|  |  |
| --- | --- |
| **Joint Collaborative Team on Video Coding (JCT-VC)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  34th Meeting: Marrakech, MA, 12–18 Jan. 2019 | Document: JCTVC-AH1012-v1 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **Annotated regions and fisheye video information SEI messages for HEVC (Draft 2)** | | |
| *Status:* | Output document approved by JCT-VC | | |
| *Purpose:* | Draft text for standardization | | |
| *Author(s) or Contact(s):* | Jill Boyce Gary J. Sullivan Ye-Kui Wang | Email: | [jill.boyce@intel.com](mailto:jill.boyce@intel.com) [garysull@microsoft.com](mailto:garysull@microsoft.com) [yekui.wang@huawei.com](mailto:yekui.wang@huawei.com) |
| *Source:* | Editors | | |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Abstract

This document contains the draft text for changes to the High Efficiency Video Coding (HEVC) standard (Rec. ITU-T H.265 | ISO/IEC 23008-2) to specify additional supplemental enhancement information (SEI) messages for fisheye and annotated regions.

*Replace D.2.1 with the following:*

**D.2.1 General SEI message syntax**

|  |  |
| --- | --- |
| sei\_payload( payloadType, payloadSize ) { | **Descriptor** |
| if( nal\_unit\_type  = =  PREFIX\_SEI\_NUT ) |  |
| if( payloadType  = =  0 ) |  |
| buffering\_period( payloadSize ) |  |
| else if( payloadType  = =  1 ) |  |
| pic\_timing( payloadSize ) |  |
| else if( payloadType  = =  2 ) |  |
| pan\_scan\_rect( payloadSize ) |  |
| else if( payloadType  = =  3 ) |  |
| filler\_payload( payloadSize ) |  |
| else if( payloadType  = =  4 ) |  |
| user\_data\_registered\_itu\_t\_t35( payloadSize ) |  |
| else if( payloadType  = =  5 ) |  |
| user\_data\_unregistered( payloadSize ) |  |
| else if( payloadType  = =  6 ) |  |
| recovery\_point( payloadSize ) |  |
| else if( payloadType  = =  9 ) |  |
| scene\_info( payloadSize ) |  |
| else if( payloadType  = =  15 ) |  |
| picture\_snapshot( payloadSize ) |  |
| else if( payloadType  = =  16 ) |  |
| progressive\_refinement\_segment\_start( payloadSize ) |  |
| else if( payloadType  = =  17 ) |  |
| progressive\_refinement\_segment\_end( payloadSize ) |  |
| else if( payloadType  = =  19 ) |  |
| film\_grain\_characteristics( payloadSize ) |  |
| else if( payloadType  = =  22 ) |  |
| post\_filter\_hint( payloadSize ) |  |
| else if( payloadType  = =  23 ) |  |
| tone\_mapping\_info( payloadSize ) |  |
| else if( payloadType  = =  45 ) |  |
| frame\_packing\_arrangement( payloadSize ) |  |
| else if( payloadType  = =  47 ) |  |
| display\_orientation( payloadSize ) |  |
| else if( payloadType  = =  56 ) |  |
| green\_metadata( payloadsize ) /\* specified in ISO/IEC 23001-11 \*/ |  |
| else if( payloadType  = =  128 ) |  |
| structure\_of\_pictures\_info( payloadSize ) |  |
| else if( payloadType  = =  129 ) |  |
| active\_parameter\_sets( payloadSize ) |  |
| else if( payloadType  = =  130 ) |  |
| decoding\_unit\_info( payloadSize ) |  |
| else if( payloadType  = =  131 ) |  |
| temporal\_sub\_layer\_zero\_index( payloadSize ) |  |
| else if( payloadType  = =  133 ) |  |
| scalable\_nesting( payloadSize ) |  |
| else if( payloadType  = =  134 ) |  |
| region\_refresh\_info( payloadSize ) |  |
| else if( payloadType  = =  135 ) |  |
| no\_display( payloadSize ) |  |
| else if( payloadType  = =  136 ) |  |
| time\_code( payloadSize ) |  |
| else if( payloadType  = =  137 ) |  |
| mastering\_display\_colour\_volume( payloadSize ) |  |
| else if( payloadType  = =  138 ) |  |
| segmented\_rect\_frame\_packing\_arrangement( payloadSize ) |  |
| else if( payloadType  = =  139 ) |  |
| temporal\_motion\_constrained\_tile\_sets( payloadSize ) |  |
| else if( payloadType  = =  140 ) |  |
| chroma\_resampling\_filter\_hint( payloadSize ) |  |
| else if( payloadType  = =  141 ) |  |
| knee\_function\_info( payloadSize ) |  |
| else if( payloadType  = =  142 ) |  |
| colour\_remapping\_info( payloadSize ) |  |
| else if( payloadType  = =  143 ) |  |
| deinterlaced\_field\_identification( payloadSize ) |  |
| else if( payloadType  = =  144 ) |  |
| content\_light\_level\_info( payloadSize ) |  |
| else if( payloadType  = =  145 ) |  |
| dependent\_rap\_indication( payloadSize ) |  |
| else if( payloadType  = =  146 ) |  |
| coded\_region\_completion( payloadSize ) |  |
| else if( payloadType  = =  147 ) |  |
| alternative\_transfer\_characteristics( payloadSize ) |  |
| else if( payloadType  = =  148 ) |  |
| ambient\_viewing\_environment( payloadSize ) |  |
| else if( payloadType  = =  149 ) |  |
| content\_colour\_volume( payloadSize ) |  |
| else if( payloadType  = =  150 ) |  |
| equirectangular\_projection( payloadSize ) |  |
| else if( payloadType  = =  151 ) |  |
| cubemap\_projection( payloadSize ) |  |
| else if( payloadType  = =  152 ) |  |
| fisheye\_video\_info( payloadSize ) |  |
| else if( payloadType  = =  154 ) |  |
| sphere\_rotation( payloadSize ) |  |
| else if( payloadType  = =  155 ) |  |
| regionwise\_packing( payloadSize ) |  |
| else if( payloadType  = =  156 ) |  |
| omni\_viewport( payloadSize ) |  |
| else if( payloadType  = =  157 ) |  |
| regional\_nesting( payloadSize ) |  |
| else if( payloadType  = =  158 ) |  |
| mcts\_extraction\_info\_sets( payloadSize ) |  |
| else if( payloadType  = =  159 ) |  |
| mcts\_extraction\_info\_nesting( payloadSize ) |  |
| else if( payloadType  = =  160 ) |  |
| layers\_not\_present( payloadSize ) /\* specified in Annex F \*/ |  |
| else if( payloadType  = =  161 ) |  |
| inter\_layer\_constrained\_tile\_sets( payloadSize ) /\* specified in Annex F \*/ |  |
| else if( payloadType  = =  162 ) |  |
| bsp\_nesting( payloadSize ) /\* specified in Annex F \*/ |  |
| else if( payloadType  = =  163 ) |  |
| bsp\_initial\_arrival\_time( payloadSize ) /\* specified in Annex F \*/ |  |
| else if( payloadType  = =  164 ) |  |
| sub\_bitstream\_property( payloadSize ) /\* specified in Annex F \*/ |  |
| else if( payloadType  = =  165 ) |  |
| alpha\_channel\_info( payloadSize ) /\* specified in Annex F \*/ |  |
| else if( payloadType  = =  166 ) |  |
| overlay\_info( payloadSize ) /\* specified in Annex F \*/ |  |
| else if( payloadType  = =  167 ) |  |
| temporal\_mv\_prediction\_constraints( payloadSize ) /\* specified in Annex F \*/ |  |
| else if( payloadType  = =  168 ) |  |
| frame\_field\_info( payloadSize ) /\* specified in Annex F \*/ |  |
| else if( payloadType  = =  176 ) |  |
| three\_dimensional\_reference\_displays\_info( payloadSize ) /\* specified in Annex G \*/ |  |
| else if( payloadType  = =  177 ) |  |
| depth\_representation\_info( payloadSize ) /\* specified in Annex G \*/ |  |
| else if( payloadType  = =  178 ) |  |
| multiview\_scene\_info( payloadSize ) /\* specified in Annex G \*/ |  |
| else if( payloadType  = =  179 ) |  |
| multiview\_acquisition\_info( payloadSize ) /\* specified in Annex G \*/ |  |
| else if( payloadType  = =  180 ) |  |
| multiview\_view\_position( payloadSize ) /\* specified in Annex G \*/ |  |
| else if( payloadType  = =  181 ) |  |
| alternative\_depth\_info( payloadSize ) /\* specified in Annex I \*/ |  |
| else if( payloadType  = =  200 ) |  |
| sei\_manifest( payloadSize ) |  |
| else if( payloadType  = =  201 ) |  |
| sei\_prefix\_indication( payloadSize ) |  |
| else if( payloadType  = =  202 ) |  |
| annotated\_regions( payloadSize ) |  |
| else |  |
| reserved\_sei\_message( payloadSize ) |  |
| else /\* nal\_unit\_type  = =  SUFFIX\_SEI\_NUT \*/ |  |
| if( payloadType  = =  3 ) |  |
| filler\_payload( payloadSize ) |  |
| else if( payloadType  = =  4 ) |  |
| user\_data\_registered\_itu\_t\_t35( payloadSize ) |  |
| else if( payloadType  = =  5 ) |  |
| user\_data\_unregistered( payloadSize ) |  |
| else if( payloadType  = =  17 ) |  |
| progressive\_refinement\_segment\_end( payloadSize ) |  |
| else if( payloadType  = =  22 ) |  |
| post\_filter\_hint( payloadSize ) |  |
| else if( payloadType  = =  132 ) |  |
| decoded\_picture\_hash( payloadSize ) |  |
| else if( payloadType  = =  146 ) |  |
| coded\_region\_completion( payloadSize ) |  |
| else |  |
| reserved\_sei\_message( payloadSize ) |  |
| if( more\_data\_in\_payload( ) ) { |  |
| if( payload\_extension\_present( ) ) |  |
| **reserved\_payload\_extension\_data** | u(v) |
| **payload\_bit\_equal\_to\_one** /\* equal to 1 \*/ | f(1) |
| while( !byte\_aligned( ) ) |  |
| **payload\_bit\_equal\_to\_zero** /\* equal to 0 \*/ | f(1) |
| } |  |
| } |  |

