|  |
| --- |
| 3GPP TR 26.956 V0.0.4(2024-08) |
| Technical Report |
| 3rd Generation Partnership Project;Technical Specification Group Services and System Aspects;Evaluation and Characterization of Beyond 2D Video Formats and Codecs(Release 19) |
|   |
| *5G-logo_175px* | 3GPP-logo_web |
|  |
| The present document has been developed within the 3rd Generation Partnership Project (3GPP TM) and may be further elaborated for the purposes of 3GPP.The present document has not been subject to any approval process by the 3GPPOrganizational Partners and shall not be implemented.This Specification is provided for future development work within 3GPPonly. The Organizational Partners accept no liability for any use of this Specification.Specifications and Reports for implementation of the 3GPP TM system should be obtained via the 3GPP Organizational Partners' Publications Offices. |

|  |
| --- |
|  |
| ***3GPP***Postal address3GPP support office address650 Route des Lucioles - Sophia AntipolisValbonne - FRANCETel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16Internethttp://www.3gpp.org |
| ***Copyright Notification***No part may be reproduced except as authorized by written permission.The copyright and the foregoing restriction extend to reproduction in all media.© 2024, 3GPP Organizational Partners (ARIB, ATIS, CCSA, ETSI, TSDSI, TTA, TTC).All rights reserved.UMTS™ is a Trade Mark of ETSI registered for the benefit of its members3GPP™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational PartnersLTE™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational PartnersGSM® and the GSM logo are registered and owned by the GSM Association |

Contents

[Foreword 5](#_Toc21294)

[Introduction 6](#_Toc31304)

[1 Scope 7](#_Toc5256)

[2 References 7](#_Toc16912)

[3 Definitions of terms, symbols and abbreviations 8](#_Toc982)

[3.1 Terms 8](#_Toc25050)

[3.2 Symbols 8](#_Toc5755)

[3.3 Abbreviations 8](#_Toc8833)

[4 Beyond 2D Video Formats 8](#_Toc32479)

[4.1 Introduction 8](#_Toc11742)

[4.2 Reference Model for Beyond 2D Video 9](#_Toc17482)

[4.2.1 Overview 9](#_Toc1982)

[4. 3 Beyond 2D Video Representation Formats 10](#_Toc5733)

[5. 3.1 Introduction 10](#_Toc22488)

[4.3.2 <Representation Format 1> 10](#_Toc8702)

[4.3.2.1 Definition 10](#_Toc21505)

[4.3.2.2 Production and Capturing Systems 10](#_Toc6344)

[4.3.2.3 Rendering and Display Systems 10](#_Toc21873)

[4.3.2.4 Supporting Information 10](#_Toc27687)

[4.3.2.5 Benefits and Limitations 10](#_Toc14563)

[5 Overview of existing "Beyond 2D" Video Capabilities in 3GPP 10](#_Toc31559)

[5.1 Introduction 11](#_Toc26816)

[5.2 AR Video Capabilities 11](#_Toc14101)

[5.3 VR Video Profiles 11](#_Toc10694)

[5.4 Messaging Services 12](#_Toc4817)

[6 Evaluation and Characterization Framework 13](#_Toc17865)

[6.1 Overview 13](#_Toc3659)

[6.2 Reference Sequences 13](#_Toc19828)

[6.3 Reference Software Tools 13](#_Toc20247)

[6.4 Metrics 13](#_Toc12611)

[6.5 Encoding Constraints 14](#_Toc10289)

[7 Considered Scenarios 14](#_Toc13060)

[7.1 Introduction 14](#_Toc11238)

[7.2 Scenario 1: <tbd> 14](#_Toc31903)

[7.3 Scenario 2: <tbd> 14](#_Toc521)

[7.4 Scenario 3: <tbd> 14](#_Toc19966)

[7.x Scenario x: <tbd> 14](#_Toc19636)

[8 Common Evaluation Features 14](#_Toc837)

[9 Evaluation of Selected Scenarios 14](#_Toc30756)

[9.1 Introduction 14](#_Toc32191)

[9.2 Scenario 1: <tbd> 14](#_Toc30329)

[9.2.1 Evaluation Overview 14](#_Toc3361)

[9.2.2 Reference Sequences 15](#_Toc11862)

[9.2.3 Performance Metrics 15](#_Toc9246)

