|  |  |
| --- | --- |
| **Joint Collaborative Team on 3D Video Coding Extension Development**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  2nd Meeting: Shanghai, CN, 13–19 Oct. 2012 | Document: JCT3V-B0142r1 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **MVC+D: On the 3D reference displays information SEI message** | | |
| *Status:* | Input Document | | |
| *Purpose:* | Proposal | | |
| *Author(s) or Contact(s):* | Steffen Kamp  Matthias Narroschke  Thomas Wedi  Monzastraße 4c 63225 Langen Germany | Tel: Email: | +49 6103 766271 steffen.kamp@eu.panasonic.com |
| *Source:* | Panasonic Corporation | | |

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

# Abstract

An extension to the 3d\_reference\_displays\_info SEI message is proposed for the MVC Extension for Inclusion of Depth Maps, which allows the explicit coding of the view\_id values of the preferred views as part of a reference display configuration. If view\_id values for both left and right view are coded for a reference display configuration in the modified SEI, no camera baseline is coded as part of the reference display configuration, making it semantically independent of the depth\_acquisition\_info SEI message. The explicit signalling of either both view\_id values without camera baseline or one view\_id value in combination with a camera baseline can be used by decoders and rendering devices to decide which camera positions should be preferred when choosing a stereo pair. Additionally, a modification of the syntax of the horizontal shift parameter and a bug fix of its semantics are proposed.

# Introduction

The 3d\_reference\_displays\_info SEI message proposed in [1] and adopted to the MVC Extension for Inclusion of Depth Maps Draft [2] can help decoders and display devices in selecting appropriate stereo view pairs based on the actual viewing conditions. It allows coding of a set of reference viewing conditions, each containing a display width, a suggested camera baseline, optionally a horizontal shift suggested to be applied to one of the views, and also optionally the viewing distance between the viewer and the display. A display device that is aware of the actual viewing conditions would e.g. select the closest matching reference viewing conditions from the SEI and extract the camera baseline and shift parameters. Given the set of views coded in the bitstream and the extracted parameters, the display can select a pair of views for display, using either explicitly coded views or synthesized views.

# Problem description

The camera baseline distance in the 3d\_reference\_displays\_info SEI message is coded as floating point value expressing the distance in the same units as the horizontal translation parameter in the depth\_acquisition\_info SEI message. Therefore, in order for the decoder/renderer to interpret the camera baseline distance signalled in the 3d\_reference\_displays\_info SEI message (i.e. to map it to actual views or synthesised view positions), access units containing a 3d\_reference\_displays\_info SEI message are required to also contain a depth\_acquisition\_info SEI message.

Additionally, the 3d\_reference\_displays\_info SEI message does not allow explicit coding of the camera positions. Therefore, the 3d\_reference\_displays\_info SEI message does not give content providers control over which specific views should be used for display, given a certain viewing condition.

# Proposal 1 – Explicit view signalling

We propose to extend the 3d\_reference\_displays\_info SEI message by an optional explicit coding of the preferred left view and/or right view using their respective view\_id values. This allows three modes that can be selected individually for each reference display coded in the 3d\_reference\_displays\_info SEI message:

1. Both, view\_id for left and right preferred views are coded. Then, the camera baseline distance is not required anymore and is therefore not coded.  
   This would allow displays that are incapable of view synthesis to easily select two of the coded views for display. It is not necessary to parse and interpret depth\_acquisition\_info SEI messages in addition to the 3d\_reference\_displays\_info SEI.
2. Only the view\_id of the left or the right preferred view is coded explicitly. The camera baseline distance is then coded in addition.  
   The explicit coding of the view\_id of either the left or right view conveys the additional information about which of the views is most suitable for display (e.g. contains the most unobstructed view of a scene).
3. If neither left nor right preferred view is explicitly coded, only the camera baseline distance is coded.  
   This conveys the same information as the current 3d\_reference\_displays\_info SEI message.

## Syntax and Semantics

[Note: Proposed text changes relative to JCT3V-A1001 are highlighted in yellow. Fixes to typos in the current specification text are highlighted in green.]