*Renumber clauses D.2.41.3 through D.2.41.5 as D.2.41.4 through D.2.41.6.*

*Add clause D.2.41.3, as follows:*

**D.2.41.3 Fisheye video information SEI message syntax**

|  |  |
| --- | --- |
| fisheye\_video\_info( payloadSize ) { | **Descriptor** |
| **fisheye\_cancel\_flag** | u(1) |
| if( !fisheye\_cancel\_flag ) { |  |
| **fisheye\_persistence\_flag** | u(1) |
| **fisheye\_view\_dimension\_idc** | u(3) |
| **fisheye\_reserved\_zero\_3bits** | u(3) |
| **fisheye\_num\_active\_areas\_minus1** | u(8) |
| for( i = 0; i  <=  fisheye\_num\_active\_areas\_minus1; i++ ) { |  |
| **fisheye\_circular\_region\_centre\_x**[ i ] | u(32) |
| **fisheye\_circular\_region\_centre\_y**[ i ] | u(32) |
| **fisheye\_rect\_region\_top**[ i ] | u(32) |
| **fisheye\_rect\_region\_left**[ i ] | u(32) |
| **fisheye\_rect\_region\_width**[ i ] | u(32) |
| **fisheye\_rect\_region\_height**[ i ] | u(32) |
| **fisheye\_circular\_region\_radius**[ i ] | u(32) |
| **fisheye\_scene\_radius**[ i ] | u(32) |
| **fisheye\_camera\_centre\_azimuth**[ i ] | i(32) |
| **fisheye\_camera\_centre\_elevation**[ i ] | i(32) |
| **fisheye\_camera\_centre\_tilt**[ i ] | i(32) |
| **fisheye\_camera\_centre\_offset\_x**[ i ] | u(32) |
| **fisheye\_camera\_centre\_offset\_y**[ i ] | u(32) |
| **fisheye\_camera\_centre\_offset\_z**[ i ] | u(32) |
| **fisheye\_field\_of\_view**[ i ] | u(32) |
| **fisheye\_num\_polynomial\_coeffs**[ i ] | u(16) |
| for( j = 0; j < fisheye\_num\_polynomial\_coeffs[ i ]; j++ ) |  |
| **fisheye\_polynomial\_coeff**[ i ][ j ] | i(32) |
| } |  |
| } |  |
| } |  |