[9.2.4 Candidate Solutions 15](#_Toc23488)

[9.2.4.1 Solution 1: <Name> 15](#_Toc14942)

[9.2.4.1.1 Introduction 15](#_Toc23719)

[9.2.4.1.2 Reference Software 15](#_Toc3919)

[9.2.4.1.3 Parameter Settings 15](#_Toc23665)

[9.2.4.1.4 Distribution 15](#_Toc21579)

[9.2.4.1.5 Evaluation Results 15](#_Toc15112)

[9.2.4.1.6 Network Requirements 15](#_Toc12065)

[9.2.4.2 Solution 2: <Name> 15](#_Toc24666)

[9.2.4.2.1 Introduction 15](#_Toc20693)

[9.2.4.2.2 Reference Software 15](#_Toc11491)

[9.2.4.2.3 Parameter Settings 15](#_Toc8812)

[8.2.4.2.4 Distribution 15](#_Toc11814)

[9.2.4.2.5 Evaluation Results 15](#_Toc27047)

[9.2.4.2.6 Network Requirements 15](#_Toc28676)

[9.2.5 Summary of Evaluation 15](#_Toc29570)

[9.3 Scenario 2: <tbd> 15](#_Toc8633)

[9.4 Scenario x: <tbd> 15](#_Toc14253)

[10 Gaps and Optimization Potential 16](#_Toc29691)

[10.1 Identified Gaps and Deficiencies with Video Capabilities 16](#_Toc25794)

[10.2 Potential Requirements for New Video Capabilities 16](#_Toc202)

[10.3 Potential Network Optimizations 16](#_Toc21227)

[11 Conclusions and Proposed Next Steps 16](#_Toc3539)

[Annex A: Scenario Template 16](#_Toc4655)

[A.1 Introduction 16](#_Toc3731)

[A.2 Template 16](#_Toc17132)

[Annex B: Data Formats and Metrics 18](#_Toc6015)

[B.1 Introduction 18](#_Toc14939)

[B.2 Raw Video Sequences 18](#_Toc20412)

[B.2.1 Overview 18](#_Toc13999)

[B.2.2 JSON Schema 19](#_Toc27077)

[Annex C: Reference Sequences 21](#_Toc16677)

[C.1 Introduction 21](#_Toc12239)

For definitive guidance on drafting 3GPP TSs and TRs, see [3GPP TS 21.801](http://www.3gpp.org/DynaReport/21801.htm) supplemented by the 3GPP web page <http://www.3gpp.org/specifications-groups/delegates-corner/writing-a-new-spec>.

Ensure all blue guidance text is removed before submitting the TS/TR to the TSG for approval.

# Foreword

This Technical Report has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

x the first digit:

1 presented to TSG for information;

2 presented to TSG for approval;

3 or greater indicates TSG approved document under change control.

y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

z the third digit is incremented when editorial only changes have been incorporated in the document.

In the present document, modal verbs have the following meanings:

**shall** indicates a mandatory requirement to do something

**shall not** indicates an interdiction (prohibition) to do something

The constructions "shall" and "shall not" are confined to the context of normative provisions, and do not appear in Technical Reports.

The constructions "must" and "must not" are not used as substitutes for "shall" and "shall not". Their use is avoided insofar as possible, and they are not used in a normative context except in a direct citation from an external, referenced, non-3GPP document, or so as to maintain continuity of style when extending or modifying the provisions of such a referenced document.

**should** indicates a recommendation to do something

**should not** indicates a recommendation not to do something

**may** indicates permission to do something

**need not** indicates permission not to do something

The construction "may not" is ambiguous and is not used in normative elements. The unambiguous constructions "might not" or "shall not" are used instead, depending upon the meaning intended.

**can** indicates that something is possible

**cannot** indicates that something is impossible

The constructions "can" and "cannot" are not substitutes for "may" and "need not".

**will** indicates that something is certain or expected to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

**will not** indicates that something is certain or expected not to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

**might** indicates a likelihood that something will happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

**might not** indicates a likelihood that something will not happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

In addition:

**is** (or any other verb in the indicative mood) indicates a statement of fact

**is not** (or any other negative verb in the indicative mood) indicates a statement of fact

The constructions "is" and "is not" do not indicate requirements.

