|  |  |
| --- | --- |
| **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**  3rd Meeting: Geneva, CH, 17–23 Jan. 2013 | Document: JCT3V-C0239 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **BoG Report on 3D High level syntax** | | |
| *Status:* | Input Document | | |
| *Purpose:* | Report | | |
| *Author(s) or Contact(s):* | Ying Chen  5775 Morehouse Drive San Diego, CA 92121 USA  Thomas Rusert  Ericsson Research 164 80 Stockholm Sweden | Tel: Email: | +1-858-845-6589 [cheny@qti.qualcomm.com](mailto:cheny@qti.qualcomm.com)  [thomas.rusert@ericsson.com](mailto:thomas.rusert@ericsson.com) |
| *Source:* | AHG | | |

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

# Abstract

BoG meetings have been organized by JCT-3V on reviewing 3D high level syntax (HLS) proposals. The results from the joint BoG report of JCT-3V and JCT-VC on HLS have also been presented in this BoG.

The first BoG meeting was held on Jan. 20th, 2013, 6:00 pm ~ 9:00 pm.

The second BoG meeting was held on Jan. 21th, 2013, 9:45 am ~ 11:45 am.

# Scope

The MVC+D proposals are not reviewed in this BoG.

Proposals discussed in the joint BoG report of JCT-3V and JCT-VC on HLS have not reviewed in the HLS. However, results of the joint BoG have been presented in this BoG and agreed.

Other MV-HEVC input documents have been introduced in this document.

# Proposal review

**JCT3V-C0082 AHG7: Reference picture marking process for MV-HEVC B. Choi, M.W. Park, J. Yoon, C. Kim, J. Park (Samsung)**

A modified reference picture marking process for inter-view reference pictures is proposed. By marking the inter-view reference pictures as “used for inter-view reference” explicitly, the inter-view reference pictures can be handled in distinction from other inter-prediction reference pictures marked as “used for short-term reference” or “used for long-term reference”.

The proposed marking process ensures that the inter-view prediction is done among view components of an access unit without incorrect references and reference picture marking errors.

To find out the correct inter-view reference picture, all reference pictures marked as “used for short-term reference” shall be inspected from the decoded picture buffer.

The inter-view reference picture, once marked as “long-term” cannot be marked as “short-term” again based on the current HEVC design.

A current decoded picture is always firstly marked as “inter-view reference”, when used as a reference picture after RPLC, it is marked as long-term.

It is marked as short-term only after it has been indicated as needed by RPS.

The claimed benefits are:

Fast access of the inter-view reference pictures in the DPB.

It seems to be a bug fix is needed, to mark an inter-view reference picture status from “long-term” status back to “short-term” before it is used as temporal reference.

A comment was given that the “fast access” is an implementation issue.

A comment is given that the proposal might change the timing of the inter-view reference pictures when the marking is delayed. This proposal might be related to HRD modifications.

Revisit: after a potentially easy editorial bug fix is provided and this solution can be compared with the bug fix.

This issue has been addressed due to the discussions on JCT3V-C0146.

**JCT3V-C0146 Unification of scalable and multi-view extensions with HLS only changes, K. Ugur, M M. Hannuksela, J. Lainema, D. Rusanovskyy (Nokia)**

In this contribution, a high-level syntax only scalable extension standardization track for HEVC is proposed, in addition to a higher performing track that includes low level changes. It is further proposed to unify the high level syntax only tracks for scalable and multi-view coding in a single extension of HEVC. The goal is to support both scalable and multiview use-cases with no low level changes to HEVC in a single extension, so that HEVC hardware encoders/decoders could be re-used with firmware updates.

In revision 1, the straw-man specification text was updated so that changes to HEVC v1 are no longer proposed.

