|  |  |
| --- | --- |
| **Joint Collaborative Team on Video Coding (JCT-VC)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  16th Meeting: San José, USA, 9 – 17 Jan. 2014 | Document: JCTVC-P0186 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **SCE1: Combined bit-depth and color gamut conversion with 3D LUT for SHVC color gamut scalability** | | |
| *Status:* | Input Document to JCT-VC | | |
| *Purpose:* | Proposal | | |
| *Author(s) or Contact(s):* | Yuwen He, Yan Ye, Jie Dong  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) [jie.dong@interdigital.com](mailto:jie.dong@interdigital.com) |
| *Source:* | InterDigital Communications, Inc. | | |

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

# Abstract

This proposal tested the combined bit-depth and color gamut conversion method with online 3D LUT derivation for SHVC color gamut scalability (CGS) proposed in JCTVC-O0161 with SCE1 test conditions. Two usecases with two tests are considered. For usecase 1 test, compared to SCE1 anchors, the proposed scheme reportedly achieves average {Y, U, V} BD rate gain of {-3.1%, -5.6%, -5.0%} and {-3.9%, -5.1%, -7.1%} for AI and RA-2x, respectively. For usecase 2 test, the proposed scheme reportedly achieves average {Y, U, V} BD rate gain of {-7.8%, -8.9%, -11.5%}, and {-6.2%, -6.5%, -10.8%} for AI and RA-2x, respectively.

# Introduction

The combined bit-depth and color gamut conversion method with 3D LUT for SHVC color gamut scalability was proposed in JCTVC-O0161 [2], where 3D LUT was trained offline. This proposal reported the performance of 3D LUT with SCE1 test conditions [1]. 3D LUT derivation process is integrated into the encoder based on SCE1 anchor software released by Technicolor.

# 3D LUT derivation and coding

The 3D LUTs used in our simulations are online trained. The flow chart of SHVC encoder with 3D LUT online training is shown in Figure 1.



Figure 1. SHVC encoder with 3D LUT online training

The target picture for 3D LUT training is the downsampled picture from the EL input. The input source for 3D LUT training is the BL reconstructed picture. The 3D LUT is trained using Least Square method. The tetrahedral interpolation used in 3D LUT based CGS is a linear function and it can be described as Equation (1) in matrix form.

(1)

is the i-th converted signal of component c, where c can be Y, U or V in BT.2020. i is the index of input pixel. N is the total number of input pixels. \* in Equation(1) is matrix multiplication. is the weighting vector for the i-th input pixel, and is expressed as below.

is the weight of the j-th output entry of 3D LUT for the i-th input pixel, which is calculated according to tetrahedral interpolation Equations in [2].

is the parameter vector to be estimated, that is the output entries of 3D LUT, and can be written as below:

M is the number of 3D LUT output entries, for example M is 729 for 9x9x9 sized 3D LUT. For simplicity, we will omit the component c in the following equations because the 3D LUT of each component can be estimated independently. If we aggregate Equation (1) for all pixels, then we get:

(2)

(3)

With a given input signal, not all 3D LUT entries can be filled. Therefore, we obtain a compact parameter set *P’* by removing all those unused 3D LUT vertices from P, and we build a mapping from *P’* to *P*. We train the entries in *P’*, and then use the mapping to reconstruct *P* after *P’* is estimated.

We can calculate *W’* and *H’* accordingly using the compacted *P’*. With Least Square estimation, the solution is defined as:

(4)

Where is calculated based on using the following:

, (5)

After the compact parameter vector *P’* is estimated, then the full parameter vector *P* can be obtained by mapping those estimated vertices from *P’* to *P*.

Those un-estimated vertices in *P* are filled using the corresponding vertices in *P’* and the trilinear interpolation in 3D LUT encoding process [4].

For use case 1, only the first picture of the sequence is used for the 3D LUT derivation. For use case 2, we use all previous IRAP pictures in encoding order for 3D LUT training and apply derived 3D LUT for current IRAP period.

We use the same entropy coding method as in [4] for 3D LUT coding, and the 3D LUT is signaled in the PPS extension, in the same way as in [4].