# Introduction

In recent years, video services are evolving from traditional two-dimensional formats to beyond 2D video, which offer users a more lifelike and immersive experience. Research studies indicate that the beyond 2D market was valued at approximately multi-million USD in 2023 and is anticipated to register a CAGR (Compound Annual Growth Rate) of over 24.5% between 2024 and 2032 [2][3][4].

A variety of beyond 2D video formats and video compression technologies are available and emerging. Therefore, in order to determine appropriate beyond 2D video formats for different services, it is essential to evaluate their feasibility and performance, considering implementation constraints, performance indicators, and interoperability considerations. In addition, advanced network capabilities and service extension also need to be investigated to meet the delay and data rate requirements of beyond 2D-related services.

This document provides an overview of available and emerging beyond 2D video formats and compression technologies, which are mostly related to specific types of capturing systems and display technologies; documents a set of end-to-end reference scenarios and workflows for beyond 2D video; analyzes 3GPP-defined video compression technologies and potential new technologies to support each documented scenario; identifies gaps and offer recommendations to potentially extend 3GPP video specifications and capabilities.

# 1 Scope

The present document collects beyond 2D video formats within 3GPP services, as well as a set of beyond 2D video end-to-end reference scenarios and corresponding workflows. It also documents relevant implementation constraints, performance characteristics, and interoperability requirements of existing 3GPP codecs as well as potentially new codecs to support these scenarios. [The primary scope of the present document includes the following aspects:

1. Identify and document beyond 2D formats, that are market-relevant within the next few years, generated from established and emerging capturing systems (including cameras for spatial video capturing), contribution, and usable on display technologies (smartphones, VR HMDs, AR glasses, autostereoscopic and multiscopic displays).

2. Establish and document a set of beyond 2D video end-to-end reference scenarios, including real-time communication, streaming services, split rendering, and messaging and corresponding workflows (capturing, encoding, packaging, delivery, decoding, rendering, including general constraints on latency, as well as complexity) to support 3GPP network related delivery and devices leveraging the generation or display technologies. This includes identifying and defining relevant beyond 2D formats in the context of above workflows, and representation technologies to support delivery of these formats within 3GPP networks.

3. Prioritize the scenarios and the associated formats based on market relevance for further evaluation.

4. Define concrete evaluation framework per scenario (test conditions, KPIs, Metrics, test sequences, agreed reference signals) based on the above prioritized reference scenarios, and evaluate the feasibility and performance of existing 3GPP codecs as well as potentially new codecs to support the scenarios.

5. Based on the findings in steps 1, 2, and 4 document (i) interoperability requirements, (ii) traffic characteristics and (iii) potential QoS optimizations or requirements, to support the above workflows and evaluate the feasibility of new formats with different services, considering the implementation constraints and performance indicators such as encoding, decoding, and rendering complexity, bandwidth utilization, and interoperability considerations.

6. Based on the findings in steps 1, 2, 4 and 5, identify potential gaps or deficiencies of existing 3GPP codecs, and offer recommendations to potentially extend 3GPP video specifications and capabilities.

1. Identify potential areas for normative work as the next phase and communicate with other 3GPP WGs regarding relevant aspects related to the study to the extent needed.]

Editor’s note: The scope may be updated as study progressed.

# 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".

[2] Allied Market Research, “3D Technology Market Size, Share, Competitive Landscape and Trend Analysis Report by Product, Application : Global Opportunity Analysis and Industry Forecast, 2021-2030.”, [www.alliedmarketresearch.com/3d-technology-market.](http://www.alliedmarketresearch.com/3d-technology-market.)

[3] Mordor Intelligence, “Mobile 3D Market Size & Share Analysis - Growth Trends & Forecasts (2024 - 2029).”, <https://www.mordorintelligence.com/industry-reports/mobile-3d-market.>

[4] Grand View Research, “Immersive Technology Market Size, Share & Trends Analysis Report By Component (Hardware, Software, Services), By Technology, By Application, By Industry, By Region, And Segment Forecasts, 2023 - 2030.”, [https://www.grandviewresearch.com/industry-analysis/immersive-technology-market-report.](https://www.mordorintelligence.com/industry-reports/mobile-3d-market.)

[5] 3GPP TS 26.119: "Media Capabilities for Augmented Reality".

[6] 3GPP TS 26.118: "Virtual Reality (VR) profiles for streaming applications".