*Renumber clause D.2.45 (Reserved SEI message syntax) as D.2.47.*

*In D.3.1, replace the following paragraphs:*

The list SingleLayerSeiList is set to consist of the payloadType values 3, 6, 9, 15, 16, 17, 19, 22, 23, 45, 47, 56, 128, 129, 131, 132, 134 to 151, inclusive, and 154 to 159, inclusive.

The list VclAssociatedSeiList is set to consist of the payloadType values 2, 3, 6, 9, 15, 16, 17, 19, 22, 23, 45, 47, 56, 128, 131, 132, 134 to 151, inclusive, and 154 to 159, inclusive.

The list PicUnitRepConSeiList is set to consist of the payloadType values 0, 1, 2, 6, 9, 15, 16, 17, 19, 22, 23, 45, 47, 56, 128, 129, 131, 132, 133, 135 to 151, inclusive, and 154 to 159, inclusive.

*with the following:*

The list SingleLayerSeiList is set to consist of the payloadType values 3, 6, 9, 15, 16, 17, 19, 22, 23, 45, 47, 56, 128, 129, 131, 132, 134 to 152, inclusive, 154 to 159, inclusive, and 200 to 202, inclusive.

The list VclAssociatedSeiList is set to consist of the payloadType values 2, 3, 6, 9, 15, 16, 17, 19, 22, 23, 45, 47, 56, 128, 131, 132, 134 to 152, inclusive, 154 to 159, inclusive, and 200 to 202, inclusive.

The list PicUnitRepConSeiList is set to consist of the payloadType values 0, 1, 2, 6, 9, 15, 16, 17, 19, 22, 23, 45, 47, 56, 128, 129, 131, 132, 133, 135 to 152, inclusive, 154 to 159, inclusive, and 200 to 202, inclusive.

*In D.3.1, in Table D.1, insert the following row immediately after the row for "Cubemap projection" in the table:*

|  |  |
| --- | --- |
| Fisheye video information | Specified by the syntax of the SEI message |

*In D.3.1, in Table D.1, append the following rows to the end of the table:*

|  |  |
| --- | --- |
| Annotated regions | Specified by the syntax of the SEI message |

*Renumber clauses D.3.41.4 through D.3.41.6 (and their subordinate subclauses) as D.3.41.5 through D.3.41.7 (and subordinate subclauses)*

*Add clause D.3.41.4, as follows:*

**D.3.41.4 Fisheye video information SEI message semantics**

The presence of the fisheye video information SEI message for any picture of a CLVS indicates that the picture is a fisheye video picture containing a number of active areas captured by fisheye camera lens. The information carried in the fisheye video information SEI message enables remapping of the colour samples of the pictures onto a sphere coordinate space in sphere coordinates (ϕ, θ), for use in omnidirectional video applications for which the viewing perspective is from the origin looking outward toward the inside of the sphere. The sphere coordinates are defined so that ϕ is the azimuth (longitude, increasing eastward) and θ is the elevation (latitude, increasing northward).

When a fisheye video information SEI message is present for any picture of a CLVS of a particular layer, a fisheye video information SEI message shall be present for the first picture of the CLVS and no equirectangular projection SEI message or cubemap projection SEI message shall be present for any picture of the CLVS.

When general\_non\_packed\_constraint\_flag is equal to 1 in the active SPS for the current layer, there shall be no fisheye video information SEI messages applicable for any picture of the CLVS of the current layer.

When aspect\_ratio\_idc is present and greater than 1 in the active SPS for the current layer, there should be no fisheye video information SEI messages applicable for any picture of the CLVS of the current layer.

When a frame packing arrangement SEI message with frame\_packing\_arrangement\_cancel\_flag equal to 0 or a segmented rectangular frame packing arrangement SEI message with segmented\_rect\_frame\_packing\_arrangement\_cancel\_flag equal to 0 that applies to the picture is present, a fisheye video information SEI message with fisheye\_cancel\_flag equal to 0 that applies to the picture shall not be present. Decoders shall ignore fisheye video information SEI messages when a frame packing arrangement SEI message with frame\_packing\_arrangement\_cancel\_flag equal to 0 or a segmented rectangular frame packing arrangement SEI message with segmented\_rect\_frame\_packing\_arrangement\_cancel\_flag equal to 0 that applies to the picture is present.

**fisheye\_cancel\_flag** equal to 1 indicates that the SEI message cancels the persistence of any previous fisheye video information SEI message in output order. fisheye\_cancel\_flag equal to 0 indicates that fisheye video information follows.

**fisheye\_persistence\_flag** specifies the persistence of the fisheye video information SEI message for the current layer.

fisheye\_persistence\_flag equal to 0 specifies that the fisheye video information SEI message applies to the current decoded picture only.

Let picA be the current picture. fisheye\_persistence\_flag equal to 1 specifies that the fisheye video information SEI message persists for the current layer in output order until one or more of the following conditions are true:

– A new CLVS of the current layer begins.

– The bitstream ends.

– A picture picB in the current layer in an access unit containing a fisheye video information SEI message that is applicable to the current layer is output for which PicOrderCnt( picB ) is greater than PicOrderCnt( picA ), where PicOrderCnt( picB ) and PicOrderCnt( picA ) are the PicOrderCntVal values of picB and picA, respectively, immediately after the invocation of the decoding process for picture order count for picB.