In JCT-3V, the 3D extensions of HEVC are currently being developed under the following tracks [JCT3V-B1006]:

* MV-HEVC: In this extension, multiple views can be coded with HEVC by extending the high-level syntax appropriately, and by rearrangement of decoded picture buffers to store the reference pictures as needed, without any changes to the core of the coding layer below the level of coded tree blocks (CTB).
* 3D-HEVC: In this design the inter-component dependencies between texture and depth and are exploited and texture and depth data are jointly coded.

In JCT-VC, the work on scalable extension of HEVC was started at the last meeting and two main approaches are currently being tested in tool experiments:

* Reference index based: In this approach, the upsampled base layer is made available for EL pictures by placing it in the enhancement layer DPB.
* Block based: In this approach, the base layer samples and syntax can be used to predict the enhancement layer and it is indicated at block level (either CU or PU level).

It is argued that the reference index based approach of SHVC could be supported with small changes to the current MV-HEVC draft text. In addition, it could be further harmonized with the MV-HEVC extension. This way, using a single extension and with no low level changes, both scalability and multiview use-cases can be supported. The benefits of this approach are as follows:

* Reduce market confusion by bringing a single high-level syntax extension instead of multiple extensions.
* Support scalability with high-level syntax only changes.
* Support mixed resolution multi-view coding (as mentioned in MPEG requirements document N12956, section 2.12.3), which is currently not possible in MV-HEVC.

Support resolution enhancement of multi-view content (as mentioned in MPEG requirements document N12956, section 2.12.3), which is currently not possible in MV-HEVC and SHVC.

The main features of the provided specification text can be summarized as follows:

1. The capability of using long-term reference pictures across layers, as motivated in JCTVC-L0170, is realized with the syntax extension mechanisms for sequence parameter set and slice segment header. The decoding of HEVC v1 bitstreams remains unchanged, while the decoding of the long-term reference picture set of scalable HEVC bitstreams is affected (in all layers).
2. Video parameter set design is taken from JCTVC-K1007 with the exclusion of standards scalability support (which was believed to be less mature than the other features).

See notes related to the joint BoG.

1. Activation for picture and sequence parameter sets for individual layers was added similarly to the SVC and MVC.

Parameter set activation. Text from MVC was integrated into the MV-HEVC.

Each layer activates one sequence parameter set (SPS).

Several layers (including the base layer) may share the same active SPS. Note that it was discussed that when the profile/level information is present in the VPS, the profile defined in the VPS applicable to the non-base view and may override that in the SPS.

Recommendation: Adopt the parameter set activation into the MV-HEVC with necessary alignment to the MV-HEVC.

1. Enhancement layer RAP picture behavior as proposed in JCTVC-L0039.

See notes related to the joint BoG.

1. The (short-term) inter-layer reference pictures are deduced from the cross-layer dependencies indicated in the VPS and appended at the end of both initial reference picture lists.

Aspect related to reference picture set, it is proposed that inter-view reference picture set is not created and are directly used to be added into the reference picture list.

It was commented this change may be editorial only. Several experts didn’t agree.

Further study.

1. Short-term inter-layer reference pictures are temporarily marked as “used for long-term reference” for the decoding of the current picture and marked back as “used for short-term reference” after the decoding of the current picture, as envisioned in several earlier contributions such as JCTVC-J0071.

Reference picture lists are checked based on layer\_id and marked back to short-term when decoding each slice for non-base view.

It was noted that after decoding a current view component, all the inter-view reference pictures can be marked as short-term.

Recommendation: Agreed: to include the above bug fix into MV-HEVC WD.

1. Non-reference pictures at the highest decoded temporal sub-layer are marked as “unused for reference” immediately after their decoding to enable reduction of the DPB usage.

In MVC DPB operation, similar optimization was done.

The view dependency is parsed to decide for each view, whether it is used for inter-view reference or not.

The view dependency is parsed to decide for each view, whether it is used for inter-view reference or not.

Recommendation: Adopt the proposed DPB size optimization scheme.