[7] 3GPP TS 26.143: "Messaging Media Profiles".

[8] 3GPP TS 26.511: "5G Media Streaming (5GMS); Profiles, codecs and formats".

[9] 3GPP TR 26.966: " Evaluation of new HEVC coding tools".

[10] 3GPP TS 26.265: "Media Delivery: Video Capabilities and Operating Points".

[11] 3GPP TR 26.955: "Video codec characteristics for 5G-based services and applications".

# 3 Definitions of terms, symbols and abbreviations

This clause and its three subclauses are mandatory. The contents shall be shown as "void" if the TS/TR does not define any terms, symbols, or abbreviations.

## 3.1 Terms

For the purposes of the present document, the terms given in 3GPP TR 21.905 [1] and the following apply. A term defined in the present document takes precedence over the definition of the same term, if any, in 3GPP TR 21.905 [1].

Definition format (Normal)

**<defined term>:** <definition>.

**example:** text used to clarify abstract rules by applying them literally.

## 3.2 Symbols

For the purposes of the present document, the following symbols apply:

Symbol format (EW)

<symbol> <Explanation>

## 3.3 Abbreviations

For the purposes of the present document, the abbreviations given in 3GPP TR 21.905 [1] and the following apply. An abbreviation defined in the present document takes precedence over the definition of the same abbreviation, if any, in 3GPP TR 21.905 [1].

<ABBREVIATION> <Expansion>

# 4 Beyond 2D Video Formats

## 4.1 Introduction

Editor’s note: This clause documents beyond 2D video formats that are market-relevant within next few years.

## 4.2 Reference Model for Beyond 2D Video

### 4.2.1 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 [11] and extending existing workflows seems to be promising way forward. However, when comparing for example to TR 26.955 [11] for 2D formats or even omnidirectional video formats as defined in TS 26.118 [6], 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 introduced 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 4.2-1 Beyond 2D Video end-to-end reference model

Figure 4.Y.1-1 illustrates a generic beyond 2D Video end-to-end reference model. For example, 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 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 identifier (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 forwarded 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.

## 3 Beyond 2D Video Representation Formats

Editor’s note: The documented format should be referenced by at least one scenario in clause 5, and the following aspects may be provided.

### 3.1 Introduction

As shown in Figure 4.2.1, beyond 2D video representation formats may originate from different production systems and have to target different rendering systems. This clause collects relevant Beyond 2D Video representation formats and provides a discussion on the relevancy of the formats. In order to assess the relevancy of the formats, for each format different aspects are collected, among others:

- Definition of the format, this is preferably backed by a specification.

- Typical applications of the format, e.g. knowledge about support of the format in workflows (tools, etc.)

- Production options of the representation format

- Rendering of the representation format

- Benefits and limitation of the format

- Supporting information

- Typical quality criteria for evaluating the format

- Existing test and reference sequences

- Conversion from other formats (lossless, lossy)

- Uncompressed data size

- Known compression technologies

- Extensibility of the format

### 4.3.2 <Representation Format 1>

#### 4.3.2.1 Definition

#### 4.3.2.2 Production and Capturing Systems

#### 4.3.2.3 Rendering and Display Systems

#### 4.3.2.4 Supporting Information

#### 4.3.2.5 Benefits and Limitations

# 5 Overview of existing "Beyond 2D" Video Capabilities in 3GPP

Editor’s note: This clause summarized existing beyond 2D video capabilities in 3GPP from at least TS.26.119 and TS.26.118.

## 5.1 Introduction

<TBD>

## 5.2 AR Video Capabilities

3GPP TS 26.119 [5] specifies the mandatory and optional media capabilities and profiles to be supported for each XR device type. These media capabilities include support for video codecs (AVC and HEVC), audio codecs (EVS, IVAS and AAC-ELDv2), scene description formats, and XR system capabilities. Table 5.2-1 summarized the Beyond 2D video capabilities defined in clause 7 of TS 26.119 [5].

NOTE: The definition of concurrent video decoder instances can be found in clause 7.1.2.1 of TS 26.119 [5].

