [26143].

This issue addresses the analysis and requirements to support 3D video in advanced messaging based on MV-HEVC as defined in 3GPP TS 26.265 [26265].

### 5.5.2 Gap Analysis and Requirements

In order to identify the gaps to add 3D video to advanced messaging, a high-level flow chart is provided in Figure 5.5.2-1 and also eloborated further below based on the architecture shown in Figure 4.1-1 and the player model shown in Figure 4.3-1.

****

**Figure 5.5.2-1 High-Level Call flow for 3D video messaging**

The following high-level call flow is executed to address the 3D video messaging and gaps are indicated in **bold**:

1) A Message Service Sender instructs and MMBP generator to **generate a message with 3D format** where the format is defined through a set of **parameters handed over** on an MMBP-GEN-API.

2) The MMBP generator **creates an MMBP that includes 3D video formats encoded with MV-HEVC**, and adds relevant **static and dynamic metadata**.

3) The generated file with associated relevant static and dynamic metadata is provided in a container message that is uploaded to a messaging server, together with a **signaling in the media type and associated parameters**. The media signaling may include that the file is suitable for different clients, for example it includes a 2D and a Beyond 2D experience.

4) A messaging service client is informed about the availability of a message in this format using the media type signaling together with appropriate sub-parameters.

5) A messaging service client **uses the MMBP-PLAY-API to query the MMBP player with it can process, decode and render the message**, possibly in 2D legacy or as 3D video.

6) If the MMBP player indicates that it supports the capability, the messaging service clients selects and downloads the message and hands it to the MMBP player for playback.

7) The MMBP player **processes, decodes and renders the 3D message** based on its decoding and rendering capabilities **together with other media types.**

Based on the analysis in the call flow, the following aspects are missing:

a) the definition of the 3D formats that are supported

b) a generator for a 3D video message including API parameters

c) a 3D video messaging format including MV-HEVC with relevant metadata, including depth, alpha. projection formats, disparity information – for details refer to TR 26.956 [26956].

d) the signaling of the media type and the associated parameters, typicaly a MIME type and the codecs parameter.

e) an API to query the MMBP player that it can process, decoder and render the message

f) an MMBP player capability that processes, decodes and renders the message.

g) an MMBP player capability renders the video message together withother media components in a synchronized and spatially aligned matter.

### 5.5.3 Potential Solutions

#### 5.5.3.1 Overview

To address the issues a) – g), a set of solutions are discussed in the following clause.

#### 5.5.3.2 3D Video Formats

For the 3D video formats according to a), the parameters are defined in TR 26.956 [26956], including different type of Beyond 2D video using projections to left and right eye:

- rectangular, traditional 3D video

- spherically-projected 3D video similar as defined in 3GPP TS 26.118 [26118].

- either of the two may be extended with additional depth data, also referred to as video contour maps [CONTOUR].

In addition, the detailed signal properties of the video each eye needs to be defined:

- Sample aspect ratio for each eye

- Picture aspect ratio for each eye.

- Resolutions per eye

- Framerates for each eye are

- Signal characteristics, Chroma subsampling, bit depth, colour primaries, transfer characteristics, matrix coefficients

- Projection parameters: rectilinear, fisheye, equirectangular, Field-of-view and restricted coverage.

- Additional metadata such as hero eye, camera parameters, disparity adjustments, disparity/depth map, camera intrinsics, line times,

- hero eye: A value that indicates which eye is the primary eye when rendering in 2D.

- camera parameters:

- distance between the camera lens centers

- disparity adjustment:

- horizontal disparity adjustment, a value that indicates a relative shift of the left and right images, which changes the zero parallax plane.

NOTE: The parameters may be aligned with TS 26.118 [B]

- 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

Editor’s Note: a more detailed subset may be selected or inherited from TS 26.265 [26265] and TR 26.956 [26956].

#### 5.5.3.3 3D Video Encoding and Decoding capabilities

To address issue b) as well as partially e) and f), video encoding and decoding capabilities, as video operation points need to be defined. A candidate solution will be available in 3GPP TS 26.265 [26265].

The capabilities are expected to include abstract APIs to provide encoding parameters.

Editor’s note: It is expected that TS 26.265 defines encoding and decoding capabilities for 3D video.

#### 5.5.3.4 3D Video Systems Operation Point

To address the integration of the video elementary stream into a container message, together with the metadata, a comprehensive 3D track format is required.

ISO/IEC 14496-12:2022 (ISOBMFF) defines the multiplexed metadata track format ('mebx'), which is capable of carrying multiple metadata items over a time range in a single track.

Additional metadata signaling for 3D video operation points are expected to be completed in 3GPP TS 26.265 [26265].

Editor’s note: It is expected that TS 26.265 defineds a track format for 3D video.

#### 5.5.3.5 3D Video Media Type Signaling

One of the possibilities to carry MV-HEVC video in mp4 is by using the 'hvc1' or 'hev1' sample entry type as specified in clause 9 of ISO/IEC 14496-15 in a backward compatible manner. Some existing products in the market are using this concept to carry stereoscopic content and alpha using the Multiview extensions of the HEVC standard as L-HEVC in mp4. Such profiles with existing support in the mobile ecosystem are expected to be defined in TS 26.265 [26265]. However, when constructing the MIME types 'codecs' parameter, according to Annex E of ISO/IEC 14496-15, it does not provide the necessary signalling for all layers. Even if additional information can be specified with other MIME type parameters, these may not be processed by certain APIs. For example, the W3C API accepts a MIME type with no extra MIME parameters except for codecs.

Furthermore, the current signalling inside the codecs string does not expose other important information such as the types of auxiliary information related to rendering the stream. Given these challenges, it is important that a standardized solution be developed to enhance interoperability, accuracy, and efficiency of multi-layer video stream handling.The following high-level requirements for such a solution have been identified:

- Comprehensive Layer Signalling: Enable the signalling of multiple video layers to be used for 3GPP-based services, including DASH MPDs (see TS 26.511), capability checks in the context of 5G Media Streaming (see TS 26.511), as well as for Media Messaging Services (see TS 26.143). This should include but not be limited to the number of layers, types of each layer, their inter-dependencies, etc.

- Backward Compatibility: Ensure that the solution maintains backward compatibility, allowing existing players and systems to continue functioning without modifications, while enabling enhanced capabilities for updated systems.

- Adaptability and Extensibility: Design the solution to be adaptable for future extensions and new types of layers or enhancements without requiring significant overhauls. Consider making the signalling codec agnostic.

Editor’s note: It is expected that MPEG provides signaling on this matter to be added here.

#### 5.5.3.6 System integration

The 3D video needs to be synchronized and spatially aligned with other media types.

Time synchronization is provided by using the ISO BMFF/MP4 file format.

Editor’s note: It is expected that the coordinate system from TS 26.118 can be re-used.

### 5.5.4 Summary and Conclusions

Editor’s note: It is expected that all functionalities are available to support 3D video and general media experiences in a future messaging service.