**fisheye\_view\_dimension\_idc** indicates the alignment and viewing direction of a fisheye lens, as follows:

– fisheye\_view\_dimension\_idc equal to 0 indicates that fisheye\_num\_active\_areas is equal to 2, and the values of fisheye\_camera\_centre\_azimuth, fisheye\_camera\_centre\_elevation, fisheye\_camera\_centre\_tilt, fisheye\_camera\_centre\_offset\_x, fisheye\_camera\_centre\_offset\_y, and fisheye\_camera\_centre\_offset\_z are such that the active areas have aligned optical axes and face opposite directions, and the sum of fisheye\_field\_of\_view values is greater than or equal to 360 \* 216.

– fisheye\_view\_dimension\_idc equal to 1 indicates that fisheye\_num\_active\_areas is equal to 2, and the values of fisheye\_camera\_centre\_azimuth, fisheye\_camera\_centre\_elevation, fisheye\_camera\_centre\_tilt, fisheye\_camera\_centre\_offset\_x, fisheye\_camera\_centre\_offset\_y, and fisheye\_camera\_centre\_offset\_z are such that the active areas have parallel optical axes that are orthogonal to the line intersecting the camera centre points, and the camera corresponding to i equal to 0 is the left view.

– fisheye\_view\_dimension\_idc equal to 2 indicates that fisheye\_num\_active\_areas is equal to 2, and the values of fisheye\_camera\_centre\_azimuth, fisheye\_camera\_centre\_elevation, fisheye\_camera\_centre\_tilt, fisheye\_camera\_centre\_offset\_x, fisheye\_camera\_centre\_offset\_y, and fisheye\_camera\_centre\_offset\_z are such that the active areas have parallel optical axes that are orthogonal to the line intersecting the camera centre points, and the camera corresponding to i equal to 0 is the right view.

– fisheye\_view\_dimension\_idc equal to 7 indicates that no additional constraints are implied for the syntax element values within the fisheye video information SEI message.

– Values of fisheye\_view\_dimension\_idc in the range of 3 to 6, inclusive, are reserved for future use by ITU-T | ISO/IEC. Decoders encountering a value of fisheye\_view\_dimension\_idc in the range of 3 to 6, inclusive, shall ignore it.

**fisheye**\_**reserved\_zero\_3bits** shall be equal to 0 in bitstreams conforming to this version of this Specification. Other values for fisheye\_reserved\_zero\_3bits are reserved for future use by ITU-T | ISO/IEC. Decoders shall ignore the value of fisheye\_reserved\_zero\_3bits.

**fisheye\_num\_active\_areas\_minus1** plus 1 specifies the number of active areas in the coded picture. The value of fisheye\_num\_active\_areas\_minus1 shall be in the range of 0 to 3, inclusive. Values of fisheye\_num\_active\_areas\_minus1 greater than 3 are reserved for future use by ITU-T | ISO/IEC. Decoders encountering a fisheye video information SEI message with fisheye\_num\_active\_areas\_minus1 greater than 3 shall ignore the fisheye video information SEI message.

**fisheye\_circular\_region\_centre\_x**[ i ] and **fisheye\_circular\_region\_centre\_y**[ i ] specify the horizontal and vertical coordinates of the centre of the circular region that contains the i-th active area in the coded picture, respectively, in units of 2−16 luma samples. The value of fisheye\_circular\_region\_centre\_x[ i ] and fisheye\_circular\_region\_centre\_y[ i ] shall be in the range of 0 to 65 536 \* 216 − 1 (i.e., 4 294 967 295), inclusive.

**fisheye\_rect\_region\_top**[ i ], **fisheye\_rect\_region\_left**[ i ], **fisheye\_rect\_region\_width**[ i ], and **fisheye\_rect\_region\_height**[ i ] specify the coordinates of the top-left corner and the width and height of the i-th rectangular region that contains the i-th active area, in units of luma samples.

The value of fisheye\_rect\_region\_top[ i ] shall be in the range of SubHeightC \* conf\_win\_top\_offset to pic\_height\_in\_luma\_samples − ( SubHeightC \* conf\_win\_bottom\_offset + 1 ), inclusive.

The value of fisheye\_rect\_region\_left[ i ] shall be in the range of SubWidthC \* conf\_win\_left\_offset to pic\_width\_in\_luma\_samples − ( SubWidthC \* conf\_win\_right\_offset + 1 ), inclusive.

The value of fisheye\_rect\_region\_width[ i ] shall be in the range of 1 to pic\_width\_in\_luma\_samples − SubWidthC \* ( conf\_win\_left\_offset + conf\_win\_right\_offset ), inclusive.

The value of fisheye\_rect\_region\_height[ i ] shall be in the range of 1 to pic\_height\_in\_luma\_samples − SubHeightC \* ( conf\_win\_top\_offset + conf\_win\_bottom\_offset ), inclusive.

The sum of fisheye\_rect\_region\_top[ i ] and fisheye\_rect\_region\_height[ i ] shall be less than pic\_height\_in\_luma\_samples − SubHeightC \* conf\_win\_bottom\_offset.

The sum of fisheye\_rect\_region\_left[ i ] and fisheye\_rect\_region\_width[ i ] shall be less than pic\_width\_in\_luma\_samples − SubWidthC \* conf\_win\_right\_offset.

**fisheye\_circular\_region\_radius**[ i ] specifies the radius of the circular region that contains the i-th active area that is defined as a length from the centre of the circular region specified by fisheye\_circular\_region\_centre\_x[ i ] and fisheye\_circular\_region\_centre\_y[ i ] to the outermost pixel boundary of the circular region, in units of 2−16 luma samples, that corresponds to the maximum field of view of the i-th fisheye lens, specified by fisheye\_field\_of\_view[ i ]. The value of fisheye\_circular\_region\_radius[ i ] shall be in the range of 0 to 65 536 \* 216 − 1 (i.e., 4 294 967 295), inclusive.

The i-th active area is defined as the intersection of the i-th rectangular region, specified by fisheye\_rect\_region\_top[ i ], fisheye\_rect\_region\_left[ i ], fisheye\_rect\_region\_width[ i ], and fisheye\_rect\_region\_height[ i ], and the i-th circular region, specified by fisheye\_circular\_region\_centre\_x[ i ], fisheye\_circular\_region\_centre\_y[ i ], and fisheye\_circular\_region\_radius[ i ].

