|  |  |
| --- | --- |
| **Joint Collaborative Team on Video Coding (JCT-VC)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  14th Meeting: Vienna, AT, 25 July – 2 Aug. 2013 | Document: JCTVC-N0204 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **ILR enhancement with differential coding for SHVC reference index framework** | | |
| *Status:* | Input Document to JCT-VC | | |
| *Purpose:* | Proposal | | |
| *Author(s) or Contact(s):* | Yuwen He, Yan Ye, Xiaoyu Xiu  9710 Scranton R-D, #250 San Diego, CA 92121 USA | Tel: Email: | +1-858-210-4819 [yuwen.he@interdigital.com](mailto:yuwen.he@interdigital.com) [yan.ye@interdigital.com](mailto:yan.ye@interdigital.com) [xiaoyu.xiu@interdigital.com](mailto:xiaoyu.xiu@interdigital.com) |
| *Source:* | InterDigital Communications, Inc. | | |

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

# Abstract

This proposal describes inter-layer reference (ILR) enhancement with differential coding for SHVC reference index (RefIdx) framework. In SHVC reference index framework, the base layer reconstructed picture (after upsampling if needed) is used as an additional reference for enhancement layer coding. In this contribution, the ILR is further enhanced by adding weighted differential signal from the temporal domain to restore high frequency information. The differential signal is generated by motion compensation in the temporal domain with the motion field from the base layer picture. Compared to the SHM-2.0 RefIdx anchor, the proposed scheme reportedly achieves average {Y, U, V} BD rate gain of {-2.3%, -6.6%, -7.4%}, {-2.9%, -7.0%, -7.6%} and {-3.6%, -6.9%, -7.3%} for RA, LD-B, and LD-P if uncompressed motion field from base layer picture is used, respectively. The results with 8x8 and 16x16 sized compressed motion field are also reported.

# Introduction

In the RefIdx framework [1], the ILR is inserted (after upsampling if needed) into the reference picture list(s) in enhancement layer coding. The ILR picture usually lacks sufficient high frequency information, because 1) the base layer picture is quantized during base layer encoding, and 2) the base layer picture has smaller spatial resolution in the case of spatial scalability. The previous proposals [3][4][5] show that the ILR picture quality can be enhanced effectively with differential coding method. This contribution updates the results of ILR enhancement with weighted differential signal based on SHM-2.0 [1][2]. The results of four test cases for enhanced ILR (EILR) are reported: (1) ILR is enhanced with the uncompressed motion field from base layer, and it is used as one additional inter-layer reference picture. (2) ILR is enhanced with the 8x8 sized compressed motion field from base layer, and it is used as one additional inter-layer reference picture. (3) ILR is enhanced with the normal 16x16 sized compressed motion field from base layer, and it is used as one additional inter-layer reference picture. (4) ILR is enhanced with the normal 16x16 sized compressed motion field from base layer, and there is only one inter-layer reference picture: ILR, EILR, or the block-based combination of ILR and EILR. It can save the memory usage for decoder with picture level ILR implementation.

# Algorithm description

## ILR enhancement with weighted differential signal

The symbols used in this proposal are defined in Table 1. shows the generation method of EILR. EILR is generated by enhancing ILR block with motion aligned differential block.

Table . Definition of symbols

|  |  |
| --- | --- |
| **Symbol** | **Definition** |
| B(P) | block of picture P |
| MC (P, MV) | motion compensation with picture P and motion vector MV; bilinear interpolation used in this proposal |
| BL’T | base layer reconstructed picture at time T |
| ILRT | inter-layer reference picture at time T |
| EILRT | Enhanced inter-layer reference picture at time T |
| ELT | enhancement layer original picture at time T |
| EL’T | enhancement layer reconstructed picture at time T |
| SMV | Scaled motion vector for inter layer reference |
| Wuni(R) | weight for the blocks with uni-prediction and the reference picture is R, (0≤ Wuni ≤1) |
| Wbi(R0, R1) | weight for the blocks with bi-prediction and the reference pictures are R0 and R1 (0≤ Wbi ≤1) |

B(BL’T) is the base layer collocated block. SMV is derived by scaling base layer motion according to spatial scalability ratio. The enhanced ILR block B(EILRT) is derived with Eq.(2) for uni-prediction mode, and Eq.(3) for bi-prediction mode.

