|  |  |
| --- | --- |
| **Joint Collaborative Team on Video Coding (JCT-VC)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  13th Meeting: Incheon, KR, 18–26 Apr. 2013 | Document: JCTVC-M0077\_r1 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **Cross-check on SCE3.5: Simplification of Generalized Residual Inter-Layer Prediction for spatial scalability** | | |
| *Status:* | Input Document to JCT-VC | | |
| *Purpose:* | Information | | |
| *Author(s) or Contact(s):* | Takeshi Tsukuba  1-9-2 Nakase, Mihama-ku,  Chiba-shi, Chiba 261-8520 JAPAN | Tel: Email: | +81-43-299-8526 [tsukuba.takeshi@sharp.co.jp](mailto:tsukuba.takeshi@sharp.co.jp) |
| *Source:* | SHARP Corporation | | |

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

# Abstract

This contribution reports cross-check results of SCE3.5 (JCTVC-M0109) on simplification of generalized residual inter-layer prediction for spatial scalability. The results of coding performance and average complexity exactly match those provided by the proponents. It is also confirmed that the worst case of complexity provided by the proponents is appropriate. Note that the results of worst case complexity were revised and also confirmed since the template was updated.

# Verification

In SCE3.5 (JCTVC-M0109) [1], the proponent proposed a simplification method to reduce memory access of generalized residual inter-layer prediction (GRP) for spatial scalability.

Basically, in GRP for uni-prediction case, the prediction signal PREDEL for the enhancement layer is formulated as follows:

PREDEL = **MC1**[ REFEL, MVEL] + **W**\*{**UP1**{RECBL} – **MC2**[ **UP2**{REFBL} , MVEL] } -(eq.1)

where, the above terms are defined as follows:

* PREDEL is the prediction signal of the enhancement layer
* REFEL and REFBL are the temporal reference signals in the enhancement and base layers
* RECBL is the base layer reconstructed signal corresponding to the current enhancement layer block
* MVEL is the enhancement layer motion vector
* UPx{.} is the up-sampling operator x
* MCx[I,MV] is the motion compensation operator x of the current block using I as reference picture and MV for the motion vector.
* W is the weighting factor for the second order residual

## Spatial Scalability case

For spatial scalability, the proposed method in [1] is summarized as follows:

* Combining **MC2** and **UP2** in eq.1 into a single step, namely as **MCoUP**, as shown in eq. 2.  
  PREDEL = **MC1**[ REFEL, MVEL] + **W**\*{**UP1**{RECBL} – **MCoUP**[REFBL, MVEL] } -(eq.2)
* Employing 8-phase-8-tap for luma and 16-phase-4-tap for chroma as **MCoUP** in Spatial 2x case and 6-phase-8 tap for luma and 12-phase-4-tap for chroma in Spatial 1.5x case
* Using bilinear interpolation filter for **MC1** in eq. 1.

## SNR Scalability case

For SNR scalability, the proposed method in [1] is summarized as follows:

* Using bilinear interpolation filter for **MC1** and **MC2** in eq. 1.

## Restriction of GRP mode

In [1], the GRP mode is disabled on PUs highlighted in yellow in Table 1 and 2 for uni/bi-prediction respectively at encoder.

Table 1: Restriction of GRP mode in uni-prediction case



Table 2: Restriction of GRP mode in bi-prediction case



# Experimental results

We inspected the source code provided by the proponents, implemented on SHM-1.0, to verify that the proposed method was implemented as described in [1]. Two evaluations for SCE3.5 are done based on the test condition described in SCE3 document [2].

## Test1: SCE3.5 (one weight case: W=1.0)

In Test1, the GRP mode which uses one weight (W=1.0) is tested.

Table 3 and 4 shows the coding performance and complexity information of the Test1 respectively. The results of BD-rate and actual complexity assessment of Test1 match those provided by the proponents. Please note that decoding times of LB case are not accurate due to our un-uniform clusters.