1. The DPB size and other characteristics (e.g. sps\_max\_num\_reorder\_pics) is indicated for each layer separately in the active (layer) sequence parameter set for that layer. The DPB operates separately for each layer except for the bumping process of the output order DPB which operates across layers. The bumping process outputs consecutively, in ascending nuh\_layer\_id order, the pictures having the same smallest POC value among all the pictures (marked as “needed for output”) in the DPB.

This sounds beneficial. It is not aligned with MVC.

Revisit after offline discussions.

Other technical aspects of this proposal

1. Change the “view dependency change SEI” to “layer dependency SEI”. The view dependency change SEI message of MV-HEVC was changed to equivalent layer dependency SEI message, which applies to all types of scalability.

Change the “view dependency change SEI” to “layer dependency SEI”.

Recommendation: Agreed: A constraint is introduced such that no non-present view dependency as indicated in the previous SEI message of the same type is included in the current SEI message.

1. Signaling of DPB size. The number of DPB frames used in HRD is proposed to be the number signalled in the active layer SPS. This value is separate from that used for base view.

It was commented that it might be useful to have the total number of the DPB frames for an operation point to be considered in the HRD.

This value might be related to level definitions.

Recommendation: Agreed: number of DPB frames used in HRD for the enhancement view is signalled in the active layer SPS.

Other discussions.

It was discussed which software may be potentially chosen for the unified design.

It was commented that MV-HEVC software is a part of 3D-HEVC software. It sounds important for MV-HEVC software to be the basis of 3D-HEVC in terms of coding performance evaluation.

Higher priority in terms of software maintenance is to migrate the software to the latest version of HM, compared with unifying MV-HEVC software and SHVC software.

It was suggested that the proponent may integrate editorial changes of this proposal to MV-HEVC.

**JCT3V-C0079 AHG7: On initialization process for reference picture lists O. Nakagami, Y. Takahashi, T. Suzuki (Sony)**In the current MV-HEVC draft (JCT3V-B1004\_d0), inter-view reference picture is appended at the end of temporal reference picture list. It is noted that reference picture list modification (RPLM) syntax is necessary to put the inter-view reference forward.

This contribution proposes to modify the initialization process for reference picture lists. An inter-view reference picture is inserted between RefPicSetStCurrBefore and RefPicSetStCurrAfter pictures in creating the temporal list. Since it is a semantics change, no syntax is added to the base spec. The change enables to skip RPLM signalling in the common reference picture structure (e.g. B-pictures in traditional N15M3 GOP structure) without extra syntax.

The proposal is implemented on top of HTM5.0.1. It is reported the BD-rate difference in common test condition is 0.0%, -0.1%, -0.1% and -0.1% for video0, video1, video2 and coded & synthesized, respectively. The difference comes from RPLM signalling bit reduction in the dependent view.

This proposal makes it more probable that the initial reference picture list becomes the final reference picture list.

Discussed together with C0060, see notes under C0060.

**JCT3V-C0060 AHG7: Reference picture list initialization for MV-HEVC, A.K. Ramasubramonian, L.Zhang, Y. Chen, Y.-K. Wang(Qualcomm)**

A reference picture list initialization method is proposed that signals the starting position of the inter-view reference pictures in the initial reference picture list. Signalling the starting position in the initial list for inter-view reference pictures is reported to avoid signalling reference picture list modification syntax in most cases. The proposal also reports modest BD-rate decrease under common test conditions.

Modification proposed to the initial reference picture list construction.

Proposed changes include:

A new syntax element in the slice header extension, at the bottom of the slice segment header

In the decoding process, insert the inter-view reference pictures at the signaled position

Changes only for list0, no changes to list 1.

It was commented that this was proposed already at the last meeting.

The reference picture list modification commands are sent before the proposed new syntax element. A question was asked why it was not included in the beginning of the slice header. The proponent responded that location could be too costly.

Another comment was made that this may have an impact on implementations of slice header parsing and reference picture list modification based on the HEVC base specification. It was responded that slice header parsing and reference picture list construction would typically be separated.

