|  |  |
| --- | --- |
| **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-M0189 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **Non-SCE3: ILR enhancement with differential coding for RefIdx framework** | | |
| *Status:* | Input Document to JCT-VC | | |
| *Purpose:* | Proposal | | |
| *Author(s) or Contact(s):* | Yuwen He, Yan Ye  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) |
| *Source:* | InterDigital Communications, LLC | | |

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

# Abstract

This proposal describes inter-layer reference (ILR) enhancement with differential coding for the RefIdx framework. In the RefIdx 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 compressed motion field from the base layer picture. Compared to the SHM1.0 RefIdx anchor, the proposed scheme reportedly achieves average {Y, U, V} BD rate gain of {-1.6%, -4.6%, -5.2%}, {-2.6%, -5.0%, -5.1%} and {-2.0%, -5.0%, -5.4%} for RA, LD-P, and LD-B, respectively. It is also reported that higher {Y, U, V} BD rate gain of {-2.5%, -6.7%, -7.3%}, {-3.7%, -7.1%, -7.2%} and {-3.0%, -6.9%, -7.4%} for RA, LD-P, and LD-B, respectively, can be achieved, if uncompressed motion field from the base layer picture is used.

# 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 differential signal used in differential coding [3][4] provides some high frequency information. In this contribution, an ILR enhancement method is proposed to restore the high frequency information in the ILR using weighted differential signal. The weighted differential signal is generated by performing motion compensation using motion information derived from the base layer motion field [2]. The enhanced ILR picture is then used as additional reference picture for enhancement layer coding.

# Algorithm description

## Differential signal generation

The symbols used in this proposal are defined in Table 1. shows the ILR block is enhanced by motion aligned differential block.

Table . Definition of symbols

|  |  |
| --- | --- |
| **Symbol** | **Definition** |
| B(P) | block of picture P |
| MCDiff(PDiff, MV) | motion compensation with differential picture PDiff 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 |
| ILDiffT | Inter-layer differential 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) |

ILDiffT is generated using Eq.(1), where Ioffset is equal to 2bitdepth-1 .

ILDiffT = Clip (EL’T – ILRT + Ioffset) (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)) \* (MCDIFF(ILDiffT-n, SMV) – Ioffset)) (2)

B(EILRT)=Clip( B(ILRT)+(1-Wbi(ILRT-n0, ILRT-n1))\*( (MCDIFF(ILDiffT-n0, SMV0)+MCDIFF(ILDiffT-n1, SMV1))/2 – Ioffset)) (3)

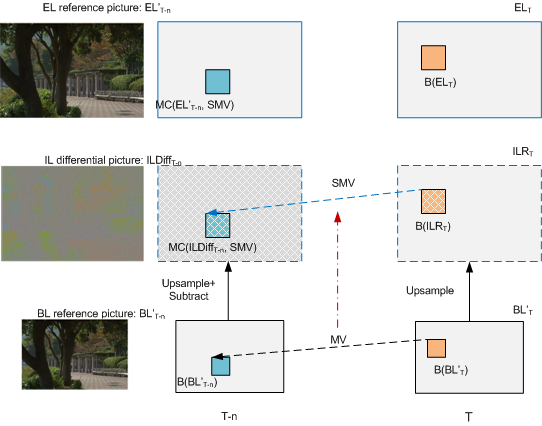


Figure . ILR enhancement with motion compensated differential signal

## 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 a weight. These two weight sets, one for uni-prediction and the other for bi-prediction, are signaled in the slice header of enhancement layer. 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) = , *Bi(BL’T)* is uni-prediction and reference is R; (5)

W(R0, R1) = , *Bi(BL’T)* is bi-prediction and references are R0 and R1 (6)

|  |  |  |
| --- | --- | --- |
| **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

shows the enhanced ILR (EILR) insertion in B and P slice coding in the simulation. For enhacement layer B slice coding, EILR is put at the end of list1. For P slice coding, it is put after temporal reference pictures.



Figure . EILR placement in the reference list

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.