B(EILRT) = Clip( B(ILRT) + (1-Wuni(ILRT-n)) \* (MC(EL’T-n, SMV) – MC(BL’T-n, SMV))) (2)

B(EILRT)=Clip( B(ILRT)+(1-Wbi(ILRT-n0, ILRT-n1))\*( (MC(EL’T-n0, SMV0)+MC(EL’T-n1, SMV1))/2 –

( (MC(BL’T-n0, SMV0)+MC(BL’T-n1, SMV1))/2)) (3)

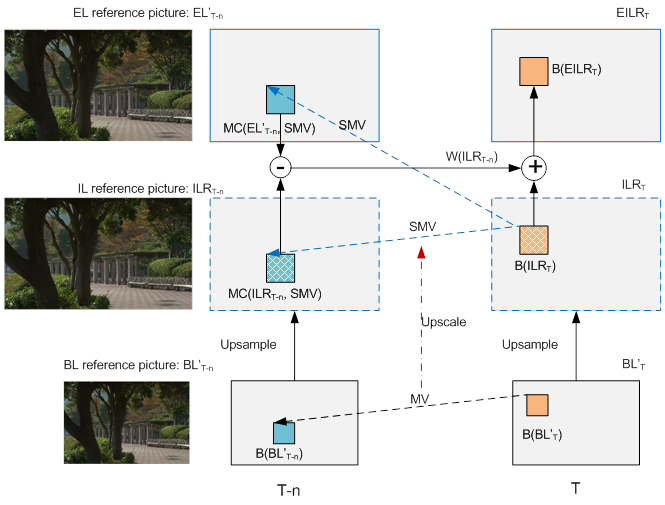


Figure . ILR enhancement with motion compensated signal

## Weight estimation

To generate the enhanced ILR picture, the encoder estimates weights that will be applied to differential signal to get good high frequency restoration for compression. For uni-prediction mode, each unique reference picture identified by POC in both list0 and list1 has a different weight. For bi-prediction mode, each unique reference picture pair identified by two POCs from list0 and list1 has one weight. These two weight sets, one for uni-prediction and the other for bi-prediction, are signaled in the slice header of enhancement layer. Figure 2 shows an example of weight list signaling, where POC of the current picture is 2, and the coding configuration is random access. Figure 2 (a) shows two reference lists. P4 appears in both lists. Figure 2 (b) is the weight lists for uni- and for bi-prediction. The weights are estimated to minimize the distortion by solving Eq(5) and Eq(6) for uni- and bi-prediction. We use the Least Square method and SSE for distortion evaluation. The estimated weights are quantized to fixed point precision and fixed length encoded. In our implementation, we used 2-bit precision for the weights.

W(R) =  (5)

where the collocated block at base layer of block Bi(ELT) is uni-prediction coded, and the reference picture is R.

W(R0, R1) =  (6)

where the collocated block at base layer of block Bi(ELT) is bi-prediction coded, and the references pictures from two lists are R0 and R1.

|  |  |  |
| --- | --- | --- |
| **List** | **Index** | |
| 0 | 1 |
| **L0** | P0 | P4 |
| **L1** | P4 | P8 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Weight set** | **Index** | | | |
| 0 | 1 | 2 | 3 |
| **weight\_list\_uni** | P0 | P4 | P8 |  |
| **weight\_list\_bi** | (P0, P4) | (P0, P8) | (P4, P4) | (P4, P8) |

(a) (b)

Figure . weight list of picture P2 for RA configuration

## EILR placement

There are two settings for EILR placement: setting one is for the case that EILR used as additional inter-layer reference picture, setting two is for the case ILR is replaced by EILR. Figure 3 shows the EILR insertion in B and P slice coding in setting one. For enhancement layer B slice coding, EILR is put at the end of list1. For P slice coding, it is put after temporal reference pictures. Figure 4 shows the EILR insertion in B and P slice coding in setting two. For setting two, there is an inter-layer reference picture type signaling to indicate inter-layer reference picture is normal ILR without enhancement, or EILR, or the block-based EILR. If it is block-based EILR, then there is a block based flag map encoded at the slice header of enhancement layer. The flag is to indicate if enhancement is applied for the luma component of the block. The flag map is coded by run-length coding with Exp-Golomb codes. The block size is 64x64 in simulation.



Figure . Two inter-layer reference pictures: EILR placement in reference lists



Figure . One inter-layer reference picture: ILR’ can be ILR, EILR, or block-based EILR

EILR will not be available at enhancement layer when the base layer picture is intra slice coding because there is no motion information at base layer. Both settings degrade to the coding with normal ILR.

