|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| MH 1 |  | ge | The comments are authored relative to HEVC SoDAM3 (JCTVC-AB1005-v1). |  |
| MH 2 | All | ge/te | There are significant similarities in clause HEVC SoDAM3 (JCTVC-AB1005-v1) compared to Omnidirectional MediaA Format (OMAF) SoDIS (MPEG N16950):   * The semantics of sample locations in a decoded picture are specified very similarly in D.3.41.5 of HEVC SoDAM3 when compared to clauses 7.2.1 and 5 of OMAF SoDIS. * Additional definitions for clause 3 in HEVC SoDAM3 are identical or similar to the respective definitions in OMAF SoDIS.   The amount of similar text is significant. For example, clause D.3.41.5 of HEVC SoDAM3 is about 6 pages in length.  There is a burden of keeping the HEVC and OMAF standardization processes aligned in time, features, and detailed design (e.g., syntax structures and elements).  Any possible corrigendum needs to be carried out in synchronization for both SEI and box definitions. | Instead of including specification text for, refer to MPEG-I metadata part when it comes to:   * equations for omnidirectional projection formats, * region-wise packing equations and conversions, * conversion from one set of coordinate axes to another, * process to specify the semantics of sample locations in a decoded picture in terms of sphere coordinates.   Refer to MPEG-I metadata part for definitions, or use identical definitions to those of OMAF. |
| MH 3 | All | ge/te | Some functionality that can be expressed by OMAF boxes cannot be expressed by omnidirectional video SEI messages (e.g., guard band indications in region-wise packing). Hence, a key design goal of the HEVC profiles of OMAF, i.e., replicating the metadata of boxes as SEI messages, is not reached. This functionality would be needed for the following purposes:   1. For encrypted video, the displaying process that interprets the omnidirectional video metadata may reside in the trusted domain and thus may only have access to the SEI messages and not the equivalent file boxes. 2. By having equivalent metadata present in both boxes and SEI messages, different player implementation architectures are enabled. In some players, the player logic has access to the file boxes and controls the displaying process, while in other players the displaying process runs independently on the basis of the decoder output only. 3. SEI messages form a convenient standardized interface between the video encoder and the file encapsulator. In other words, the content of the file boxes can be generated based on the SEI messages, and no implementation-specific interface to the file encapsulator is needed. | Provide the same functionality in the omnidirectional video SEI messages as available in the OMAF boxes. Design the syntax and semantics of the SEI messages so that a straightforward conversion from the SEI messages to the OMAF boxes is enabled. |
| MH 4 | D.2.41, D.3.41 |  | The equirectangular projection SEI message and the cubemap projection SEI message include the rotation information. The rotation information is independent of which projection format is used and would also apply to any projection formats that might be specified in the future. | Exclude rotation syntax elements from the equirectangular projection SEI message and the cubemap projection SEI message.  Specify a rotation SEI message that includes the rotation syntax elements.  Align the syntax elements with OMAF's RotationBox and with OMAF's clause 5.3 for conversion from the local coordinate axes to the global coordinate axes. |
| MH 5 | D.2.41, D.3.41 | ge/te | Frame packing arrangement SEI can be misinterpreted by decoders that omit the decoding of the region-wise packing SEI message.  When region-wise packing SEI message is present in the bitstream, the arrangement indicated by the frame packing arrangement SEI message is intended to indicate the frame packing of the projected picture, i.e., the picture that results when the regions of the decoded picture have been "unpacked" according to the region-wise packing SEI message. However, the semantics of the frame packing arrangement SEI message, as specified in clause D.3.16 of HEVC, apply to the decoded picture. Decoders are not required to decode all SEI messages for conforming output. Decoders not parsing the region-wise packing SEI message would interpret the frame packing arrangement SEI message incorrectly, when both SEI messages are present in the bitstream. | Include syntax elements for indicating the frame packing arrangement scheme of the projected pictures and the association of the left or right view to the specific constituent pictures within the region-wise packing SEI message.  Require the region-wise packing SEI message and the frame packing arrangement SEI messages be mutually exclusive: when region-wise packing SEI message is present, frame packing arrangement SEI message shall not be present, and vice versa.  Consider input contributions on this subject. |