Table 5.2-1: Summary of Operation Points

|  |  |  |
| --- | --- | --- |
| Operation Point Name | Max Concurrent Video Decoder Instances | Decoding Capabilities |
| AVC-FullHD-Dec-2 | 2 | Aggregate decoding capabilities of H.264/AVC HP@L4.0 |
| AVC-UHD-Dec-4 | 4 | Aggregate decoding capabilities of H.264/AVC HP@L5.1 |
| HEVC-UHD-Dec-4 | 4 | Aggregate decoding capabilities of H.265/HEVC MP10@L5.1 |
| UHD-Dec-4 | 4 | Aggregate capabilities of *AVC-UHD-Dec-4* |
| Aggregate capabilities of *HEVC-UHD-Dec-4* |
| Decoding up to 4 bitstreams, each not exceeding the capabilities of H.264/AVC HP@L4.0 or H.265/HEVC MP10@L4.1. |
| AVC-8K-Dec-8 | 8 | Aggregate capabilities of H.264/AVC HP@L6.1 |
| HEVC-8K-Dec-8 | 8 | Aggregate capabilities of H.265/HEVC MP10@L6.1 |
| 8K-Dec-8 | 8 | Aggregate capabilities of *AVC-8K-Dec-8* |
| Aggregate capabilities of *HEVC-8K-Dec-8* |
| Decoding up to 8 bitstreams, each not exceeding the capabilities of H.264/AVC HP@L4.0 or H.265/HEVC MP10@L4.1. |
| Decoding up to 4 bitstreams, each not exceeding the capabilities of H.264/AVC HP@L5.1 or H.265/HEVC MP10@L5.1. |

## 5.3 VR Video Profiles

The VR profiles for streaming services are defined in TS 26.118 [6], specifying the coded representation and media profile of 360 VR distribution signals. Table 5.3-1 provides an overview of the 360 VR relevant formats considered in the context of 3GPP VR Profiles.

For restrictions on source formats such as resolution and frame rates, content generation and encoding guidelines, refer to TS 26.118 [6], Annex A.

Table 5.3-1: High-level Summary of Operation Points

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Operation Point name | Decoder | Bit depth | TypicalOriginalSpatialResolution | FrameRate | Colour space format | TransferCharacteristics | Projection | Rotation | RWP | Stereo |
| Basic H.264/AVC | H.264/AVC HP@L5.1 | 8 | Up to 4k | Up to 60 Hz | BT.709 | BT.709 | ERP w/o padding | No | No | No |
| Main H.265/HEVC | H.265/HEVC MP10@L5.1 | 8, 10 | Up to 6k in mono and 3k in stereo | Up to 60 Hz | BT.709BT.2020 | BT.709 | ERP w/o padding | No | Yes | Yes |
| Flexible H.265/HEVC | H.265/HEVC MP10@L5.1 | 8, 10 | Up to 8k in mono and 3k in stereo | Up to 120 Hz | BT.709 BT.2020 | BT.709, BT.2100 PQ, BT.2100 HLG | ERP w/o paddingCMP | No | Yes | Yes |
| Main 8K H.265/HEVC | H.265/HEVC MP10@L6.1 | 10 | Up to 8k in mono and 6k in stereo | Up to 60 Hz for 8K and 120 Hz for 4k | BT.709BT.2020 | BT.709,BT.2100 PQ, BT.2100 HLG | ERP w/o padding | No | Yes, but restricted to coverage | Yes |

Table 5.3-2 summarizes the video operation point, sample entry, and DASH integration associated with each video media profiles defined in clause 5.2 of TS 26.118 [6].

Table 5.3-2 Video Media Profiles

|  |  |  |  |
| --- | --- | --- | --- |
| Media Profile | Operation Point | Sample Entry | DASH Integration |
| Basic Video | Basic H.264/AVC | resvavc1 | Single Adaptation SetSingle Representation streaming |
| Main Video | Main H.265/HEVC or Main 8K H.265/HEVC  | resvhvc1  | Single or Multiple independent Adaptation Sets offeredSingle Representation streaming |
| Advanced Video  | Flexible H.265/HEVC | resvhvc1, hvc2 | Single or Multiple dependent Adaptation Sets offeredSingle or Multiple representation streaming |

## 5.4 Messaging Services

3GPP TS 26.143 [7] specifies the media types, formats, codecs capabilities and profiles for the messaging applications used over the 5G System. The document extends to codecs for speech, audio, video, still images, bitmap graphics, 3D scenes and assets, and other media in general, as well as scene description.