# Simulation results

The compression performance is measured using BD rate compared with SHM-2.0 RefIdx anchor. There are three coding configurations: random access, low-delay B and low-delay P. Table 2 gives a summary of four test cases under the CTC [2]. Table 3 gives the detailed average BD rate reduction for test case 1. In test case 1, the motion applied in ILR enhancement is derived from BL uncompressed motion. In test case 2, the motion applied in EILR generation is derived from BL 8x8 compressed motion. In test case 3 and 4, the motion applied in EILR generation is derived from BL 16x16 compressed motion. Both EILR and ILR are used as inter-layer reference pictures for test case 1, 2, 3. In test case 4, only one inter-layer reference picture is used and it can be ILR, EILR or block-based EILR signaled at slice header. The interpolation filter applied to generate differential signal for all test cases is bilinear. The scaling factor used for fixed point representation of the weights is 4, and the weights are coded with variable length coding. The weights are only applied to luma component; for chroma components, the differential signal is added directly without weights.

As shown in Table 2, the proposed scheme reportedly achieves average {Y, U, V} BD rate gain of {-2.3%, -6.6%, -7.4%}, {-2.9%, -7.0%, -7.6%} and {-3.6%, -6.9%, -7.3%} for RA, LD-B, and LD-P, respectively for test case 1. Readers are referred to the accompanying spreadsheets for further details.

“JCTVC-M1009-SHVC-template\_HEVCBase\_r1\_idcc\_E2UCMV\_FR\_Testcase1.xlsm” is for the results of test case 1.

“JCTVC-M1009-SHVC-template\_HEVCBase\_r1\_idcc\_E2CMV8x8\_FR\_Testcase2.xlsm” is for the results of test case 2.

“JCTVC-M1009-SHVC-template\_HEVCBase\_r1\_idcc\_E2CMV\_FR\_Testcase3.xlsm” is for the results of test case 3.

“JCTVC-M1009-SHVC-template\_HEVCBase\_r1\_idcc\_E3CMV\_FR\_Testcase4.xlsm” is for the results of test case 4.

Table . Summary of simulation results

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Index** | **Test cases** | **Coding configurations** | **Aver Results** | | | | |
| Y | U | V | EncT | DecT |
| 1 | Uncompressed motion | RA | -2.3% | -6.6% | -7.4% | 109% | 107% |
| LD-B | -2.9% | -7.0% | -7.6% | 101% | 103% |
| LD-P | -3.6% | -6.9% | -7.3% | 102% | 102% |
| Average | -2.9% | -6.8% | -7.4% | 104% | 104% |
| 2 | 8x8 Compressed motion | RA | -2.1% | -6.0% | -6.9% | 123% | 123% |
| LD-B | -2.6% | -6.4% | -7.0% | 102% | 107% |
| LD-P | -3.2% | -6.2% | -6.5% | 100% | 101% |
| Average | -2.6% | -6.2% | -6.8% | 108% | 110% |
| 3 | 16x16 Compressed motion | RA | -1.5% | -4.5% | -5.4% | 126% | 121% |
| LD-B | -2.0% | -5.1% | -5.6% | 102% | 101% |
| LD-P | -2.5% | -4.9% | -5.1% | 101% | 92% |
| Average | -2.0% | -4.8% | -5.4% | 110% | 105% |
| 4 | 16x16 Compressed motion + one ILR | RA | -1.2% | -4.7% | -5.9% | 107% | 117% |
| LD-B | -1.6% | -4.4% | -5.1% | 90% | 94% |
| LD-P | -1.8% | -3.2% | -3.7% | 91% | 92% |
| Average | -1.5% | -4.1% | -4.9% | 96% | 101% |