# Simulation results

The compression performance is measured using BD rate compared with SCE1 anchors, using the SCE1 test conditions [1]. There are two tests (test A/B) listed in Table 2 of SCE1 document [1] for each of these two use cases. In use case 1, only the first IRAP picture is used for 3D LUT training. The derived 3D LUT is sent once at the beginning of the sequence and applied to all pictures in the sequence. In use case 2, 3D LUTs are derived periodically at about 1 second interval (64 pictures for 60Hz and 48 pictures for 50Hz); the 3D LUTs are sent at ~1 second interval and applied to all pictures in that ~1 second period. The 3D LUT size is 9x9x9 for all tests. Table 1 gives the detailed average BD rate reduction for use case 1 compared with SCE1 anchors. Table 2 gives the detailed average BD rate reduction for use case 2 compared with SCE1 anchors.

As shown in Table 1, compared with SCE1 anchors, the proposed 3D LUT with the first picture online training achieves average {Y, U, V} BD rate gain of {-3.1%, -5.6%, -5.0%}, and {-3.9%, -5.1%, -7.1%} for AI and RA-2x, respectively. As shown in Table 2, the proposed 3D LUT with periodical updating method achieves average {Y, U, V} BD rate gain of {-7.8%, -8.9%, -11.5%}, and {-6.2%, -6.5%, -10.8%} for AI and RA-2x, respectively.

Readers are referred to the accompanying spreadsheets for further details. The encoding time is not reliable because the cores in grid for simulations are not homogeneous. The decoding time for both anchor and test was collected from the same workstation.

“SHM4.0\_irap\_UC1\_anchor\_vs\_3DLut.xls” is the data sheet for the 3D LUT results of usecase 1 compared with SCE1 anchors.

“SHM4.0\_irap\_UC2\_anchor\_vs\_3DLut.xls” is the data sheet for the 3D LUT results of usecase 2 compared with SCE1 anchors.

Table 1. Average BD rate reduction for usecase 1 compared with SCE1 anchors

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Test B**  **AI HEVC 2x 10-bit base** | | | **Test A**  **AI HEVC 2x 8-bit base** | | |
|  | Y | U | V | Y | U | V |
|  |  |  |  |  |  |  |
| Class A+ | -2.9% | -5.5% | -5.0% | -3.4% | -5.8% | -5.1% |
| **Overall (Test vs Ref)** | -2.9% | -5.5% | -5.0% | -3.4% | -5.8% | -5.1% |
| **Overall (Test vs single layer)** | 15.3% | 16.1% | 12.7% | 17.3% | 17.5% | 13.9% |
| **Overall (Ref vs single layer)** | 18.5% | 22.8% | 18.1% | 21.2% | 24.7% | 19.4% |
| **EL only (Test vs Ref)** | -6.0% | -8.5% | -7.4% | -6.9% | -9.2% | -7.9% |
| **Overall (Test EL+BL vs single EL+BL)** | -23.2% | -22.9% | -25.3% | -22.0% | -22.1% | -24.8% |
| Enc Time[%] | 99.9% | | | 79.2% | | |
| Dec Time[%] | 100.7% | | | 101.1% | | |
| BL Match | Matched | | | Matched | | |
|  |  |  |  |  |  |  |
|  | **Test B**  **RA HEVC 2x 10-bit base** | | | **Test A**  **RA HEVC 2x 8-bit base** | | |
|  | Y | U | V | Y | U | V |
|  |  |  |  |  |  |  |
| Class A+ | -3.8% | -5.1% | -7.3% | -4.0% | -5.2% | -6.9% |
| **Overall (Test vs Ref)** | -3.8% | -5.1% | -7.3% | -4.0% | -5.2% | -6.9% |
| **Overall (Test vs single layer)** | 22.6% | 27.1% | 17.4% | 24.0% | 28.4% | 18.9% |
| **Overall (Ref vs single layer)** | 27.5% | 33.7% | 26.9% | 29.2% | 35.2% | 27.8% |
| **EL only (Test vs Ref)** | -6.9% | -7.9% | -10.0% | -7.2% | -8.1% | -9.7% |
| **Overall (Test EL+BL vs single EL+BL)** | -17.7% | -14.0% | -20.7% | -17.0% | -13.5% | -20.0% |
| Enc Time[%] | 69.7% | | | 57.7% | | |
| Dec Time[%] | 107.1% | | | 107.4% | | |
| BL Match | Matched | | | Matched | | |