**I.13.1.6 3D reference displays information SEI message syntax**

|  |  |  |
| --- | --- | --- |
| 3d\_reference\_displays\_info( payloadSize ) { | **C** | **Descriptor** |
| **prec\_ref\_baseline** | 5 | ue(v) |
| **prec\_ref\_display\_width** | 5 | ue(v) |
| **ref\_viewing\_distance\_flag** | 5 | u(1) |
| if( ref\_viewing\_distance\_flag ) |  |  |
| **prec\_ref\_viewing\_dist** | 5 | ue(v) |
| **num\_ref\_displays\_minus1** | 5 | ue(v) |
| numRefDisplays = num\_ref\_displays\_minus1 + 1 |  |  |
| for( i = 0; i < numRefDisplays; i++ ) { |  |  |
| **ref\_left\_view\_id\_present\_flag** | 5 | u(1) |
| if( ref\_left\_view\_id\_present\_flag ) |  |  |
| **ref\_left\_view\_id**[ i ] | 5 | ue(v) |
| **ref\_right\_view\_id\_present\_flag** | 5 | u(1) |
| if( ref\_right\_view\_id\_present\_flag ) |  |  |
| **ref\_right\_view\_id**[ i ] | 5 | ue(v) |
| if( !ref\_left\_view\_id\_present\_flag | | !ref\_right\_view\_id\_present\_flag ) { |  |  |
| **exponent\_ref\_baseline**[ i ] | 5 | u(6) |
| **mantissa\_ref\_baseline**[ i ] | 5 | u(v) |
| } |  |  |
| **exponent\_ref\_display\_width[ i ]** | 5 | u(6) |
| **mantissa\_ref\_display\_width[ i ]** | 5 | u(v) |
| if( ref\_viewing\_distance\_flag ) { |  |  |
| **exponent\_ref\_viewing\_distance[ i ]** | 5 | u(6) |
| **mantissa\_ref\_viewing\_distance[ i ]** | 5 | u(v) |
| **}** |  |  |
| **additional\_shift\_present\_flag**[ i ] | 5 | u(1) |
| if ( additional\_shift\_present[ i ] ) |  |  |
| **num\_pixel\_shift\_plus128**[ i ] | 5 | u(8) |
| } |  |  |
| } |  |  |

**I.13.2.6 3D reference displays information SEI message semantics**

When present, this SEI message shall be associated with an IDR access unit. A reference displays information message contains information about the reference display width(s) and reference viewing distance(s) as well as information about the corresponding baseline distance(s), preferred view(s) and additional horizontal image shift(s), which form a stereo-pair for the reference display width and the reference viewing distance. This information enables a view renderer to produce a proper stereo-pair for the target screen width and the viewing distance. The reference display width and viewing distance values are signalled in units of centimet~~er~~res. The reference baseline values shall be signalled in the same units as the x component of the translation vector in the ~~SEI message~~ depth acquisition information SEI message that is valid for the same access unit. When a reference displays information SEI message containing reference baseline information is present in an access unit, the depth acquisition information SEI message shall also be present in the same access unit. ~~The baseline and shift information signaled for the reference display is valid for all access units they associated with and until the next IDR access unit or the next access unit containing depth acquisition information SEI message or reference displays information SEI message.~~ The information signalled in the SEI message applies to the coded video sequence. If an access unit following, in decoding order, the access unit the SEI message is associated with contains a depth acquisition information SEI message, the baseline information of the reference displays, if present in the SEI message, should be updated accordingly.

NOTE – The reference displays information SEI message specifies display parameters for which the 3D sequence was optimized and the corresponding reference parameters. Each reference display (i.e. a reference display width and possibly a corresponding viewing distance) is associated with one reference baseline distance and optionally one preferred left or right view, or no reference baseline distance and one preferred left view and one preferred right view.

[…]

**num\_ref\_displays\_minus1** plus 1 specifies the number of reference displays that are signalled in the bitstream. The value of num\_ref\_displays\_minus1shall be in the range of 0 to 31, inclusive.

**ref\_left\_view\_id\_present\_flag** equal to 1 indicates that the syntax element ref\_left\_view\_id[ i ] is present. ref\_left\_view\_id\_present\_flagequal to 0 indicates that the syntax element ref\_left\_view\_id[ i ] is not present.

**ref\_left\_view\_id[ i ]** specifies that the view with view\_id equal to ref\_left\_view\_id[ i ] is the preferred left view for the i-th reference display. The value of ref\_left\_view\_id[ i ] shall be in the range of 0 to 1023, inclusive. When ref\_left\_view\_id[ i ] is not present, its value is unspecified.

**ref\_right\_view\_id\_present\_flag[ i ]** equal to 1 indicates that the syntax element ref\_right\_view\_id[ i ] is present. ref\_left\_view\_id\_present\_flagequal to 0 indicates that the syntax element ref\_right\_view\_id[ i ] is not present.

**ref\_right\_view\_id[ i ]** specifies that the view with view\_id equal to ref\_right\_view\_id[ i ] is the preferred right view for the i-th reference display. The value of ref\_right\_view\_id[ i ] shall be in the range of 0 to 1023, inclusive. When ref\_right\_view\_id[ i ] is not present, its value is unspecified.