Table . Average BD rate reduction for ILR enhancement with uncompressed motion from base layer (test case 1)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **RA HEVC 2x** | | | **RA HEVC 1.5x** | | | **RA HEVC SNR** | | |
|  | Y | U | V | Y | U | V | Y | U | V |
| Class A | -2.3% | -7.6% | -7.2% |  |  |  | -2.6% | -9.0% | -9.5% |
| Class B | -1.8% | -4.5% | -5.1% | -2.4% | -6.0% | -7.0% | -2.6% | -8.0% | -9.7% |
| **Overall (Test vs Ref)** | -2.0% | -5.4% | -5.7% | -2.4% | -6.0% | -7.0% | -2.6% | -8.3% | -9.6% |
| **Overall (Test vs single layer)** | 16.9% | 26.1% | 24.4% | 13.4% | 21.2% | 20.3% | 11.5% | 20.8% | 20.8% |
| **Overall (Ref vs single layer)** | 19.2% | 33.3% | 32.0% | 16.2% | 28.8% | 29.1% | 14.4% | 32.1% | 34.1% |
| **EL only (Test vs Ref)** | -4.0% | -7.4% | -7.7% | -7.0% | -10.6% | -11.5% | -4.9% | -10.9% | -12.4% |
| Enc Time[%] | 110.4% | | | 104.5% | | | 111.9% | | |
| Dec Time[%] | 103.3% | | | 107.6% | | | 110.9% | | |
| Enc Mem[%] | #VALUE! | | | #VALUE! | | | #VALUE! | | |
| BL Match | Matched | | | Matched | | | Matched | | |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **LD-B HEVC 2x** | | | **LD-B HEVC 1.5x** | | | **LD-B HEVC SNR** | | |
|  | Y | U | V | Y | U | V | Y | U | V |
| Class A | -2.8% | -8.4% | -7.8% |  |  |  | -3.0% | -9.1% | -9.6% |
| Class B | -2.4% | -4.7% | -4.9% | -3.3% | -7.1% | -7.7% | -2.9% | -7.8% | -9.4% |
| **Overall (Test vs Ref)** | -2.5% | -5.8% | -5.8% | -3.3% | -7.1% | -7.7% | -3.0% | -8.2% | -9.5% |
| **Overall (Test vs single layer)** | 25.3% | 30.9% | 31.6% | 20.6% | 23.8% | 25.6% | 20.7% | 23.4% | 25.8% |
| **Overall (Ref vs single layer)** | 28.5% | 39.0% | 39.7% | 24.8% | 33.0% | 35.9% | 24.3% | 34.7% | 39.5% |
| **EL only (Test vs Ref)** | -4.7% | -7.9% | -7.9% | -8.8% | -12.2% | -12.8% | -5.3% | -10.7% | -12.0% |
| Enc Time[%] | 103.8% | | | 97.4% | | | 102.3% | | |
| Dec Time[%] | 94.9% | | | 102.8% | | | 110.4% | | |
| Enc Mem[%] | #VALUE! | | | #VALUE! | | | #VALUE! | | |
| BL Match | Matched | | | Matched | | | Matched | | |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **LD-P HEVC 2x** | | | **LD-P HEVC 1.5x** | | | **LD-P HEVC SNR** | | |
|  | Y | U | V | Y | U | V | Y | U | V |
| Class A | -3.3% | -8.2% | -7.4% |  |  |  | -3.6% | -9.2% | -9.4% |
| Class B | -3.0% | -4.2% | -4.2% | -3.8% | -6.7% | -6.7% | -4.0% | -8.7% | -10.1% |
| **Overall (Test vs Ref)** | -3.1% | -5.3% | -5.1% | -3.8% | -6.7% | -6.7% | -3.9% | -8.8% | -9.9% |
| **Overall (Test vs single layer)** | 22.7% | 30.6% | 32.0% | 18.2% | 24.1% | 26.7% | 18.7% | 22.1% | 24.7% |
| **Overall (Ref vs single layer)** | 26.6% | 38.0% | 39.1% | 22.8% | 32.8% | 35.6% | 23.4% | 34.6% | 39.4% |
| **EL only (Test vs Ref)** | -5.5% | -7.6% | -7.4% | -9.6% | -12.2% | -12.2% | -6.5% | -11.6% | -12.7% |
| Enc Time[%] | 101.4% | | | 103.2% | | | 101.4% | | |
| Dec Time[%] | 95.1% | | | 104.1% | | | 105.3% | | |
| Enc Mem[%] | #VALUE! | | | #VALUE! | | | #VALUE! | | |
| BL Match | Matched | | | Matched | | | Matched | | |

# Conclusions

In this proposal, an ILR enhancement technology is proposed for the RefIdx framework of SHVC. The proposed method uses weighted differential signal to enhance the high frequency information in the inter-layer reference (ILR) picture. Four test cases with different complexity settings are evaluated. Simulation results show substantial coding performance improvements with reasonable complexity increase.

# Patent rights declaration(s)

**InterDigital Communications, Inc. 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).**

# References

1. J. Chen, J. Boyce, Y. Ye, M. M. Hannuksela, “SHVC Test Model 2”, JCTVC-M1007, Apr. 2013.
2. X. Li, J. Boyce, P. Onno, Y. Ye, “Common SHM test conditions and software reference configurations”, JCTVC-M1009, Apr. 2013.
3. Y. He, Y. Ye, “Non-SCE3: ILR enhancement with differential coding for RefIdx framework”, JCTVC-M0189, Apr. 2013.
4. A. Aminlou, J. Lainema, K. Ugur, M. Hannuksela, “Non-CE3: Enhanced inter layer reference picture for RefIdx based scalability”, JCTVC-M0155, Apr. 2013.
5. X. Li, E. Francois, P. Lai, D.-K. Kwon, A. Saxena, “SCE3: Summary Report of SHVC Core Experiment on Combined Inter- and Interlayer Prediction”, JCTVC-M0023, Apr. 2013.