|  |  |
| --- | --- |
| **Joint Collaborative Team on Video Coding (JCT-VC)**  **of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11**  12th Meeting: Geneva, CH, 14–23 Jan. 2013 | Document: JCTVC-L0178 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **Legacy base layer codec support in SHVC** | | |
| *Status:* | Input Document to JCT-VC | | |
| *Purpose:* | Proposal | | |
| *Author(s) or Contact(s):* | Jill Boyce  Danny Hong  Wonkap Jang | Tel: Email: | +1 201 478 6145 {jill, danny, wonkap}@vidyo.com |
| *Source:* | Vidyo | | |

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

# Abstract

SVC uses a single-loop coder design, which precludes the use of a separate legacy codec for the base layer with the scalable codec. However, it is expected that the SHVC will require multiple-loop decoding. In theory, multi-loop decoding enables a separate legacy codec to be used for the base layer, with a scalable HEVC enhancement layer codec, for either an AVC base layer or an HEVC base layer.

A legacy base layer codec outputs a cropped output picture rather than the full size decoded picture, when cropping is used. So when a legacy base layer codec is used, the scalable HEVC enhancement layer codec will not have access to the coded samples outside the cropping window of the base layer resolution picture. It is proposed that in the SHVC test model the enhancement layer codec predict from the conformance cropping window output base layer picture rather than the decoded base layer picture. It is also proposed to apply padding to the cropped base layer picture when there is no available sample in the scaled reference layer corresponding to the enhancement layer CU sample, so that a corresponding reference layer sample can always be considered to be available.

# Problem description

No test model yet exists for the HEVC scalability extension, but the SMuC reference software is available, and a test model document should be created soon. We propose that the test model design and corresponding reference software allow a mode of operation where a legacy base layer codec is supported, for both encoding and decoding. This is of particular interest for an AVC base layer, but is also applicable to a black box HEVC base layer codec. At this stage of the SHVC development, it is uncertain whether motion vector and mode information from the base layer will be used in coding the scalable enhancement layer, and this may be a profiling issue. However, it is expected that the base layer decoded or output samples will be used when coding the enhancement layer. We propose using the cropping window output samples rather than the decoded samples, in order to support a legacy base layer codec.

When the base layer resolution is not a multiple of the smallest CU size, mismatch can occur between the decoding processes for a separate legacy base layer codec vs. an integrated base layer codec. For example, consider a 2x spatial scalability case with a base layer of resolution 480x270 and an enhancement layer of 960x540. There could be a 3rd spatial layer with resolution of 1920x1080, which is why these particular resolutions are of interest, but that will be neglected in this discussion, since the issue of concern arises between the base layer and first enhancement layer.

When coding a base layer with input resolution of 480x270, the coded resolution is 480x272, but a legacy decoder would crop the decoded picture and output a 480x270 cropped output picture. For an enhancement layer input/cropped output resolution of 960x540, the coded resolution is 960x544. When decoding the bottom row of CUs of the enhancement layer, for the lowest 4 sample rows of the enhancement layer would have corresponding scaled reference layer samples available in the decoded picture but not in the cropped output picture. In order to allow for a legacy base layer decoder, it is proposed to not perform inter-layer prediction using the bottom rows of the decoded base layer picture, but instead to apply padding to the cropped output picture, and perform inter-layer prediction with respect to the padded cropped output picture. The proposed padding is similar to that used for motion prediction with motion vectors that point outside the reference picture.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| (a) | (b) | (c) |

**Figure 1. (a) Decoded base layer picture, (b) Cropped output base layer picture, (c) Cropped and padded base layer picture used for scaled reference**

Figure 1 illustrates base layer pictures for the proposed process. Fig 1(a) shows the decoded base layer picture, which for this example has resolution of 480x272. Fig 1(b) shows the cropped output base layer picture, with a resolution of 480x270. The two bottom lines (red and yellow), have been removed. The bottom line of the cropped output picture is blue. Fig 1(c) shows the cropped and padded base layer which is upsampled and then used as the scaled reference for inter-layer prediction. It has a resolution of 480x272, with padding to fill the bottom two lines with the bottom line within the cropped picture, shown in blue.