Each active area shall contain at least one sample location. There shall not be any sample location that is within more than one active area.

**fisheye\_scene\_radius**[ i ] specifies the radius of a circular region within the i-th active area in units of 2−16 luma samples, where the obstruction, such as the camera body, is not included in the region specified by fisheye\_circular\_region\_centre\_x[ i ], fisheye\_circular\_region\_centre\_y[ i ], and fisheye\_scene\_radius[ i ]. The value of fisheye\_scene\_radius[ i ] shall be less than or equal to fisheye\_circular\_region\_radius[ i ], and shall be in the range of 0 to 65 536 \* 216 − 1 (i.e., 4 294 967 295), inclusive. The enclosed area is the suggested area for stitching as recommended by the encoder.

**fisheye\_camera\_centre\_azimuth**[ i ] and **fisheye\_camera\_centre\_elevation**[ i ] indicate the spherical coordinates that correspond to the centre of the circular region that contains the i-th active area in the cropped output picture, in units of 2−16 degrees. The value of fisheye\_camera\_centre\_azimuth[ i ] shall be in the range of −180 \* 216 (i.e., −11 796 480) to 180 \* 216 − 1 (i.e., 11 796 479), inclusive, and the value of fisheye\_camera\_centre\_elevation[ i ] shall be in the range of −90 \* 216 (i.e., −5 898 240) to 90 \* 216 (i.e., 5 898 240), inclusive.

**fisheye\_camera\_centre\_tilt**[ i ] indicates the tilt angle of the sphere region that corresponds to the i-th active area of the cropped output picture, in units of 2−16 degrees. The value of fisheye\_camera\_centre\_tilt[ i ] shall be in the range of −180 \* 216 (i.e., −11 796 480) to 180 \* 216 − 1 (i.e., 11 796 479), inclusive.

**fisheye\_camera\_centre\_offset\_x**[ i ], **fisheye\_camera\_centre\_offset\_y**[ i ] and **fisheye\_camera\_centre\_offset\_z**[ i ] indicate the XYZ offset values, in units of 2−16 millimeters, of the focal centre of the fisheye camera lens corresponding to the i-th active area from the focal centre origin of the overall fisheye camera configuration. The value of each of fisheye\_camera\_centre\_offset\_x[ i ], fisheye\_camera\_centre\_offset\_y[ i ], and fisheye\_camera\_centre\_offset\_z[ i ] shall be in the range of 0 to 65 536 \* 216 − 1 (i.e., 4 294 967 295), inclusive.

**fisheye\_field\_of\_view**[ i ] specifies the field of view of the lens that corresponds to the i-th active area in the coded picture, in units of 2−16 degrees. The value of fisheye\_field\_of\_view[ i ] shall be in the range of 0 to 360 \* 216 (i.e., 23 592 960), inclusive.

**fisheye\_num\_polynomial\_coeffs**[ i ] specifies the number of polynomial coefficients for the circular region corresponding to the i-th active area. The value of fisheye\_num\_polynomial\_coeffs[ i ] shall be in the range of 0 to 8, inclusive. Values of fisheye\_num\_polynomial\_coeffs[ i ] greater than 8 are reserved for future use by ITU-T | ISO/IEC. Decoders encountering a fisheye video information SEI message with fisheye\_num\_polynomial\_coeffs[ i ] greater than 8 shall ignore the fisheye video information SEI message.

**fisheye\_polynomial\_coeff**[ i ][ j ] specifies the j-th polynomial coefficient value, in units of 2−24, of the curve function that maps the normalized distance of a luma sample from the centre of the circular region corresponding to the i-th active area to the angular value of a sphere coordinate from the normal vector of a nominal imaging plane that passes through the centre of the sphere coordinate system for the i-th active region. The value of fisheye\_polynomial\_coeff[ i ][ j ] shall be in the range of −128 \* 224 (i.e., 2 147 483 648) to 128 \* 224 − 1 (i.e., 2 147 483 647), inclusive.

*Renumber clause D.3.47 (Reserved SEI message semantics) as D.3.48.*

*Add clauses D.3.47 as follows:*

**D.3.47 Annotated regions SEI message semantics**

|  |  |
| --- | --- |
| annotated\_regions( payloadSize ) { | **Descriptor** |
| **ar\_cancel\_flag** | u(1) |
| if (!ar\_cancel\_flag) { |  |
| **ar\_not\_optimized\_for\_viewing\_flag** | u(1) |
| **ar\_true\_motion\_flag** | u(1) |
| **ar\_occluded\_object\_flag** | u(1) |
| **ar\_partial\_object\_flag\_present\_flag** | u(1) |
| **ar\_object\_label\_present\_flag** | u(1) |
| **ar\_object\_confidence\_info\_present\_flag** | u(1) |
| if( ar\_object\_confidence\_info\_present\_flag ) |  |
| **ar\_object\_confidence\_length\_minus1** | u(4) |
| if( ar\_object\_label\_present\_flag ) { |  |
| **ar\_object\_label\_language\_present\_flag** | u(1) |
| if( ar\_object\_label\_language\_present\_flag ) { |  |
| while( !byte\_aligned( ) ) |  |
| **ar\_bit\_equal\_to\_zero** /\* equal to 0 \*/ | f(1) |
| **ar\_object\_label\_language** | st(v) |
| } |  |
| **ar\_num\_label\_updates** | ue(v) |
| for( i = 0; i < ar\_num\_ label\_updates; i++ ) { |  |
| **ar\_label\_idx**[ i ] | ue(v) |
| **ar\_label\_cancel\_flag** | u(1) |
| LabelAssigned[ ar\_label\_idx[ i] ] = !ar\_label\_cancel\_flag |  |
| if ( !ar\_label\_cancel\_flag) { |  |
| while( !byte\_aligned( ) ) |  |
| **ar\_bit\_equal\_to\_zero** /\* equal to 0 \*/ | f(1) |
| **ar\_label**[ ar\_label\_idx[ i ] ] | st(v) |
| } |  |
| } |  |
| } |  |
| **ar\_num\_object\_updates** | ue(v) |
| for( i = 0; i  <=  ar\_num\_object\_updates; i++ ) { |  |
| **ar\_object\_idx**[ i ] | ue(v) |
| **ar\_object\_cancel\_flag** | u(1) |
| ObjectTracked[ ar\_object\_idx[ i ] ] = !ar\_object\_cancel\_flag |  |
| if (!ar\_object\_cancel\_flag) { |  |
| if( ar\_object\_label\_present\_flag ) { |  |
| **ar\_object\_label\_update\_flag** | u(1) |
| if( ar\_object\_label\_update\_flag ) |  |
| **ar\_object\_label\_idx**[ ar\_object\_idx[ i ] ] | ue(v) |
| } |  |
| **ar\_bounding\_box\_update\_flag** | u(1) |
| if( ar\_bounding\_box\_update\_flag { |  |
| **ar\_bounding\_box\_top[** ar\_object\_idx[ i ] ] | u(16) |
| **ar\_bounding\_box\_left**[ ar\_object\_idx[ i ] ] | u(16) |
| **ar\_bounding\_box\_width**[ ar\_object\_idx[ i ] ] | u(16) |
| **ar\_bounding\_box\_height**[ ar\_object\_idx[ i ] ] | u(16) |
| if( ar\_partial\_object\_flag\_present\_flag ) |  |
| **ar\_partial\_object\_flag**[ ar\_object\_idx[ i ] ] | u(1) |
| if( ar\_object\_confidence\_info\_present\_flag ) |  |
| **ar\_object\_confidence**[ ar\_object\_idx[ i ] ] | u(v) |
| } |  |
| } |  |
| } |  |
| } |  |
| } |  |