**exponent\_ref\_baseline[ i ]** specifies the exponent part of the reference baseline ~~of the~~ for the i-th reference display. The value of exponent\_ref\_baseline[ i ] shall be in the range of 0 to 62, inclusive. The value 63 is reserved for future use by ITU‑T | ISO/IEC. Decoders shall treat the value 63 as indicating an unspecified reference baseline. If exponent\_ref\_baseline[ i ] is not present, the reference baseline for the i-th reference display is unspecified.

**mantissa\_ref\_baseline[ i ]** specifies the mantissa part of the reference baseline ~~of the~~ for the i-th reference display. If mantissa\_ref\_baseline[ i ] is not present, the reference baseline for the i-th reference display is unspecified. Otherwise, the length of the mantissa\_ref\_baseline[ i ] syntax element is variable and determined as follows.

# Proposal 2 – Improved coding of horizontal shift

In the current 3d\_reference\_displays\_info SEI message, the horizontal shift parameter num\_pixel\_shift\_plus128 is coded as 8-bit wide unsigned value, resulting in an allowed value range of [−127, 128] which may be insufficient to represent required horizontal shift values. One example would be a 2-view video signal with 1920 columns per view and a parallel camera setup (i.e. a point at infinity in 3D scene coordinates would end up at the same 2D view coordinates in both views). The maximum shift that can be signalled corresponds to 1/15 of the video width or equivalently 1/15 of the display screen width. If the sequence was shown on a handheld tablet with a screen width of 20 cm, the maximum possible shift would be 1.33 cm. With the assumption that the display is held at a distance of 40 cm from the eyes and an interpupillary distance of 6.5 cm, points at infinity in scene coordinates are projected to a depth of

which is just 10 cm behind the screen. For higher resolution sequences, this maximum depth is further reduced.

It is therefore proposed to code the horizontal shift value using a signed Exp-Golomb value which allows values exceeding the [−127, 128] range and restricting the value range based on the sequence width.

## Originally proposed Syntax and Semantics

The syntax and semantics in this section were originally proposed in this document. During discussion in the BoG on high-level syntax, different syntax and semantics were agreed upon which is found in section 4.2.

[Note: Proposed text changes relative to JCT3V-A1001 are highlighted in yellow.]

**I.13.1.6 3D reference displays information SEI message syntax**

|  |  |  |
| --- | --- | --- |
| 3d\_reference\_displays\_info( payloadSize ) { | **C** | **Descriptor** |
| **prec\_ref\_baseline** | 5 | ue(v) |
| **prec\_ref\_display\_width** | 5 | ue(v) |
| **ref\_viewing\_distance\_flag** | 5 | u(1) |
| if( ref\_viewing\_distance\_flag ) |  |  |
| **prec\_ref\_viewing\_dist** | 5 | ue(v) |
| **num\_ref\_displays\_minus1** | 5 | ue(v) |
| numRefDisplays = num\_ref\_displays\_minus1 + 1 |  |  |
| for( i = 0; i < numRefDisplays; i++ ) { |  |  |
| **exponent\_ref\_baseline[ i ]** | 5 | u(6) |
| **mantissa\_ref\_baseline[ i ]** | 5 | u(v) |
| **exponent\_ref\_display\_width[ i ]** | 5 | u(6) |
| **mantissa\_ref\_display\_width[ i ]** | 5 | u(v) |
| if( ref\_viewing\_distance\_flag ) { |  |  |
| **exponent\_ref\_viewing\_distance[ i ]** | 5 | u(6) |
| **mantissa\_ref\_viewing\_distance[ i ]** | 5 | u(v) |
| **}** |  |  |
| **~~additional\_shift\_present\_flag~~**~~[ i ]~~ | ~~5~~ | ~~u(1)~~ |
| ~~if ( additional\_shift\_present[ i ] )~~ |  |  |
| **~~num\_pixel\_shift\_plus128~~**~~[ i ]~~ | ~~5~~ | ~~u(8)~~ |
| **num\_pixel\_shift**[ i ] | 5 | se(v) |
| } |  |  |
| } |  |  |

**I.13.2.6 3D reference displays information SEI message semantics**

**~~additional\_shift\_present\_flag[ i ]~~** ~~equal to 1 indicates that the information about additional horizontal shift of the left and right views for the i-th reference display is present in the bitstream. additional\_shift\_present\_flag[ i ] equal to 0 indicates that the information about additional horizontal shift of the left and right views for the i-th reference display is not present in the bitstream.~~