Question was asked how the signaled location was derived at the encoder. It was responded that it was selected using the same scheme as the RPLM in the current software.

Comparison of C0079 and C0060:

* Neither has changes to list1 construction
* Similar coding efficiency gains under CTC
* C0060 introduces a new syntax element in the slice segment header extension. It allows omit all reference picture list modification commands under CTC.
* C0079 does not introduce new syntax elements. For the CTC it is still necessary to send some RPLM commands.

It was suggested that it may be beneficial to combine both approaches. When the C0060 syntax element is equal to 0, the C0079 approach would be used, otherwise the C0060 approach would be used. It was suggested that such combined functionality should be tested.

It is claimed that in configurations other than CTC, use of C0079 only would require significantly more use of RPLM commands than with C0060, which could lead to coding efficiency penalty.

It was suggested that a more balanced approach to handle also list1 may be desirable, and that this might need testing with an I-B-P configuration.

It was agreed that a combination of C0060 and C0079 may be useful, however coding efficiency results should be provided.

Action: Offline discussion to combine C0060 and C0079. Revisit combined approach subject to availability of a solution for list1 construction and testing results for both CTC and I-B-P.

**JCT3V-C0061 AHG7: Slice header prediction for MV-HEVC, A.K. Ramasubramonian, Y. Chen, Y.-K. Wang(Qualcomm)**

The values of many slice header parameters of the view component of the non-base views in an access unit are identical to those of the base view in the same access unit. A slice header prediction scheme is proposed in this document in order to take advantage of this observation. In the proposed scheme, for non-base views, some of the slice header parameters may not be explicitly present, but are rather predicted or inferred from the first slice header of the view component of the base view.

Slice header parameters are grouped into 4 groups, and the presence of syntax elements in each of the groups is gated by a flag.

* Common information (including pps\_id, poc, rps)
* Reference picture list information (list sizes, RPLM)
* Deblocking parameters
* Prediction weight table elements

Some similar technique exists in 3D-AVC. It is proposed for MV-HEVC, but may also apply to 3D-HEVC. Similar functionality could also be used in SHVC (JCTVC-L0231). It was commented that JCT3V-C0223 (which is proposed for 3D-HEVC) is also related.

This seems to be an item that is of common interest for both 3D and scalable extensions.

The prediction is done only between view components, not between slices within the same view component. Reference view component is always the first slice of the base view. It was commented that intuitively, the reference might be selected to be according to the layer dependencies signaled in the VPS.

Benefit of the method is coding efficiency. No coding efficiency numbers have been provided,. It was commented that this method may particularly be useful when MTU size limits are present.

There was discussion on potential implications on error resilience. It was commented that it may be a problem that the first slice of the reference view needs to be present. If the first slice of the reference view component is lost, error concealment needs to be applied. A possible solution would be to require that the slice segment headers of the base view are identical.

Action: Further study, including analysis of coding efficiency gains and other benefits (MTU size matching), as well as error resilience aspects. This should also be considered in the context of SHVC.

**JCT3V-C0062 AHG7: Parallel decoding SEI for MV-HEVC, Y. Chen, V. Seregin, A. K.Ramasubramonian, L. Zhang, Y.-K. Wang (Qualcomm)**

A parallel decoding SEI message for MV-HEVC similar to the parallel decoding SEI message in MVC is proposed with three modifications: 1) the delay required for parallel decoding is signalled in unit of coding tree units (CTU); 2) the horizontal CTU delay is also signalled to avoid large increased delay caused by the size of a CTU; 3) the delay is signalled once for all the non-base views and is applicable to any view that utilizes inter-view prediction.

MVC has parallel decoding SEI, specifying how many MB rows delay are required for base and dependent view. This is signaled for each dependent view.

A worst case assumption for the decoder implementation is made. That requires that the reference block has been fully reconstructed (transform, deblocking, SAO). In the worst-case assumption, SAO requires that deblocking of the neighboring blocks to the reference block is done. Further in the worst-case assumption, the further neighboring blocks need to be at least transformed. Minimum delay is two CTUs horizontally and two vertically.