*Add clause D.3.47, as follows:*

**D.3.47 Annotated regions SEI message semantics**

The annotated regions SEI message carries parameters that identify annotated regions using bounding boxes representing the size and location of identified objects.

**ar\_cancel\_flag** equal to 1 indicates that the annotated regions SEI message cancels the persistence of any previous annotated regions SEI message that is associated with one or more layers to which the annotated regions SEI message applies. ar\_cancel\_flag equal to 0 indicates that annotated regions information follows.

When ar\_cancel\_flag equal to 1 or a new CLVS of the current layer begins, the variables LabelAssigned[ i ] and ObjectTracked[ i ] are set equal to 0 for i in 0..255.

Let picA be the current picture. Each region identified in the annotated regions SEI message persists for the current layer in output order until any of the following conditions are true:

– A new CLVS of the current layer begins.

– The bitstream ends.

– A picture picB in the current layer in an access unit containing an annotated regions SEI message that is applicable to the current layer is output for which PicOrderCnt( picB ) is greater than PicOrderCnt( picA ), where PicOrderCnt( picB ) and PicOrderCnt( picA ) are the PicOrderCntVal values of picB and picA, and the semantics of the annotated regions SEI message for PicB cancels the persistence of the region identified in the annotated regions SEI message for PicA.

**ar\_not\_optimized\_for\_viewing\_flag** equal to 1 indicates that the decoded pictures that the annotated regions SEI message applies to are not optimized for user viewing, but rather are optimized for some other purpose such as algorithmic object classification performance. ar\_not\_optimized\_for\_viewing\_flagequal to 0 indicates that the decoded pictures that the annotated regions SEI message applies to may or may not be optimized for user viewing.

**ar\_true\_motion\_flag** equal to 1 indicates that the motion information in the coded pictures that the annotated regions SEI message applies to was selected with a goal of accurately representing object motion for objects in the annotated regions. ar\_true\_motion\_flag equal to 0 indicates that the motion information in the coded pictures that the annotated regions SEI message applies to may or may not be selected with a goal of accurately representing object motion for objects in the annotated regions.

**ar\_occluded\_object\_flag** equal to 1 indicates that the ar\_bounding\_box\_top[ ar\_object\_idx[ i ] ], ar\_bounding\_box\_left[ ar\_object\_idx[ i ] ], ar\_bounding\_box\_width[ ar\_object\_idx[ i ] ], and ar\_bounding\_box\_height[ ar\_object\_idx[ i ] ] syntax elements each represent the size and location of an object or a portion of an object that may not be visible or may be only partially visible within the cropped decoded picture. ar\_occluded\_object\_flagequal to 0 indicates that the ar\_bounding\_box\_top[ ar\_object\_idx[ i ] ], ar\_bounding\_box\_left[ ar\_object\_idx[ i ] ], ar\_bounding\_box\_width[ ar\_object\_idx[ i ] ], and ar\_bounding\_box\_height[ ar\_object\_idx[ i ] ] syntax elements represent the size and location of an object that is entirely visible within the cropped decoded picture. It is a requirement of bitstream conformance that the value of ar\_occluded\_object\_flag shall be the same for all annotated\_regions( ) syntax structures within a CLVS.

**ar\_partial\_object\_flag\_present\_flag** equal to 1 indicates that ar\_partial\_object\_flag[ ar\_object\_idx[ i ] ] syntax elements are present. ar\_partial\_object\_flag\_present\_flag equal to 0 indicates that ar\_partial\_object\_flag[ ar\_object\_idx[ i ] ] syntax elements are not present. It is a requirement of bitstream conformance that the value of ar\_partial\_object\_flag\_present\_flag shall be the same for all annotated\_regions( ) syntax structures within a CLVS.

**ar\_object\_label\_present\_flag** equal to 1 indicates that label information corresponding to objects in the annotated regions is present. ar\_object\_label\_present\_flag equal to 0 indicates that label information corresponding to the objects in the annotated regions is not present.

**ar\_object\_confidence\_info\_present\_flag** equal to 1 indicates that ar\_object\_confidence[ ar\_object\_idx[ i ] ] syntax elements are present. ar\_object\_confidence\_info\_present\_flag equal to 0 indicates that ar\_object\_confidence[ ar\_object\_idx[ i ] ] syntax elements are not present. It is a requirement of bitstream conformance that the value of ar\_object\_confidence\_present\_flag shall be the same for all annotated\_regions( ) syntax structures within a CLVS.