When coding a CU in the enhancement layer, prediction is performed from the corresponding CU in the base layer, i.e. in IntraBL mode. Mismatch occurs between prediction using the decoded base layer picture illustrated in Figure 1(a) vs. using the cropped and padded base layer picture illustrated in Figure 1(c), for those samples in the cropped region of the base layer picture.

When performing upsampling, neighboring samples are used by the upsampling filter. So, anytime the base layer resolution uses output cropping, the upsampled base layer picture could have different values for samples near the edge of the cropped region, depending on whether the upsampling filter is applied to the decoded base layer picture or to the cropped and padded base layer picture.

The SMuC software uses samples in the cropping region during the upsampling operation when an HEVC base layer is used, e.g. it uses the coded reference picture. However, the AVC base layer support in the SMuC software has already been modified to use the proposed method, e.g. it applies upsampling to the cropped and padded base layer picture. This means that that HEVC and AVC base layers are currently treated differently in the SMuC. Most of the common test conditions currently in use for the SHVC TEs do not include resolutions that require output cropping. All picture dimensions of enhancement layers in the common test conditions are divisible by 8. However, for the 1920x1080 2x spatial scalability test, the 960x540 base layer is coded as 960x544 and uses output cropping. This is the only common conditions test case where the proposed method would have an impact.

Padding can be further used in SHVC in other cases where the scaled base layer and enhancement layer do not exactly correspond, which is allowed in SVC, for example for a Region-of-Interest in the base layer. The current SMuC does not provide support for base and enhancement layer resolutions that do not correspond exactly.

## SVC background

SVC does not require that the reference and enhancement layers precisely correspond, and scaling factors between the reference and enhancement layers are not explicitly transmitted. Instead, for each layer the coded resolution is signaled in the SPS, and the scaling factor between layers can be calculated from those values. In the enhancement layer, offsets are signaled in the SVC SPS extension and/or in the slice header extension to indicate the offset between the corner luma samples of a scaled reference layer picture used for inter-layer prediction and the corner luma samples of the enhancement layer picture: scaled\_ref\_layer\_left\_offset, scaled\_ref\_layer\_top\_offset, scaled\_ref\_layer\_right\_offset, scaled\_ref\_layer\_bottom\_offset. The values may be either positive or negative, indicating that the scaled reference layer may be larger or smaller than the enhancement layer. For some samples in the enhancement layer picture there may be no available corresponding sample in the scaled reference layer picture.

In SVC, when coding an individual MB, the corresponding reference layer sample locations are identified. When coding an MB in which any of the corresponding reference sample locations are unavailable, as determined using the InCropWindow( ) function, the MB cannot be coded using inter-layer prediction. In that case, syntax elements used for signaling inter-layer prediction modes are not present in the bitstream for the MB, such as base\_mode\_flag, residual\_prediction\_flag,motion\_prediction\_flag\_l0 and motion\_prediction\_flag\_\_l1. An excerpt from the SVC MB layer syntax is copied below, with the InCropWindow( ) function highlighted.

|  |  |  |
| --- | --- | --- |
| macroblock\_layer\_in\_scalable\_extension( ) { | C | Descriptor |
| if( InCropWindow( CurrMbAddr ) && adaptive\_base\_mode\_flag ) |  |  |
| **base\_mode\_flag** | 2 | u(1) | ae(v) |
| … |  |  |
| if( adaptive\_residual\_prediction\_flag && slice\_type != EI &&  ( base\_mode\_flag | |  ( MbPartPredMode( mb\_type, 0 ) != Intra\_16x16 &&  MbPartPredMode( mb\_type, 0 ) != Intra\_8x8 &&  MbPartPredMode( mb\_type, 0 ) != Intra\_4x4 &&  InCropWindow( CurrMbAddr ) ) ) ) |  |  |
| **residual\_prediction\_flag** | 2 | u(1) | ae(v) |
| … |  |  |

With this conditional syntax presence, the parsing of the scalable enhancement layer is dependent upon having determined the availability of corresponding reference layer samples for each sample in the MB. Coded MBs that contain mostly samples with corresponding reference layer sample locations available but some unavailable sample locations cannot benefit from inter-layer prediction, because it is not allowed.