Specifically, the 2D video capabilities defined in TS 26.143 [7] clause 6.2 are fully aligned with 5G Media Streaming in 3GPP TS 26.511 [8]:

- **AVC with HD** and **Full-HD resolutions**

- **HEVC with HD**, **Full-HD** and **UHD resolutions**

For Beyond 2D video capabilities, as HEVC simulcast and HEVC frame packing already been included in SA4 specifications and given the coding benefits MV-HEVC provides compared to these solutions, the support for stereoscopic MV-HEVC for low delay applications of stereoscopic 3D video was recommended by TR 26.966 [9]. This aspect is being addressed in a Rel-19 work TS 26.265 [10].

#

# 6 Evaluation and Characterization Framework

## 6.1 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 6.1-1. Representative reference sequences are collected and stored in a well-defined B2D format. For a video encoder, the 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.



Figure 6.1-1 B2D Evaluation framework

## 6.2 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.

## 6.3 Reference Software Tools

Editor’s Note: For further study

## 6.4 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

## 6.5 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

# 7 Considered Scenarios

## 7.1 Introduction

Editor’s note: This clause collects end-to-end scenarios and corresponding workflows for beyond 2D video, based on the template defined in Annex A. Alignment with the generalized media delivery architecture defined in TS 26.501/506 is expected, primarily addressing reference points M2 and M4.

## 7.2 Scenario 1: <tbd>

## 7.3 Scenario 2: <tbd>

## 7.4 Scenario 3: <tbd>

## 7.x Scenario x: <tbd>

# 8 Common Evaluation Features

Editor’s note: Documents common metrics, software, etc..

# 9 Evaluation of Selected Scenarios

Editor’s note: This clause defines test conditions and parameters, KPIs, Metrics, test sequences, agreed reference signals per scenario.

## 9.1 Introduction

Editor’s note: Identifies the preferred scenarios

## 9.2 Scenario 1: <tbd>

### 9.2.1 Evaluation Overview

Editor’s note: Based on scenario in clause 6, summarizes the source formats parameters used for evaluation, the encoding and decoding constraints, interoperability considerations and the general idea of the performance metrics.

### 9.2.2 Reference Sequences

### 9.2.3 Performance Metrics

### 9.2.4 Candidate Solutions

#### 9.2.4.1 Solution 1: <Name>

##### 9.2.4.1.1 Introduction

##### 9.2.4.1.2 Reference Software

##### 9.2.4.1.3 Parameter Settings

##### 9.2.4.1.4 Distribution

##### 9.2.4.1.5 Evaluation Results

##### 9.2.4.1.6 Network Requirements

Editor’s note: Documents required bitrates as well as possibly other aspects.

#### 9.2.4.2 Solution 2: <Name>

##### 9.2.4.2.1 Introduction

##### 9.2.4.2.2 Reference Software

##### 9.2.4.2.3 Parameter Settings

##### 8.2.4.2.4 Distribution

##### 9.2.4.2.5 Evaluation Results

##### 9.2.4.2.6 Network Requirements

### 9.2.5 Summary of Evaluation

## 9.3 Scenario 2: <tbd>

## 9.4 Scenario x: <tbd>

# 10 Gaps and Optimization Potential

## 10.1 Identified Gaps and Deficiencies with Video Capabilities

## 10.2 Potential Requirements for New Video Capabilities

## 10.3 Potential Network Optimizations

# 11 Conclusions and Proposed Next Steps

Editor’s note: This clause provides conclusion and potential areas for normative work as the next phase.

# Annex A: Scenario Template

## A.1 Introduction

This annex provides a proposed template to introduce a Scenario for Beyond 2D Video. This template has been used to collect the scenarios in this report. The text in blue corresponds to guidelines on the information to be provided with a scenario proposal.