| MH 6 | D.2.41.1, D.3.41.1, D.3.41.5 | te | erp\_azimuth\_min, erp\_azimuth\_max, erp\_elevation\_min, and erp\_elevation\_max supposedly indicate azimuth and elevation angles of the center point of the top-left, top-right, bottom-left, and bottom-right samples of the cropped output picture. This conclusion can be drawn from the equations of the equirectangular projection in D.3.41.5.2.  Consequently, in order to set erp\_azimuth\_min, erp\_azimuth\_max, erp\_elevation\_min, and erp\_elevation\_max to correctly, the encoder needs to be careful. For example, if picture consists exactly the left half of an original ERP picture of size 4096x2048, the intuitive choices erp\_azimuth\_min equal to −180 (in degrees) and erp\_azimuth\_max equal to 0 are not correct values. erp\_azimuth\_min should be set equal to −180 + 360÷4096÷2 in degrees (i.e., the azimuth angle of the and center point of the left-most sample column) and erp\_azimuth\_max should be set equal to −360÷4096÷2 in degrees. The values in degrees should be further converted to the units of 2−16 degrees.  It is believed that indicating the coverage as angular ranges is vulnerable to implementation mistakes. This could lead to interoperability problems between encoders and decoders/renderers.  The correct choices of erp\_azimuth\_min, erp\_azimuth\_max, erp\_elevation\_min, and erp\_elevation\_max depend on the picture width and height of the ERP picture from which the content represented by the bitstream is extracted, or equivalently the width and height of the samples in degrees. For example, if the pictures in bitstream A contain exactly the left half of original ERP pictures of size 4096x2048 and if pictures in bitstream B contain exactly the left half of original ERP pictures of size 3072x1536, erp\_azimuth\_min and erp\_azimuth\_max values for these two bitstreams would respectively differ. This is unintuitive and should be avoided. | It is proposed to use relative integer sample coordinates for indicating the coverage. In other words, rather than using erp\_azimuth\_min, erp\_azimuth\_max, erp\_elevation\_min, and erp\_elevation\_max, it is proposed to indicate proj\_picture\_width, proj\_picture\_height, proj\_reg\_left, proj\_reg\_top, proj\_reg\_width, and proj\_reg\_height to indicate the coverage. proj\_picture\_width and proj\_picture\_height specify the width and height of a conceptual projected picture. proj\_reg\_left and proj\_reg\_top specify the sample location of the top-left sample of the cropped output picture on the conceptual projected picture. proj\_reg\_width and proj\_reg\_height specify the width and height of the cropped output picture in units of samples of the conceptual projected picture.  For example, if the pictures contain exactly the left half of original ERP pictures, the proposed syntax elements can be set to the following values as an example: proj\_picture\_width = 4096, proj\_picture\_height = 2048, proj\_reg\_left = 0, proj\_reg\_top = 0, proj\_reg\_width = 2048, proj\_reg\_height = 2048.  It is noted that proj\_reg\_left + proj\_reg\_width is allowed to be greater than proj\_picture\_width (and is handled in wraparound manner).  It is also remarked that this design is aligned with OMAF's way of specifying coverage with RegionWisePackingBox, which when used with one region (for monoscopic video) specifies coverage on projected picture domain (i.e., on local coordinates) in manner that affects the mapping of locations on the decoded picture to sphere coordinates. |
| MH 7 | D.2.41.1, D.3.41.1, D.3.41.5 | te | Multiple syntax structures that achieve the same functionality should be avoided. As discussed in the previous comments erp\_azimuth\_min, erp\_azimuth\_max, erp\_elevation\_min, and erp\_elevation\_max achieve the functionality offered by region-wise packing of one region (for monoscopic video) but are more vulnerable to implementation errors and less intuitive. | Exclude erp\_azimuth\_min, erp\_azimuth\_max, erp\_elevation\_min, and erp\_elevation\_max from the equirectangular projection SEI message and their impact from the equations in D.3.41.5.2.  When region-wise packing SEI message is present, require it to be present in the same SEI NAL unit as the equirectangular projection SEI message or the cubemap projection SEI message. Consequently, it is reasonably simple to discover that region-wise packing is applied.  Provide in the region-wise packing SEI message a mechanism to indicate early in the message that the values of the syntax elements comply with certain pre-defined constraints. Specify a code point corresponding to the constraints of 'erpv' scheme type of OMAF. This suggestion aims at:   * providing pre-defined constrained sets of region-wise packing so that decoders/renderers need not support its full flexibility, and * simplifying encapsulation of bitstreams into OMAF-compliant files. For example, the file encapsulator does not need to check the bitstream whether all constraints of the 'erpv' scheme type are fulfilled, since the SEI message already asserts that. |