Table 2. Average BD rate reduction for usecase 2 compared with SCE1 anchors

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Test B**  **AI HEVC 2x 10-bit base** | | | **Test A**  **AI HEVC 2x 8-bit base** | | |
|  | Y | U | V | Y | U | V |
|  |  |  |  |  |  |  |
| Class A+ | -7.8% | -8.9% | -11.6% | -7.8% | -8.9% | -11.5% |
| **Overall (Test vs Ref)** | -7.8% | -8.9% | -11.6% | -7.8% | -8.9% | -11.5% |
| **Overall (Test vs single layer)** | 9.2% | 11.8% | 4.7% | 11.8% | 13.6% | 6.1% |
| **Overall (Ref vs single layer)** | 18.5% | 22.8% | 18.1% | 21.2% | 24.7% | 19.4% |
| **EL only (Test vs Ref)** | -15.2% | -15.9% | -18.6% | -15.2% | -15.9% | -18.4% |
| **Overall (Test EL+BL vs single EL+BL)** | -28.3% | -26.5% | -31.9% | -26.6% | -25.4% | -31.1% |
| Enc Time[%] | 90.9% | | | 93.0% | | |
| Dec Time[%] | 98.3% | | | 98.6% | | |
| BL Match | Matched | | | Matched | | |
|  |  |  |  |  |  |  |
|  | **Test B**  **RA HEVC 2x 10-bit base** | | | **Test A**  **RA HEVC 2x 8-bit base** | | |
|  | Y | U | V | Y | U | V |
|  |  |  |  |  |  |  |
| Class A+ | -6.3% | -6.5% | -11.0% | -6.2% | -6.5% | -10.6% |
| **Overall (Test vs Ref)** | -6.3% | -6.5% | -11.0% | -6.2% | -6.5% | -10.6% |
| **Overall (Test vs single layer)** | 19.3% | 25.1% | 12.7% | 21.2% | 26.6% | 14.3% |
| **Overall (Ref vs single layer)** | 27.5% | 33.7% | 26.9% | 29.2% | 35.2% | 27.9% |
| **EL only (Test vs Ref)** | -11.7% | -11.4% | -16.0% | -11.5% | -11.4% | -15.4% |
| **Overall (Test EL+BL vs single EL+BL)** | -20.5% | -15.7% | -24.6% | -19.4% | -15.0% | -23.8% |
| Enc Time[%] | 64.0% | | | 65.1% | | |
| Dec Time[%] | 106.1% | | | 106.3% | | |
| BL Match | Matched | | | Matched | | |

The gain of use case 1 is much lower than that of use case 2, primarily because the hue of two test sequences “BirthdayFlashPart1” and “BirthdayFlashPart2” changed in the middle of the sequence, as shown in Figure 2 and Figure 3. Therefore, the 3D LUT derived with the first picture in use case 1 is not so effective for those pictures after the hue changed.



1. Picture 0 (b) Picture 140

Figure 2. Hue changes of BirthdayFlashPart1



1. Picture 0 (b) Picture 100

Figure 3. Hue changes of BirthdayFlashPart2

# Conclusions

In this proposal, the bit-depth conversion and color gamut conversion are combined into one combined 3D LUT for SHVC color gamut scalability coding. Compared with SCE1 anchors, which uses weighted prediction the ILR pictures and simple bit-shifting if the bit depth is different between the two layers, 3D LUT method can achieve higher coding efficiency, with average BD rate gains of {-3.1%, -5.6%, -5.0%} for AI, {-3.9%, -5.1%, -7.1%} for RA-2x for use case 1, and {-7.8%, -8.9%, -11.5%} for AI, {-6.2%, -6.5%, -10.8%} for RA-2x for use case 2.

# 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. P. Bordes, Y. Ye, E. Alshina, X. Li, S.-H. Kim, A. Duenas, K. Ugur, K. Sato, “Description of HEVC Scalable Extensions Core Experiment SCE1: Color Gamut and Bit-Depth Scalability”, JCTVC-O1101, Oct. 2013, CH.
2. Y. He, Y. Ye, J. Dong, “Non-SCE4/AHG14: Combined bit-depth and color gamut conversion with 3D LUT for SHVC color gamut scalability”, JCTVC-O0161, Oct. 2013, CH.
3. P. Bordes, P. Andrivon, F. Hiron, P. Salmon, R. Boitard, “SCE4: Results on 5.3-test1 and 5.3-test2”, JCTVC-O0159, Oct. 2013, CH.
4. P. Bordes, P. Andrivon, F. Hiron, “AHG14: Color Gamut Scalable Video Coding using 3D LUT: New Results”, JCTVC-N0168, Jul. 2013.