## A.2 Template

The following aspects are considered for a scenario:

1. **Scenario name**
2. **Motivation for the scenario**

*What is the market relevance of the proposed scenario within the next few years? Are there any commercially available or pre-released products or prototypes?*

*Market relevance key indicators:*

1. *Technology evaluation on the market*

*Are there indications of pre-evaluation by service providers, device manufacturers, and/or network operators?*

1. *Industry activities*

*Is there relevant work in 3GPP MRPs, industry collaborations or among market stakeholders?*

1. *Production tools/companies*

*What is the availability of capturing setups, and production software? Are there endorsed formats for representation, contribution, compression, and storage? Is there an ecosystem of content creators?*

1. *Delivery solutions*

*Which delivery type is expected to be used? What are the expected transport formats? Is there SW or HW support and providers?*

1. *Content decoding and rendering*

*Is there decoding SW/HW support, and providers? Are there rendering devices and displays available yet?*

1. **Description of the scenario**

*This provides a description of beyond 2D video end-to-end workflows, which includes identifying and defining beyond 2D formats being used in the context and representation technologies to delivery these formats. The following aspects may be considered for each workflow:*

1. *Capturing and processing*
2. *Encoding*
3. *Packaging and delivery*
4. *Decoding*
5. *\*Post-processing*
6. *Rendering*
7. *General constraints on latency, bandwidth, reliability and complexity*
8. **Supporting companies and 3GPP members**
9. *This documents the 3GPP members that support this scenario in terms of providing the information, test material, test requirements and the characterization for the tests. For each of the identified necessities, a tick box is created in the template.*
10. *Preferably several 3GPP members are included in the support, and in addition a video service provider may be included (not necessarily a 3GPP member).*
11. *Cross-verification is preferably done by the supporters of the scenario.*
12. **Source format properties**

*This defines a clear range of the considered and relevant source formats, including the signal properties, but also the characteristics of the content. As an example, the texture and depth format properties of the source may be used which include:*

1. *Spatial resolutions*
2. *Chroma Format*
3. *Chroma Subsampling*
4. *Aspect ratios*
5. *Frame rates*
6. *Colour space formats*
7. *Transfer Characteristics*
8. *Bit depth*
9. *Viewpoints*
10. *Other signal properties*
11. **Encoding and decoding constraints and settings**

*Typical encoding constraints and settings such as:*

* 1. *Relevant Codec and Codec Profile/Levels according to 3GPP TS (e.g., TS 26.119),*
	2. *Random access frequency*
	3. *Error resiliency requirements*
	4. *Bitrates and quality requirements*
	5. *Bitrate parameters (CBR, VBR, CAE, HRD parameters)*
	6. *ABR encoding requirements (switching frequency, etc.)*
	7. *Latency requirements and specific encoding settings*
	8. *Encoding context: real-time encoding, on device encoding, cloud-based encoding, offline encoding, etc.*
	9. *Required decoding capabilities*
	10. *Synchronization requirements*
1. **Performance Metrics and Requirements**
2. *A clear definition on how the performance needs to be evaluated including metrics, etc addressing the main KPIs of the scenario.*
3. *Objective measures such as PSNR, VMAF, etc, may be used*
4. *Justification on whether objective metrics are sufficient and representative of the subjective performance.*
5. **Interoperability Considerations for the application**
6. *Streaming with DASH/HLS/CMAF/QUIC*
7. *RTP based delivery*
8. **Test Sequences**

*A set of selected test sequences that are provided by the proponents in order to do the evaluation. They should cover a set of source format properties*

1. **Detailed test conditions**

*Provides a proposal for detailed test conditions, for example based on a reference software together with the sequences and configuration parameters.*

1. **External Performance data**

*References to external performance data that can be added, for example other SDOs, public documents and so on.*

1. **Additional Information**
2. *Industry activities*

*Is there Relevant work in industry forums?*

1. *Implementation constraints*

*Are there any indications about scalability of the technology with regards to network and devices?*

1. *Innovation*

*Does the technology address a current or a future need on the market? Can it potentially disrupt existing markets?*

# Annex B: Data Formats and Metrics

## B.1 Introduction

<TBD>

## B.2 Raw Video Sequences

### B.2.1 Overview

<TBD>

### B.2.2 JSON Schema

JSON schema for the raw format is here

< [https://dash-large-files.akamaized.net/WAVE/3GPP/Beyond2D/ReferenceSequence](https://dash-large-files.akamaized.net/WAVE/3GPP/5GVideo/Beyond2D)/raw-schema.json>