# Simulation results

The compression performance is measured using BD rate compared with SHM-1.0 RefIdx anchor. There are three coding configurations: random access, low-delay P and low-delay B. Table 2 and Table 3 give the average BD rate reduction under the CTC; in Table 2, the motion applied in ILR enhancement is derived from BL compressed motion, and in Table 3, the motion applied in ILR enhancement is derived from BL uncompressed motion. The interpolation filter applied to differential signal (MCDiff) is bilinear. The number of bits used for the weights is 2. 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 {-1.6%, -4.6%, -5.2%}, {-2.6%, -5.0%, -5.1%} and {-2.0%, -5.0%, -5.4%} for RA, LD-P, and LD-B, respectively. If uncompressed motion field from base layer is used. higher {Y, U, V} BD rate gain of {-2.5%, -6.7%, -7.3%}, {-3.7%, -7.1%, -7.2%} and {-3.0%, -6.9%, -7.4%} for RA, LD-P, and LD-B, respectively, can be achieved. Readers are referred to the accompanying spreadsheet for further details.

Table . Average BD rate reduction for ILR enhancement with compressed motion from base layer

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **RA HEVC 2x** | | | **RA HEVC 1.5x** | | | **RA HEVC SNR** | | |
|  | Y | U | V | Y | U | V | Y | U | V |
| Class A | -0.8% | -3.1% | -3.0% |  |  |  | -1.7% | -6.1% | -6.3% |
| Class B | -1.1% | -3.0% | -3.4% | -1.7% | -4.5% | -5.2% | -2.1% | -6.3% | -7.6% |
| **Overall (Test vs Ref)** | -1.0% | -3.0% | -3.3% | -1.7% | -4.5% | -5.2% | -2.0% | -6.3% | -7.3% |
| **Overall (Test vs single layer)** | 18.9% | 29.2% | 29.9% | 15.4% | 22.4% | 24.1% | 13.6% | 21.5% | 24.4% |
| **EL only (Test vs Ref)** | -2.2% | -4.1% | -4.4% | -5.4% | -8.1% | -8.8% | -4.1% | -8.6% | -9.6% |
| Enc Time[%] | 118.6% | | | 114.1% | | | 110.1% | | |
| Dec Time[%] | 138.4% | | | 132.7% | | | 131.0% | | |
| Enc Mem[%] | #NUM! | | | #NUM! | | | #NUM! | | |
| 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 | -1.4% | -3.6% | -3.4% |  |  |  | -2.1% | -5.6% | -5.7% |
| Class B | -2.2% | -2.8% | -2.6% | -2.9% | -5.0% | -5.0% | -3.4% | -7.3% | -8.1% |
| **Overall (Test vs Ref)** | -2.0% | -3.1% | -2.8% | -2.9% | -5.0% | -5.0% | -3.0% | -6.8% | -7.4% |
| **Overall (Test vs single layer)** | 24.1% | 33.5% | 35.2% | 19.4% | 26.3% | 29.1% | 19.8% | 24.9% | 28.1% |
| **EL only (Test vs Ref)** | -3.3% | -4.3% | -4.1% | -7.2% | -9.1% | -9.1% | -5.0% | -8.9% | -9.6% |
| Enc Time[%] | 110.1% | | | 105.8% | | | 104.9% | | |
| Dec Time[%] | 128.6% | | | 124.9% | | | 128.8% | | |
| Enc Mem[%] | #NUM! | | | #NUM! | | | #NUM! | | |
| 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 | -1.0% | -3.6% | -3.4% |  |  |  | -1.8% | -5.8% | -6.1% |
| Class B | -1.6% | -3.1% | -3.3% | -2.4% | -5.6% | -5.8% | -2.4% | -6.4% | -7.5% |
| **Overall (Test vs Ref)** | -1.4% | -3.2% | -3.4% | -2.4% | -5.6% | -5.8% | -2.3% | -6.2% | -7.1% |
| **Overall (Test vs single layer)** | 26.7% | 34.2% | 35.5% | 21.8% | 25.9% | 28.0% | 21.5% | 26.0% | 29.2% |
| **EL only (Test vs Ref)** | -2.6% | -4.4% | -4.5% | -6.6% | -9.5% | -9.7% | -4.0% | -8.1% | -9.0% |
| Enc Time[%] | 128.2% | | | 117.6% | | | 116.9% | | |
| Dec Time[%] | 157.9% | | | 154.6% | | | 159.1% | | |
| Enc Mem[%] | #NUM! | | | #NUM! | | | #NUM! | | |
| BL Match | Matched | | | Matched | | | Matched | | |

Table . Average BD rate reduction for ILR enhancement with uncompressed motion from base layer