**num\_pixel\_shift~~\_plus128~~[ i ]** indicates the recommended additional horizontal shift for a stereo-pair corresponding to the i-th reference baseline and the i-th reference display. If ~~(~~num\_pixel\_shift~~\_plus128~~[ i ]~~− 128 )~~ is less than 0, it is recommended that the left view of the stereo-pair corresponding to the i-th reference baseline and the i-th reference display is shifted in the left direction by ~~( 128~~−num\_pixel\_shift~~\_plus128~~[ i ]~~)~~ pixels with respect to the right view of the stereo-pair; if num\_pixel\_shift~~\_plus128~~[ i ] is equal to ~~128~~ 0, it is recommended that shifting is not applied; if ~~(~~num\_pixel\_shift~~\_plus128~~[ i ]~~− 128 )~~ is greater than 0, it is recommended that the left view of the stereo-pair corresponding to the i-th reference baseline and the i-th reference display should be shifted in the right direction by ~~(~~num\_pixel\_shift~~\_plus128~~[ i ] ~~− 128 )~~ pixels with respect to the right view of the stereo-pair. The value of num\_pixel\_shift~~\_plus128~~[ i ] shall be in the range of ~~0 to 255~~ −16 \* ( pic\_width\_in\_mbs\_minus1 + 1 ) to 16 \* ( pic\_width\_in\_mbs\_minus1 + 1 ), inclusive.

## Syntax and Semantics as decided by the BoG on high-level syntax

During discussion of the proposal in the BoG on high-level syntax on Monday, 16th October, 16:00 – 18:00h the BoG recommended to adopt the second aspect of the proposal in spirit but using the following different syntax and semantics.

[Note: Text changes relative to JCT3V-A1001 are highlighted in yellow.]

**I.13.1.6 3D reference displays information SEI message syntax**

|  |  |  |
| --- | --- | --- |
| 3d\_reference\_displays\_info( payloadSize ) { | **C** | **Descriptor** |
| **prec\_ref\_baseline** | 5 | ue(v) |
| **prec\_ref\_display\_width** | 5 | ue(v) |
| **ref\_viewing\_distance\_flag** | 5 | u(1) |
| if( ref\_viewing\_distance\_flag ) |  |  |
| **prec\_ref\_viewing\_dist** | 5 | ue(v) |
| **num\_ref\_displays\_minus1** | 5 | ue(v) |
| numRefDisplays = num\_ref\_displays\_minus1 + 1 |  |  |
| for( i = 0; i < numRefDisplays; i++ ) { |  |  |
| **exponent\_ref\_baseline[ i ]** | 5 | u(6) |
| **mantissa\_ref\_baseline[ i ]** | 5 | u(v) |
| **exponent\_ref\_display\_width[ i ]** | 5 | u(6) |
| **mantissa\_ref\_display\_width[ i ]** | 5 | u(v) |
| if( ref\_viewing\_distance\_flag ) { |  |  |
| **exponent\_ref\_viewing\_distance[ i ]** | 5 | u(6) |
| **mantissa\_ref\_viewing\_distance[ i ]** | 5 | u(v) |
| **}** |  |  |
| **additional\_shift\_present\_flag**[ i ] | 5 | u(1) |
| if ( additional\_shift\_present[ i ] ) |  |  |
| **num\_pixel\_shift\_plus512**[ i ] | 5 | u(10) |
| } |  |  |
| **3d\_reference\_displays\_extension\_flag** | 5 | u(1) |
| } |  |  |

**I.13.2.6 3D reference displays information SEI message semantics**

**num\_pixel\_shift\_plus512[ i ]** indicates the recommended additional horizontal shift for a stereo-pair corresponding to the i-th reference baseline and the i-th reference display. If ( num\_pixel\_shift\_plus512[ i ] − 512 ) is less than 0, it is recommended that the left view of the stereo-pair corresponding to the i-th reference baseline and the i-th reference display is shifted in the left direction by ( 512 − num\_pixel\_shift\_plus512[ i ] ) pixels with respect to the right view of the stereo-pair; if num\_pixel\_shift\_plus512[ i ] is equal to 512, it is recommended that shifting is not applied; if ( num\_pixel\_shift\_plus512[ i ] − 512 ) is greater than 0, it is recommended that the left view of the stereo-pair corresponding to the i-th reference baseline and the i-th reference display should be shifted in the right direction by ( num\_pixel\_shift\_plus512[ i ] − 512 ) pixels with respect to the right view of the stereo-pair. The value of num\_pixel\_shift\_plus512[ i ] shall be in the range of 0 to 1023, inclusive.

