**Source:** InterDigital Inc.

**Title: Interest of transparency information in the context of Edge Architecture**

**Document for** Discussion and agreement

**Agenda item:** 9.5 - MeCAR (Media Capabilities for Augmented Reality)

# Introduction

This contribution takes place in the context of an edge architecture, where the rendering function is split between the Edge device, located in close network proximity, and the AR glasses. The principle of this architecture is to offload from the AR glasses the computationally intensive tasks to an edge device. The purpose is to decrease device heating and battery drain, which are two main concerns for the design of AR glasses.

In the context of the edge Architecture, it may be of interest to only rely on 2D encoding and decoding capabilities for the AR glasses. The 3D related-functions are ended in the Edge device and 3D rendering is propagated to the AR glasses thanks to a stereoscopic effect by the transmission of appropriate 2D videos. In that context, in addition to the 2D video, it may also be of interest to transmit transparency information to the AR glasses.

This contribution discusses the interest to transmit transparency information and documents how it can be realized with 2D video codeoc capabilities (AVC and HEVC) which have been referenced in TS 26.511 (“5G Media Streaming, Profiles, Codecs and Formats”).

# Interest of transparency information

## 3D rendering based on 2D video streaming

The figure here below depicts a possible distribution of functions where the AR glasses only implement 2D video codec capabilities. The Edge concentrates most of the AR related functions. The final viewport rendering is performed on the Edge which then sends one or several video streams (depending on the solution for carrying one picture per eye, multi-view codec, legacy side by side or top-bottom stereoscopic 3D video,…), corresponding to the viewport to be rendered to the user.



Figure 1 : Possible use of 2D video codecs for 3D rendering in a split context

Corresponding viewport rendering methods have been described in clauses 6.2.4 and 6.2.5 of TS 26.928.

## Use of Transparency information

In addition, it is desirable to permit the transmission of transparency information (alpha\_channel) in addition to the RGB information. Augmented reality may consist of an overlay of a virtual object on the real world which is accessed directly through the “optical-see through” glasses. The overlay is not a full picture but only part of it, the other pixels of the picture being transparent or partially transparent, in case of a shadow effect for instance.

The below pictures depict the overlay of a virtual dragon on the table of a real living room. If the whole video is overlayed, the dragon may appear in the middle of a rectangle corresponding to the video size.This is illustrated on the left picture. With additional transparency information, only the part of the video corresponding to the dragon is overlayed, as illustrated on the right picture.



Figure 2 : video overlay without or with transparency information

# Alpha channel for transparency information in AVC and HEVC

The carriage of transparency information may be ensured by using the concept of auxiliary picture offered by both AVC and HEVC.

## AVC

AVC specification ([H.264](https://www.itu.int/rec/dologin_pub.asp?lang=f&id=T-REC-H.264-202108-I!!PDF-E&type=items)) providesguidelines for carrying transparency information.

It defines the concept of alpha-blending in clause 3.5 : *“A process not specified by this Recommendation | International Standard, in which an auxiliary coded picture is used in combination with a primary coded picture (…) the samples of an auxiliary coded picture are interpreted as indications of the degree of opacity (or, equivalently, the degrees of transparency) associated with the corresponding luma samples of the primary coded picture*.” AVC specification precises in clause 3.7 that “*An auxiliary coded picture must contain the same number of macroblocks as the primary coded picture. Auxiliary coded pictures have no normative effect on the decoding process*.” It also mentions (clause 3.1) that “*In addition to the primary coded picture, an access unit may also contain (…) one auxiliary coded picture*”.

In clause 7.3.2.1.2 (Sequence parameter set extension RBSP syntax), it defines fields related to alpha blending (alpha\_incr\_flag, alpha\_opaque\_value and alpha\_transparent \_value) which semantics are detailed in clause 7.4.2.1.2 (Sequence parameter set extension RBSP semantics). The same clause also precises that “aux\_format\_idc equal to 1 indicates that exactly one auxiliary coded picture is present in each access unit of the coded video sequence, and that for alpha blending purposes the decoded samples of the associated primary coded picture in each access unit should be multiplied by the interpretation sample values of the auxiliary coded picture in the access unit in the display process after output from the decoding process.”