| MH 8 | D.2.41.1, D.3.41.1, D.3.41.5 | te | In the equirectangular projection SEI message, erp\_azimuth\_min and erp\_azimuth\_max are required to be in the range of −360°, inclusive, to 360°, exclusive, erp\_azimuth\_max is required to be greater than erp\_azimuth\_min. This provides the flexibility of content with greater than 360° azimuth range, supposedly intended to provide a guard band for the back seam of ERP and thus making it unperceivable.  However, it is unclear which operation should be used in displaying decoded pictures that have content overlapping on sphere domain. For example, in stereoscopic ERP packed side-by-side, the guard band could simply represent a boundary pixel extension in order to improve inter prediction and should not be used for displaying.  The syntax is not flexible enough to indicate stereoscopic ERP packed side-by-side and a guard in between the constituent pictures.  The syntax is not flexible enough to indicate stereoscopic ERP packed in top-bottom manner with a guard band between the constituent pictures. | The proposal for MH 7 applies. |
| MH 9 | D.2.41.1, D.3.41.1, D.3.41.5 | te | It is not specified in D.3.41.1 whether erp\_azimuth\_min, erp\_azimuth\_max, erp\_elevation\_min, and erp\_elevation\_max are relative to the global or local coordinate axes. Based on the equations of the equirectangular projection in D.3.41.5.2, erp\_azimuth\_min, erp\_azimuth\_max, erp\_elevation\_min, and erp\_elevation\_max are relative to the local coordinate axes. This design choice differs from CoverageInformationBox of OMAF, which indicates the coverage relative to the global coordinate axes. Moreover, CoverageInformationBox does not affect the projection equations mapping a location of 2D picture to sphere coordinates.  In OMAF it is intended to use the coverage information for player's content selection. The player has access to the current viewing orientation relative to the global coordinate axes. In order to avoid conversion between coordinate axes, the coverage is indicated relative to the global coordinate axes. | Exclude erp\_azimuth\_min, erp\_azimuth\_max, erp\_elevation\_min, and erp\_elevation\_max from the equirectangular projection SEI message and their impact from the equations in D.3.41.5.2.  It is noted that it is not considered useful to specify a coverage information SEI message that would correspond to OMAF's CoverageInformationBox, i.e., would not affect the projection equations and would be relative to global coordinate axes. The coverage information SEI message is not envisioned to be used for content selection and is not used for rendering (since it does not affect the projection equations). |
| MH 10 | D.2.41.2, D.2.41.3, D.3.41.2 D.3.41.3 | te | The cubemap projection SEI message includes some padding functionality, which is similar to but not aligned with OMAF's guard bands and is more constrained. The differences include the following:   1. OMAF's guard bands are generic to any omnidirectional projection format, which is asserted to be a useful feature. 2. cmp\_padding\_type values of the cubemap projection SEI message do not have the same semantics as gb\_type[i][j] values of RegionWisePackingStruct. Alignment of the semantics would be desirable. 3. cmp\_padding\_type is the same for each padding region of each cube face in the cubemap projection SEI message. gb\_type[i][j] values can be freely selected per each side of each packed region. 4. The cubemap projection SEI message specifies padding that have equal thickness in each side of each cube face, whereas the thickness of guard bands on each side of a packed region can be freely selected in RegionWisePackingStruct. 5. The padding specified in the cubemap projection SEI message is not reflected in the equations in D.3.41.5.2 and hence specified incorrectly. | Include guard band signalling into region-wise packing SEI message with syntax structures and elements aligned with OMAF.  Remove the padding related syntax elements from the cubemap projection SEI message.  Provide in the region-wise packing SEI message a mechanism to indicate early in the message that the values of the syntax elements comply with certain pre-defined constraints. Specify a code point corresponding to a cube map that is otherwise identical to "unpacked" cube map but has a guard band between the rows of cube faces. This suggestion aims at providing pre-defined constrained sets of region-wise packing so that decoders/renderers need not support its full flexibility. |
| MH 11 | D.2.41.3, D.3.41.3 | te | The region-wise packing SEI message lacks guard band indications, which are included in OMAF's RegionWisePackingBox. Guard bands are asserted to be useful when region-wise packing is in use, e.g. in assisting the rendering process, regardless which omnidirectional projection format is in use. | Include guard band signalling into region-wise packing SEI message with syntax structures and elements aligned with OMAF. |