**ar\_object\_confidence\_length\_**minus + 1 specifies the length, in bits, of the ar\_object\_confidence[ ar\_object\_idx[ i ] ] syntax elements. It is a requirement of bitstream conformance that the value of ar\_object\_confidence\_length shall be the same for all annotated\_regions( ) syntax structures within a CLVS.

**ar\_object\_label\_language\_present\_flag** equal to 1 indicates that the ar\_object\_label\_language syntax element is present. ar\_object\_label\_language\_present\_flag equal to 0 indicates that the ar\_object\_label\_language syntax element is not present.

**ar\_bit\_equal\_to\_zero** shall be equal to zero.

**ar\_object\_label\_language** contains a language tag as specified by IETF RFC 5646 followed by a null termination byte equal to 0x00. The length of the ar\_object\_label\_language syntax element shall be less than or equal to 255 bytes, not including the null termination byte. When not present, the language of the label is unspecified.

**ar\_num\_label\_updates** indicates the total number of labels associated with the annotated regions that will be signalled. The value of ar\_num\_label\_updates shall be in the range of 0 to 255, inclusive.

**ar\_label\_idx**[ i ] indicates the index of the signalled label . The value of ar\_label\_idx[ i ]shall be in the range of 0 to 255, inclusive.

**ar\_label\_cancel\_flag** equal to 1 cancels the persistence scope of the ar\_label\_idx[ i ]-th label. ar\_label\_cancel\_flag equal to 0 indicates that the ar\_label\_idx[ i ]-th label will be assigned a signalled value.

**ar\_label**[ ar\_label\_idx[ i ] ] specifies the contents of the ar\_label\_idx[ i ] –th label. The length of the ar\_label[ ar\_label\_idx[ i ] ] syntax element shall be less than or equal to 255 bytes, not including the null termination byte.

**ar\_num\_object\_updates** indicates the number of object updates to be signalled. ar\_num\_cancelled\_objects shall be in the range of 0 to 255, inclusive.

**ar\_object\_idx**[ i ] is the index of the object parameters to be signalled. ar\_object\_idx[ i ]shall be in the range of 0 to 255, inclusive.

**ar\_object\_cancel\_flag** equal to 1 cancels the persistence scope of the ar\_object\_idx[ i ]-th object. ar\_object\_cancel\_flag equal to 0 indicates that parameters associated with the ar\_object\_idx[ i ]-th object tracked object will be signalled..

**ar\_object\_label\_update\_flag** equal to 1 indicates that an object label will be signalled. ar\_object\_label\_update\_flag equal to 0 indicates that an object label will not will not be signalled.

**ar\_object\_label\_idx**[ ar\_object\_idx[ i ] ] indicates the index of the label corresponding to the ar\_object\_idx[ i ]-th object. When ar\_object\_label\_idx[ ar\_object\_idx[ i ] ] is not present, its value is inferred from a previous annotated regions SEI messages in output order in the same CLVS, if any.

**ar\_bounding\_box\_update\_flag** equal to 1 indicates that an object bounding box will be signalled. ar\_bounding\_box\_update\_flag equal to 0 indicates that an object bounding box will not be signalled.

**ar\_bounding\_box\_top**[ ar\_object\_idx[ i ] ], **ar\_bounding\_box\_left**[ ar\_object\_idx[ i ] ], **ar\_bounding\_box\_width**[ ar\_object\_idx[ i ] ] and **ar\_bounding\_box\_height**[ ar\_object\_idx[ i ] ] specify the coordinates of the top-left corner and the width and height, respectively, of the bounding box of the ar\_object\_idx[ i ]-th object in the cropped decoded picture, relative to the conformance cropping window specified by the active SPS.

Let croppedWidth and croppedHeight be the width and height, respectively, of the cropped decoded picture in units of luma samples, as specified by Equations D-28 and D-29.

The value of ar\_bounding\_box\_left[ ar\_object\_idx[ i ] ] shall be in the range of 0 to croppedWidth / SubWidthC − 1, inclusive.

The value of ar\_bounding\_box\_top[ ar\_object\_idx[ i ] ] shall be in the range of 0 to croppedHeight / SubHeightC − 1, inclusive.

The value of ar\_bounding\_box\_width[ ar\_object\_idx[ i ] ] shall be in the range of 0 to croppedWidth / SubHeightC − ar\_bounding\_box\_left[ ar\_object\_idx[ i ] ], inclusive.

The value of ar\_bounding\_box\_height[ ar\_object\_idx[ i ] ] shall be in the range of 0 to croppedHeight / SubHeightC − ar\_bounding\_box\_top[ ar\_object\_idx[ i ] ], inclusive.

The identified object rectangle contains the luma samples with horizontal picture coordinates from SubWidthC \* ( conf\_win\_left\_offset + ar\_bounding\_box\_left[ ar\_object\_idx[ i ] ] ) to SubWidthC \* ( conf\_win\_left\_offset + ar\_bounding\_box\_left[ ar\_object\_idx[ i ] ] + ar\_bounding\_box\_width[ ar\_object\_idx[ i ] ] ) − 1, inclusive, and vertical picture coordinates from SubHeightC \* ( conf\_win\_top\_offset + ar\_bounding\_box\_top[ ar\_object\_idx[ i ] ] ) to SubWidthC \* ( conf\_win\_top\_offset + ar\_bounding\_box\_top[ ar\_object\_idx[ i ] ] + ar\_bounding\_box\_height[ ar\_object\_idx[ i ] ] ) − 1, inclusive.

The values of ar\_bounding\_box\_top[ ar\_object\_idx[ i ] ], ar\_bounding\_box\_left[ ar\_object\_idx[ i ] ], ar\_bounding\_box\_width[ ar\_object\_idx[ i ] ] and ar\_bounding\_box\_height[ ar\_object\_idx[ i ] ] persist in output order within the CLVS for each value of ar\_object\_idx[ i ]. When not present, the values of ar\_bounding\_box\_top[ ar\_object\_idx[ i ] ], ar\_bounding\_box\_left[ ar\_object\_idx[ i ] ], ar\_bounding\_box\_width[ ar\_object\_idx[ i ] ] or ar\_bounding\_box\_height[ ar\_object\_idx[ i ] ] are inferred from a previous annotated regions SEI message in output order in the CLVS, if any.