**3d\_reference\_displays\_extension\_flag** equal to 0 indicates that no additional data follows within the reference displays SEI message. The value of 3d\_reference\_displays\_extension\_flag shall be equal to 0. The value of 1 for 3d\_reference\_displays\_extension\_flag is reserved for future use by ITU-T | ISO/IEC. Decoders shall ignore all data that follows the value of 1 for 3d\_reference\_displays\_extension\_flag in a reference displays SEI message.

# Proposal 3 – Bug fix for the horizontal shift parameter semantics

As part of a reference display configuration, the suggested camera baseline distance and a horizontal shift parameter are coded. In Section I.13.2.6 there is a note detailing what adjustments a decoder should perform, if a depth\_acquisition\_info SEI message is received in an access unit following the access unit containing the 3d\_reference\_displays\_info SEI in decoding order. The last sentence of this note suggests that the horizontal shift parameter should be scaled according to the camera baseline ratio if the camera baseline distance between views is changed by the newly received depth\_acquisition\_info SEI message. We believe that this is an error in the specification and that the scaling should not actually be performed. The input documents to previous JCT-3V or MPEG meetings leading to the inclusion of the 3d\_reference\_displays\_info SEI message also do not suggest that such a scaling is necessary.

Take, for example, the case where according to the 3d\_reference\_displays\_info SEI the view with view\_id equal to 0 is used as left view and the view with view\_id equal to 1 is used as right view. Now, if a scene change occurs and a different camera setup has been used for the shot (i.e. different camera baselines), a depth\_acquisition\_info SEI reflecting the updated camera setup is transmitted. However, the rendering device should keep using the views with view\_id 0 and 1 for left and right view, as it can be expected that the content was produced such that no view switching is required if the viewing conditions do not change. The scaling in the Note in Section I.13.2.6 is required because the suggested camera baseline distance as coded in the 3d\_reference\_displays\_info should be updated to reflect the new camera setup after the scene change. Staying with the current example, the result of the suggested scaling is the continued usage of views with view\_id 0 and 1 for left and right view. According to this understanding, the horizontal shift value should be kept constant when the camera setup is changed (just as the left and right views should continue being view\_id 0 and 1).

## Semantics

[Note: Proposed text changes relative to JCT3V-A1001 are highlighted in yellow. Fixes to typos in the current specification text are highlighted in green.]

**I.13.2.6 3D reference displays information SEI message semantics**

[…]

NOTE – […]

When camera parameters are updated by a depth acquisition information SEI message in a following access unit and the baseline between the views used in the view synthesis process in the following access unit changes relative to that ~~in the~~ in the access unit which the reference displays information SEI belongs to, the baseline ~~and the horizontal shift~~ for the receiver's display in the following access unit should be modified accordingly. Let the scaling factor *s* be equal to the ratio of the baseline between two views in the following access unit and the baseline between the same two views in the access unit, which the reference displays information SEI message belongs to, where the two views are used in the view synthesis process. Then the baseline distance for the receiver's display in the following access unit should be modified with the scaling factor *s* relative to the baseline distance for the receiver's display in the access unit which the reference displays information SEI belongs to. ~~The hortiontal shift for the receiver’s display should also be modified by scaling it with the same factor as that used to scale the baseline distance.~~

# Conclusion

The proposed explicit coding of preferred views as part of the reference display configuration in the 3d\_reference\_displays\_info SEI message allows a coding of the SEI that makes it semantically independent of the depth\_acquisition\_info SEI message. The coding of explicit view\_id values furthermore helps decoders or rendering devices to decide which views from a MVC stream to choose for displaying a stereo pair. Additionally, an improved syntax for the horizontal shift parameter and a bug fix to its semantics are proposed. We suggest adopting the proposed modifications to the MVC Extension for Inclusion of Depth Maps draft text.

# References

1. A. Norkin, I. Girdziajuskas, Y. Zhao, Y. Lu, “Show-case and syntax for SEI message on reference display information signaling”, JCT3V-A0163, Stockholm, 2012.
2. T. Suzuki, S. Hattori, M. Hannuksela, Y. Chen, G. Sullivan, “MVC Extension for Inclusion of Depth Maps Draft Text 4”, JCT3V-A1001, Stockholm, 2012.

# Patent rights declaration(s)

**Panasonic Corporation may have current or pending patent rights relating to the technology described in this contribution and, conditioned on reciprocity, is prepared to grant licenses under reasonable and non-discriminatory terms as necessary for implementation of the resulting ITU-T Recommendation | ISO/IEC International Standard (per box 2 of the ITU-T/ITU-R/ISO/IEC patent statement and licensing declaration form).**