| MH 12 | D.3.41.1 | te | The semantics of erp\_azimuth\_min, erp\_azimuth\_max, erp\_elevation\_min, and erp\_elevation\_max are unclear in the following aspects:   * It is not specified whether these syntax elements are relative to the global or local coordinate axes. Based on the equations of the equirectangular projection in D.3.41.5.2, erp\_azimuth\_min, erp\_azimuth\_max, erp\_elevation\_min, and erp\_elevation\_max are relative to the local coordinate axes, but this should be explicitly stated also in the semantics. * erp\_azimuth\_min, erp\_azimuth\_max, erp\_elevation\_min, and erp\_elevation\_max supposedly indicate azimuth and elevation angles of the center point of the top-left, top-right, bottom-left, and bottom-right samples of the cropped output picture. This conclusion can be drawn from the equations of the equirectangular projection in D.3.41.5.2. | If these syntax elements remain in the equirectangular projection SEI message, clarify their semantics as commented. |
| MH 13 | D.3.41.4 | te | The omnidirectional viewport SEI message lacks the indication whether the viewport is taken from the left view or the right view of stereoscopic content, or if the metadata is meant for displaying on a stereoscopic display. | Include this functionality in the omnidirectional viewport SEI message. |
| MH 14 | D.3.41.4 | te | The definition of OMAF's recommended viewport timed metadata differs from that of the omnidirectional viewport SEI message.  In OMAF the intent is clear:  The recommended viewport timed metadata track indicates the viewport that should be displayed when the user does not have control of the viewing orientation or has released control of the viewing orientation.  The following is specified in JCTVC-AB1005-v1:  The omnidirectional viewport SEI message specifies the coordinates of one or more regions of spherical-coordinate geometry, bounded by four great circles, corresponding to viewports recommended for display.  It is not specified in which situation or use case the regions are recommended for display. | Align the definition of omnidirectional viewport SEI message with that of OMAF's recommended viewport timed metadata. |
| MH 15 | D.3.41.4 | te | Collection of user statistics on viewing orientations for omnidirectional video typically results into a heat map sequence without clear distinct viewport boundaries. It is unclear   * how a heat map sequence or such statistics are to be converted to the omnidirectional viewport SEI message; * whether the most-viewed viewport sequence would be the most suited form of representing such statistical data for target use cases. For example, would it be beneficial to allow excluding the width (in azimuth range) and height (in elevation range) of the viewport from the SEI message?   If statistics have been collected that a viewport at orientation (0,0) has the highest popularity and a viewport at orientation (0,1) (in azimuth, elevation in units of units of 2−16 degrees) has the second highest popularity, which options does the content author have for generating the omnidirectional viewport SEI message?  If statistics have been collected that a viewport at orientation (0,0) with 95° field of view has the highest popularity and a viewport at orientation (0,0) with 100° field of view has the second highest popularity, which options does the content author have for generating the omnidirectional viewport SEI message? | Please clarify. |
| MH 16 | D.3.41.4 | te | Are the viewports continuous in time, i.e., does the N-th viewport within an omnidirectional viewport SEI message in access unit A correspond to the N-th viewport within an omnidirectional viewport SEI message applicable to the next access unit in output order? In other words, when a player chooses to follow the N-th viewport in displaying, should it display the N-th viewport consistently? If yes, is there consistency across CVS boundaries too?  It is specified that omni\_viewport\_cancel\_flag equal to 1 indicates that the SEI message cancels the persistence of any previous omnidirectional viewport SEI message in output order. It could be desirable to have a cancelling mechanism that is specific to a particular viewport rather than all of them. | Please clarify. |
| MH 17 | D.3.41.4 | te | How is the omnidirectional viewport orientation SEI message used if two or more viewports have an unequal or unspecified priority relative to each other? | Please clarify. |
| MH 18 | D.3.41.4 | ed | The phrase director's cut is used in the normative semantics but not defined. | If the phrase "director's cut" is kept as part of normative text, it is proposed to clarify its meaning or avoid the use of the term. |
| MH 19 | D.3.41.4 | te | It is not specified whether omni\_viewport\_azimuth\_centre[ i ] and omni\_viewport\_elevation\_centre[ i ] are relative to the local or global coordinate axes. | Specify omni\_viewport\_azimuth\_centre[ i ] and omni\_viewport\_elevation\_centre[ i ] to be relative to the global coordinate axes. |