**ar\_partial\_object\_flag**[ ar\_object\_idx[ i ] ] equal to 1 indicates that the ar\_bounding\_box\_top[ ar\_object\_idx[ i ] ], ar\_bounding\_box\_left[ ar\_object\_idx[ i ] ], ar\_bounding\_box\_width[ ar\_object\_idx[ i ] ] and ar\_bounding\_box\_height[ ar\_object\_idx[ i ] ] syntax elements represent the size and location of an object that is only partially visible within the cropped decoded picture. ar\_partial\_object\_flag[ ar\_object\_idx[ i ] ] equal to 0 indicates that the ar\_bounding\_box\_top[ ar\_object\_idx[ i ] ], ar\_bounding\_box\_left[ ar\_object\_idx[ i ] ], ar\_bounding\_box\_width[ ar\_object\_idx[ i ] ] and ar\_bounding\_box\_height[ ar\_object\_idx[ i ] ] syntax elements represent the size and location of an object that may or may not be only partially visible within the cropped decoded picture. When not present, the value of ar\_partial\_object\_flag[ ar\_object\_idx[ i ] ] is inferred from a previous annotated regions SEI message in output order in the CLVS, if any.

**ar\_object\_confidence**[ ar\_object\_idx[ i ] ] indicates the degree of confidence associated with the ar\_object\_idx[ i ]-th object, in units of 2−( ar\_object\_confidence\_length\_minus1 + 1 ), such that a higher value of **ar\_object\_confidence**[ ar\_object\_idx[ i ] ] indicates a higher degree of confidence. The length of the ar\_object\_confidence[ ar\_object\_idx[ i ] ] syntax element is ar\_object\_confidence\_length\_minus1 + 1 bits. When not present, the value of\_object\_confidence[ ar\_object\_idx[ i ] ] is inferred from a previous annotated regions SEI message in output order in the CLVS, if any.

*In F.14.3.1 (General SEI payload semantics), replace the following paragraphs:*

The list VclAssociatedSeiList is set to consist of the payloadType values 2, 3, 6, 9, 15, 16, 17, 19, 22, 23, 45, 47, 56, 128, 131, 132, 134 to 152, inclusive, 154 to 159, inclusive, 161, 165, 167, 168, and 200 to 201, inclusive.

The list PicUnitRepConSeiList is set to consist of the payloadType values 0, 1, 2, 6, 9, 15, 16, 17, 19, 22, 23, 45, 47, 56, 128, 129, 131, 132, 133, 135 to 152, inclusive, 154 to 168, inclusive, and 200 to 201, inclusive.

*with the following:*

The list VclAssociatedSeiList is set to consist of the payloadType values 2, 3, 6, 9, 15, 16, 17, 19, 22, 23, 45, 47, 56, 128, 131, 132, 134 to 152, inclusive, 154 to 159, inclusive, 161, 165, 167, 168, and 200 to 202, inclusive.

The list PicUnitRepConSeiList is set to consist of the payloadType values 0, 1, 2, 6, 9, 15, 16, 17, 19, 22, 23, 45, 47, 56, 128, 129, 131, 132, 133, 135 to 152, inclusive, 154 to 168, inclusive, and 200 to 202, inclusive.

*In G.14.3.1 (General SEI payload semantics), replace the following paragraphs:*

The list VclAssociatedSeiList is set to consist of payloadType values 2, 3, 6, 9, 15, 16, 17, 19, 22, 23, 45, 47, 56, 128, 131, 132, 134 to 152, inclusive, 154 to 159, inclusive, 161, 165, 167, 168, 177, 178, 179, and 200 to 201, inclusive.

The list PicUnitRepConSeiList is set to consist of payloadType values 0, 1, 2, 6, 9, 15, 16, 17, 19, 22, 23, 45, 47, 56, 128, 129, 131, 132, 133, 135 to 152, inclusive, 154 to 168, inclusive, 176 to 180, inclusive, and 200 to 201, inclusive.

*with the following:*

The list VclAssociatedSeiList is set to consist of payloadType values 2, 3, 6, 9, 15, 16, 17, 19, 22, 23, 45, 47, 56, 128, 131, 132, 134 to 152, inclusive, 154 to 159, inclusive, 161, 165, 167, 168, 177, 178, 179, and 200 to 202, inclusive.

The list PicUnitRepConSeiList is set to consist of payloadType values 0, 1, 2, 6, 9, 15, 16, 17, 19, 22, 23, 45, 47, 56, 128, 129, 131, 132, 133, 135 to 152, inclusive, 154 to 168, inclusive, 176 to 180, inclusive, and 200 to 202, inclusive.

*In I.14.3.1 (General SEI payload semantics), replace the following paragraphs:*

The list VclAssociatedSeiList is set to consist of payloadType values 2, 3, 6, 9, 15, 16, 17, 19, 22, 23, 45, 47, 56, 128, 131, 132, 134 to 152, inclusive, 154 to 159, inclusive, 161, 165, 167, 168, 177, 178, 179, and 200 to 201, inclusive.

The list PicUnitRepConSeiList is set to consist of payloadType values 0, 1, 2, 6, 9, 15, 16, 17, 19, 22, 23, 45, 47, 56, 128, 129, 131, 132, 133, 135 to 152, inclusive, 154 to 168, inclusive, 176 to 181, inclusive, and 200 to 201, inclusive.

*with the following:*

The list VclAssociatedSeiList is set to consist of payloadType values 2, 3, 6, 9, 15, 16, 17, 19, 22, 23, 45, 47, 56, 128, 131, 132, 134 to 152, inclusive, 154 to 159, inclusive, 161, 165, 167, 168, 177, 178, 179, and 200 to 202, inclusive.

The list PicUnitRepConSeiList is set to consist of payloadType values 0, 1, 2, 6, 9, 15, 16, 17, 19, 22, 23, 45, 47, 56, 128, 129, 131, 132, 133, 135 to 152, inclusive, 154 to 168, inclusive, 176 to 181, inclusive, and 200 to 202, inclusive.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_