The decoding of the dependent view can start after the vertical delay of 3 CTUs rows or after a vertical delay of 2 rows and additional horizontal delay of 2 CTUs, assuming that no vertical disparity vector components are present. If vertical disparity vector components are present, then the SEI message would indicate a higher delay, which is a function of the maximum vertical disparity vector component.

There was a question what would happen if the CTU sizes would not be the same across views. The proponent responded that the worst case CTU size could be assumed. Alternatively a loop over the view components could be present.

Another question was how this would deal with tile partitioning in the reference view. It was commented that this case was not covered by the proposed SEI. It could be further studied if something more useful could be signaled if tiles were present.

There is a relationship with AhG on disparity vector constraints. Such constraints are likely to be defined in the context of profile definitions. It was commented that these functionalities could co-exist.

A question was asked why the vps\_id is signaled in the SEI.

This feature was generally considered useful.

Recommendation: Revisit subject to availability solutions for aspects of tiles partitioning in the reference view, different CTU sizes across view, need for signaling vps\_id.

**Recommendations in the joint HLS BoG of JCT-3V and JCT-VC have been reviewed in this BoG and**

The following have been recommended for adoption into the combined high-level syntax design, to be included in working drafts for both SHVC and MV-HEVC.

* JCTVC-L0039/JCTVC-C0165: several RAP picture related aspects
  + On EL CRA pictures:
    - CRA NAL unit type can be used when nuh\_layer\_id is greater than 0.
    - Inter-layer prediction is allowed for CRA NAL units with nuh\_layer\_id greater than 0, while inter prediction is disallowed
    - CRA NAL units need not be aligned across layers. In other words, a CRA NAL unit type can be used for all VCL NAL units with a particular value of nuh\_layer\_id while another NAL unit type can be used for all VCL NAL units with another particular value of nuh\_layer\_id in the same access unit.
  + On IDR and BLA pictures:
    - IDR pictures may have nuh\_layer\_id greater than 0 and they may be inter-layer predicted while inter prediction is disallowed.
    - IDR pictures shall be present in an access unit either in no layers or in all layers, i.e. an IDR nal\_unit\_type indicates a complete IDR access unit where decoding of all layers can be started.
* JCT3V-C0081/JCT3V-C0084: POC for all HEVC layers in an access unit shall be the same
* JCT3V-C0085: Specific editorial improvement as part of same adoption reflected in an aspect of JCTVC-L0039
* JCTVC-L0263: Editorial bug fix to ensure that coded picture in a layer can only reference pictures in a lower layer
* JCTVC-L0200: Add a splitting\_flag to the VPS extension, which imposes a constraint that bit mapping of layer\_id is supported, but otherwise doesn’t change existing syntax and semantics
* JCTVC-L0180: Profile tier level signaling per operation point, and optionally referencing the profile and tier from an earlier operation point while sending level
* JCTVC-L0446: Layer dependency signaling using mask approach
* JCTVC-L0188: Several aspects relating to combination of SHVC and MV-HEVC
  + Activation process for picture and sequence parameter sets for individual layers
  + Non-reference pictures at the highest decoded temporal sub-layer are marked as “unused for reference” immediately after their decoding to enable reduction of the DPB usage.
  + Change MV-HEVC’s view dependency change SEI to generic layer dependency SEI message, and include in combined text

The following contributions are suggested to be revisited by the JCTVC track so that additional experts are available:

* JCTVC-L0226: Aspect relating to byte alignment at end of VPS or beginning of VPS extension
* JCTVC-L0262: Signaling of required DPB size in VPS
* JCTVC-L0188: BoG recommends that the track revisit for consideration in version 1 the aspect related to marking of non-reference pictures at highest temporal sub-layer

This BoG group reviewed the results and agreed with the suggested recommendations.