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **RA HEVC 2x** | | | **RA HEVC 1.5x** | | | **RA HEVC SNR** | | |
|  | Y | U | V | Y | U | V | Y | U | V |
| Class A | -2.5% | -8.0% | -7.3% |  |  |  | -2.8% | -9.7% | -9.8% |
| Class B | -1.8% | -4.7% | -5.1% | -2.6% | -6.3% | -7.0% | -2.7% | -7.7% | -9.0% |
| **Overall (Test vs Ref)** | -2.0% | -5.6% | -5.7% | -2.6% | -6.3% | -7.0% | -2.7% | -8.2% | -9.2% |
| **Overall (Test vs single layer)** | 17.7% | 25.6% | 26.5% | 14.3% | 20.2% | 21.7% | 12.7% | 18.8% | 21.7% |
| **EL only (Test vs Ref)** | -4.3% | -7.8% | -7.9% | -7.8% | -11.4% | -12.1% | -5.5% | -11.2% | -12.3% |
| Enc Time[%] | 113.3% | | | 108.1% | | | 102.3% | | |
| Dec Time[%] | 135.8% | | | 130.9% | | | 131.1% | | |
| Enc Mem[%] | #NUM! | | | #NUM! | | | #NUM! | | |
| 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 | -3.4% | -8.3% | -7.6% |  |  |  | -3.8% | -9.4% | -9.5% |
| Class B | -3.1% | -4.3% | -3.9% | -3.9% | -6.8% | -7.0% | -4.2% | -8.7% | -9.7% |
| **Overall (Test vs Ref)** | -3.2% | -5.5% | -5.0% | -3.9% | -6.8% | -7.0% | -4.1% | -8.9% | -9.7% |
| **Overall (Test vs single layer)** | 22.6% | 30.1% | 32.1% | 18.1% | 23.9% | 26.5% | 18.5% | 22.0% | 24.9% |
| **EL only (Test vs Ref)** | -5.6% | -7.7% | -7.2% | -9.7% | -12.3% | -12.4% | -6.7% | -11.7% | -12.5% |
| Enc Time[%] | 108.7% | | | 108.0% | | | 110.0% | | |
| Dec Time[%] | 130.0% | | | 129.2% | | | 137.4% | | |
| Enc Mem[%] | #NUM! | | | #NUM! | | | #NUM! | | |
| 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 | -2.9% | -8.4% | -7.7% |  |  |  | -3.1% | -9.1% | -9.3% |
| Class B | -2.4% | -4.6% | -5.0% | -3.4% | -7.2% | -7.5% | -3.0% | -7.4% | -8.7% |
| **Overall (Test vs Ref)** | -2.5% | -5.7% | -5.7% | -3.4% | -7.2% | -7.5% | -3.0% | -7.9% | -8.9% |
| **Overall (Test vs single layer)** | 25.3% | 30.7% | 32.2% | 20.6% | 23.7% | 25.6% | 20.6% | 23.7% | 26.6% |
| **EL only (Test vs Ref)** | -4.8% | -7.8% | -7.8% | -8.9% | -12.5% | -12.7% | -5.4% | -10.4% | -11.4% |
| Enc Time[%] | 124.1% | | | 114.9% | | | 110.7% | | |
| Dec Time[%] | 159.0% | | | 156.5% | | | 158.4% | | |
| Enc Mem[%] | #NUM! | | | #NUM! | | | #NUM! | | |
| BL Match | Matched | | | Matched | | | Matched | | |

# Conclusions

In this proposal, an ILR enhancement technology was proposed for the RefIdx framework. The proposed method uses weighted differential signal to enhance the high frequency information in the inter-layer reference (ILR) picture. Simulation results showed substantial coding performance improvements. We propose to adopt the ILR enhancement method into SHVC.

# Patent rights declaration(s)

**InterDigital Communications, LLC 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 1 (SHM 1)”, JCTVC-L1007, Jan. 2013.
2. X. Xiu, Y. He, Y. He, Y. Ye, “TE5: Results on test 5.4.1 on motion field mapping”, JCTVC-L0052, Jan. 2013.
3. A. Aminlou, J. Lainema, K. Ugur, M. Hannuksela, “Differential coding for RefIdx based scalability”, JCTVC-L0213, Jan. 2013.
4. X. Li, E. François, P. Lai, D. Kwon, A. Saxena, “Description of Core Experiment SCE 3: Combined Inter and Inter-Layer Prediction in SHVC”, JCTVC-1103, Jan. 2013.