Table : Coding performance of Test 1 (ref. SHM1.0)



Table : Complexity information of Test 1 (ref. SHM1.0)



## Test2:SCE3.5 (two weight case: W=0.5, and 1.0)

In Test2, the GRP mode which uses two weights (W=0.5, and 1.0) is tested.

Table 5 and 6 shows the coding performance and complexity information of the Test1 respectively. The results of BD-rate and complexity assessment of Test1 match those provided by the proponents. Please note that decoding times of LB case are not accurate due to our un-uniform clusters.

Table 5: Coding performance of Test 2 (ref. SHM1.0)



Table 6: Complexity information of Test 2 (ref. SHM1.0)



## Worst case complexity assessment

The following table summarizes the worst case complexity for the two tests (Test1 and Test2). The worst case complexity results match those provided by the proponents.

Table 7: Worst case complexity for SCE3.5

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Test1  ratio 2x | Test1  ratio 1.5x | Test1  all ratios | Test2  all ratios |
| **Mult** | 115% | 115% | 115% | 115% |
| **Adds** | 113% | 113% | 113% | 113% |
| **MemBand(2D:4x2)** | 83% | 99% | 99% | 99% |
| **MemBand(2D:8x2)** | 81% | 100% | 100% | 100% |
| **Number of Ref Frames** | 120% | 136% | 180% | 180% |
| **Tables Size : max case** | 229% | 229% | 229% | 229% |
| **min case** | 200% | 200% | 200% | 200% |

Regarding the number of Ref Frames, it is estimated depending on the ratio between BL and EL resolution as follows;

In the case ratio 2x: nb of ref frames = 5 + 4 / (2\*2) = 6 ( 120% of the anchor )

In the case ratio 1.5x: nb of ref frames = 5 + 4 / (1.5\*1.5) = 7 ( 136% of the anchor )

In the case ratio 1x(SNR): nb of ref frames = 5 + 4 = 9 ( 180% of the anchor)

Regarding the table size, two cases (max and min case) are estimated respectively.

In the max case, the table size is estimated as follows;

Tables size = anchor + MCoUp in 2x + MCoUp in 1.5x + MC( bilinear )   
= 8-tap \* 16 phases for luma + 4-tap \*16phases for chroma // anchor (up-sampling)  
 + 8-tap \* 8 phases for luma + 4-tap\*16 phases for chroma // MCoUP in 2x  
 + 8-tap \* 6 phases for luma + 4-tap\*12 phases for chroma // MCoUP in 1.5x  
 +2-tap \* 4 phases for luma + 2-tap \* 8 phases for chroma // MC (bilinear )  
= 440 ( 229% of the anchor )

In the min case, since some filters (0-phase, 0.5-phase) used for MCoUP are the same as regular MC and luma filters of bilinear are the same as chroma’s, the table size is estimated as follows;

Tables size = anchor + MCoUp in 2x + MCoUp in 1.5x + MC( bilinear )   
= 8-tap \* 16 phases for luma + 4-tap \*16phases for chroma // anchor (up-sampling)  
+ 8-tap \* (8-2) phases for luma + 4-tap\*(16-2) phases for chroma // MCoUP in 2x  
+ 8-tap \* (6-2) phases for luma + 4-tap\*(12-2) phases for chroma // MCoUP in 1.5x  
+ 2-tap \* 8 phases // MC (bilinear )  
= 384 ( 200% of the anchor )

# Conclusion

This contribution reports cross-check results of SCE3.5 (JCTVC-M0109) on simplification of generalized residual inter-layer prediction for spatial scalability. The results of coding performance and average complexity exactly match those provided by the proponents. It is also confirmed that the worst case of complexity assessment provided by the proponents is appropriate.

# Reference

1. [E. François](mailto:edouard.francois@crf.canon.fr), et.al, “SCE3.5: Simplification of Generalized Residual Inter-Layer Prediction for spatial scalability”, JCTVC-M0109, Incheon, KR, 18–26 Apr. 2013.
2. X. Li, et.al, “Description of Tool Experiment SCE3: Combined Inter and Inter-Layer Prediction in SHVC,” JCTVC-L1103, Geneva, CH, Jan. 2013.