## HEVC

HEVC also defines how to carry an alpha channel in the same video track as the base video.In this case, each frame contains two parts, one base layer containing the video, and one alpha layer containing the alpha channel. Both layers are compressed using the HEVC codec. The two layers are identified by a specific HEVC syntax: a specific Alpha channel information SEI message has to be added, so that the decoder knows how to interpret the auxiliary pictures. A decoder uncapable of handling this SEI message only decodes the base layer.

The concept of auxiliary picture is defined in annex F of HEVC specification ([H.265](https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-H.265-202108-I!!PDF-E&type=items)) :

F.3.5 auxiliary picture: A *picture* that has no normative effect on the *decoding process* of *primary pictures*, and with a nuh\_layer\_id value such that AuxId[ nuh\_layer\_id ] is greater than 0*.*

In the same annex, Table F.2 details the different types of auxiliary pictures:

Table F.2 – Mapping of AuxId to the type of auxiliary pictures

|  |  |  |  |
| --- | --- | --- | --- |
| **AuxId** | **Name of AuxId** | **Type of auxiliary pictures** | **SEI message describing interpretation of auxiliary pictures** |
| 1 | AUX\_ALPHA | Alpha plane | Alpha channel information |
| 2 | AUX\_DEPTH | Depth picture | Depth representation information |
| 3..127 |  | Reserved |  |
| 128..159 |  | Unspecified |  |
| 160..255 |  | Reserved |  |

## System

ISO/IEC 14496 also refers to the carriage of transparency information as an auxiliary picture in the video elementary stream, in particular for HEVC codecs. In ISO/IEC 14496-1 (Systems), it is mentioned in clause 9.2.2.12.1 (transparency of visual object) that “*Content complying with ISO/IEC 14496-1 may include still images or video sequences with representations that include alpha values. These values provide transparency information and are to be treated as specified in ISO/IEC 14772-1:1998, subclause 4.14. For video sequences represented according to ISO/IEC 14496-2, transparency is handled as specified in ISO/IEC 14496-2.*”

ISO/IEC 14496-12 (ISO Base Media file Formats) doesn’t mention anything specific for the transport of auxiliary pictures (which are supposed to be part of the elementary stream). However, it defines the concept of auxiliary video track in clause 12.1.1 (“An auxiliary video track is coded the same as a video track, but uses this different handler type, and is not intended to be visually displayed (e.g. it contains depth information, or other monochrome or color two-dimensional information)"). And in clause 8.3.3 (Track Reference box), it defines ‘auxl’ as specifi value for the *reference\_type* field, with the following semantics : “this track contains auxiliary media for the indicated track (e.g. depth map or alpha plane for video).”

ISO/IEC 14496-15 (Carriage of NAL unit structured video in the ISO basemedia format), describing video track structure, mentions in clause 4.4 : “*Tracks containing video data may use the following template fields: (...) ‘depth’ in the VisualSampleEntry to document the presence of alpha (…) takes one of the following values: (…) 0x20 – the video sequence has alpha (gray or colour)*”. It further mentions in a note in clause 9 devoted to Layered HEVC elementary streams : “*For each auxiliary picture layer included in the track, it is recommended to include, within nalUnit, an SEI NAL unit containing a declarative SEI message, such as the depth representation information SEI message for depth auxiliary picture layers, specifying characteristics of the auxiliary picture layer*.”

The carriage of the transparency information is not necessarily ensured via an auxiliary picture. As mentioned above with ISOBMFF, a multitrack solution, where auxiliary media containing transparency information is carried on a different track, is also possible. It is out of scope of this contribution to discuss the pros and cons of the different solution

# Proposal

We propose to include clause 3 of this contribution as an informative part in the permanent document.

# References

1. TS 26.511 “5GMS Profiles, Codecs and Formats”.
2. H.264 “Advanced video coding for generic audio visual services”
3. H.265 “High Efficiency Video Coding (HEVC) Recommandation ITU\_T H.265”.
4. ISO/IEC 14496 “Coding of audio-visual objects”