# Proposal

## Motivation

SHVC should allow for a legacy base layer codec, and define the decoding operation with no mismatch between a legacy base layer codec and an integrated base layer codec. Also, SHVC should provide a similar degree of flexibility to SVC in regards to allowing the scaled reference layer size to not directly correspond to the enhancement layer, to support base layer ROI regions. Specific proposals to enable these features are described below.

## Discussion

In the SHVC design, inter-layer sample prediction is proposed to be performed using the output cropping window decoded reference layer picture, rather than the decoded reference layer picture itself.

The CU syntax in the current SMuC contains only a single inter-layer prediction mode syntax element, intra\_bl\_flag, and it is always present for enhancement layer CUs. The common test conditions do not include region-of-interest scalability, and always provide a direct correspondence between the scaled base layer and enhancement layer pictures’ samples, so the case where a corresponding scaled reference layer sample is not present is not exercised. However, when the test model is written, arbitrary resolutions for each layer should be handled.

Rather than following the SVC methodology and adopting a function similar to InCropWindow( ) and making the presence of intra\_bl\_flag (and any other future inter-layer syntax elements) in the CU syntax conditional on the availability of corresponding samples, padding should be applied to the scaled reference layer as necessary, so that a corresponding sample is always available. The proposed padding is very similar to that used for motion vectors that point outside of a reference picture.

Padding can also be applied when scaled reference layer offsets are used to indicate lack of direct correspondence and alignment between the scaled reference layer picture and the enhancement layer picture.

Figure 2 illustrates a reference layer picture with a cropping window parameters CT (top), CR (right), CB (bottom), and CL (left), and an enhancement layer picture with scaled reference layer offsets, OT, OR, OB, and OL. Padding is performed when coding enhancement layer CUs in the shaded region.

**Figure 2. Reference layer picture with cropping window and enhancement layer picture with scaled reference layer offsets**

Using padding has several possible advantages. First it simplifies the specification of the scalable enhancement layer syntax, semantics, and decoding process because a corresponding sample is always available, eliminating the need to describe the special case when the corresponding sample is not available. Also, it simplifies parsing complexity in the decoder, by reducing the need to check for the special case.

Additionally, there are potentially coding efficiency benefits by allowing inter-layer prediction for CUs that are partially inside and partially outside of the corresponding scaled reference layer region. However, these gains may be offset by the additional cost of sending the inter-layer prediction mode syntax elements for those regions with at least some samples outside the corresponding scaled reference layer region.

## Test model support

The syntax in the test model should include signaling for all enhancement layer CUs for inter-layer prediction, e.g. intra\_bl\_flag, without adding a function similar to SVC’s InCropWindow( ) to make those syntax elements conditional on the presence of a corresponding CU in the reference layer. Syntax should be added in the SPS extension, and possibly the slice segment header extension, similar to SVC, to indicate scaled ref layer offsets.

The decoding process to be described in the test model, should include the following steps when describing the decoding of a current layer:

1. Decode the reference layer picture
2. For each CU (or PU) in the current layer picture, when inter-layer prediction is indicated
   1. Find the corresponding CU in the cropped and padded scaled reference layer picture
   2. Use that corresponding CU for inter-layer prediction

Test model text cannot be proposed at this stage because the SHVC test model document is not yet available. However, it is anticipated that the proposed method will be easier to specify than would a method similar to SVC’s InCropWindow( ).

# Patent rights declaration(s)

**Vidyo may have IPR 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).**