1. {
2. "Sequence": {
3. "Name": "Example",
4. "Background": "This is a B2DV format example",
5. "Scenario": "On-demand",
6. "Key": "Identifier",
7. "TR26.956": "Annex X.Y.Z"
8. },
9. "Views": [
10. {
11. "ViewId": "v0",
12. "Extrinsics": {
13. "orientation": {
14. "qw": 0.9999915361,
15. "qx": 0.0024327517,
16. "qy": 0.0024349121,
17. "qz": -0.0022688841
18. },
19. "position": [
20. -0.0006123598,
21. 0.3035059273,
22. 0.0012498678
23. ]
24. },
25. "Intrinsics": {
26. "focalLength": 1002.349976,
27. "principalPoint": {
28. "horizontalNorm": 960.0,
29. "vertical": 540.0
30. }
31. },
32. "ProjectionPlaneSize": {
33. "columnCount": 1920,
34. "rowCount": 1080
35. },
36. "Quantization": {
37. "highNormDisp": 2.000000,
38. "lowNormDisp": 0.200000
39. },
40. "Components": [
41. {
42. "ComponentId": "texture",
43. "Data": {
44. "URI": "https://dash-large-files.akamaized.net/WAVE/3GPP/some/url/file.yuv",
45. "md5": "e537665c18e32bbaf8e5e9d63e18dd2c",
46. "thumbnail": "https://dash-large-files.akamaized.net/WAVE/3GPP/some/url/file.png",
47. "preview": "https://dash-large-files.akamaized.net/WAVE/3GPP/some/url/file.mp4",
48. "size": 7962624000,
49. "md5-10": "1c3550197120f95502c4add38d7ebd33"
50. },
51. "Properties": {
52. "width": 1920,
53. "height": 1080,
54. "format": "yuv",
55. "packing": "planar",
56. "scan": "progressive",
57. "subsampling": "420",
58. "bitDepth": 8,
59. "frameRate": 30,
60. "colourPrimaries": "1",
61. "transferCharacteristics": "1",
62. "matrixCoefficients": "1",
63. "sampleAspectRatio": "1",
64. "duration": 10,
65. "frameCount": 600,
66. "startFrame": 1,
67. "videoFullRangeFlag": "0",
68. "chromaSampleLocType": "0"
69. }
70. },
71. {
72. "ComponentId": "depth",
73. "Data": {
74. "URI": "https://dash-large-files.akamaized.net/WAVE/3GPP/some/url/file.yuv",
75. "md5": "e537665c18e32bbaf8e5e9d63e18dd2c",
76. "thumbnail": "https://dash-large-files.akamaized.net/WAVE/3GPP/some/url/file.png",
77. "preview": "https://dash-large-files.akamaized.net/WAVE/3GPP/some/url/file.mp4",
78. "size": 7962624000,
79. "md5-10": "1c3550197120f95502c4add38d7ebd33"
80. },
81. "Properties": {
82. "width": 1920,
83. "height": 1080,
84. "format": "yuv",
85. "packing": "planar",
86. "scan": "progressive",
87. "subsampling": "420",
88. "bitDepth": 16,
89. "frameRate": 30,
90. "colourPrimaries": "2",
91. "transferCharacteristics": "8",
92. "matrixCoefficients": "0",
93. "sampleAspectRatio": "1",
94. "duration": 10,
95. "frameCount": 600,
96. "startFrame": 1,
97. "videoFullRangeFlag": "1",
98. "chromaSampleLocType": "0"
99. }
100. }
101. ]
102. }
103. ],
104. "copyRight": "Conditions that are suitable for this study",
105. "Contact": {
106. "Name": "Bart Kroon",
107. "Company": "Philips",
108. "e-mail": "bart.kroon@philips.com",
109. "generation": "provided by contact"
110. }
111. }

# 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.

Annex <X> (informative):
Change history

|  |
| --- |
| **Change history** |
| **Date** | **Meeting** | **TDoc** | **CR** | **Rev** | **Cat** | **Subject/Comment** | **New version** |
| 2024-04 | SA4#127-bis | S4-240825 |  |  |  | Initial Version | 0.0.1 |
| 2024-05 | SA4#128 | S4-240947 |  |  |  | Updated version based on SA4-post 127-bis, 24,May,2024 | 0.0.2 |
| 2024-05 | SA4#128 | S4-241319 |  |  |  | Update style and include agreed content in S4-241266, S4-241336 and S4-241318 | 0.0.3 |
| 2024-08 | SA4#129-e | S4-241491 |  |  |  | Updated version based on agreed Tdoc S4aV240023, S4aV240040In SA4-post 128 meeting.  | 0.0.4 |