|  |  |
| --- | --- |
| **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-M0197-r2 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **AHG14: Color Gamut Scalable Video Coding using 3D LUT** | | |
| *Status:* | Input Document to JCT-VC | | |
| *Purpose:* | Proposal | | |
| *Author(s) or Contact(s):* | Philippe Bordes,  Pierre Andrivon,  Roshanak Zakizadeh  975 avenue des champs blancs  CS 17616, 35576, Cesson-Sévigné Cedex, France | Tel: Email: | +33-2-99-27-32-42 [philippe.bordes@technicolor.com](mailto:philippe.bordes@technicolor.com) |
| *Source:* | Technicolor | | |

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

# Abstract

This contribution proposes a new model of inter-layer prediction for color gamut scalable video coding based on 3D color Look-Up Tables (LUT). It is asserted the application requirements in term of color gamut scalable video coding is not limited to simple transformation between Base layer color space (ex: Rec.709) and Enhancement layer color space (ex: Rec.2020), but also the cases the Base layer and the Enhancement layer has been color graded differently.

It is asserted the size of the 3D LUT can be chosen to meet both the application complexity requirements and the BD-rate distortion trade-off: small size for describing simple color mapping transformation between Base layer and Enhancement layer, and larger size for representing more complex color differences in-between the layers.

It is reported the proposed 3D LUT model can be equivalent to the Gain-Offset model proposed in JCTVC-L0334, with appropriate 3D LUT parameters settings.

Some results are presented in 8 bits with two examples of color grading functions showing enhancement layer bit-rate savings between 13% and 29% (luma) and between 12% and 39% (chroma) compared to SHM1.0 for AI and RA SNR scalability scenarios, and enhancement layer bit-rate savings between 15% and 32% (luma) and between 15% and 41% (chroma) compared to SHM1.0 for AI and RA spatial scalability scenarios. Comparisons made with Gain-Offset model show enhancement layer bit rate saving between 5% and 20% (luma) and between 6% and 32% (chroma).

Comparisons made with Gain-Offset model in case of simple Rec.709/Rec.2020 color space conversion show the performances are equivalent.

# Introduction / Problem Statement

Color Gamut Scalability has been identified as one requirement of Scalable Coding Extension of HEVC [1]. It allows addressing the cases the original Enhancement Layer uses a different color gamut than the Base Layer. This can be useful for instance in case of deployment of UHD services compatible with legacy HD devices: HD is using the Rec.709 [2], while UHD is likely to use some of the parameters defined in the Rec.2020 [3].

The general diagram of a scalable video encoder including a prediction tool for color differences between the base layer (BL) and enhancement layer (EL) has been proposed in [4] and is shown in Figure 1.



Figure : Color Space Scalable Encoder (courtesy from Sharp).

Basically, the role of the Color Difference Predictor module is to predict the EL color samples from the collocated BL color samples. However, for a given pair of BL and EL video sequence, the determination of this color transfer function is not straightforward because the content creation workflows may include deterministic processing (Color Space 1 vs Color Space 2 conversion) but also non-deterministic operations, for the reasons explained below:

* The last digital cameras used in Digital Cinema (DC) allow to capture video signal with wide color and luminance range. The output raw data are Wide Color Gamut (WCG) that may be beyond DCI-P3-gamut, with potentially extended dynamic range (more than 14 f-stops). The floating point or 16 bits raw data represent much more information than what will be distributed and displayed finally.

This trends will probably increase in the coming years, the new captors being able to capture several exposure directly for instance.

* Bit-depth scaling / tone mapping: the choice of the luminance range mapping versus the Reference Output Display characteristics (ODT) (ex: 16 bits vs 8 or 10 bits) is made by human operator depending on artistic intents.
* Color grading (artistic intent): the color balancing is of key importance in content creation, because it traduces the artistic intent of the film director and has major impact on the final rendering. It is performed by high-skilled graphic operators (colorists), using a reference display. Then, if two targeted displays with different characteristics are used (e.g. DCI projectors for DC and Rec.709 TV for HDTV), the artistic intent may be different and the color grading may be different too. These graphic designers use special color authoring tools using 3D LUTs to represent/output their color processing operations.
* Color Space conversion: currently and pragmatically Rec.709-Rec.2020 has been identified as a probable color space conversion use case (e.g. scalable HDTV and UHDTV). This conversion is basically not linear. Besides, new video signal definitions may appear in the coming decade, fueled by the increasing display technology capabilities (e.g. OLED displays…). Some applications should require to adapt content to the end-device rendering characteristics/capabilities consequently.



Figure : Hypothetical scalable HD/UHD processing workflow, inspired from (simplified) Digital Cinema workflow.

An hypothetical scalable HD/UHD processing workflow is depicted in Figure 2. A more precise definition of a digital motion picture workflow is proposed by the Academy Color Encoding System (ACES). It may be used to create content for movie theater or for physical media distribution such as DVD or Blu-ray disc typically.

Consequently, the “Color Difference Predictor” function model is highly unpredictable and may have very various shape. This justifies to describe it based on a generic and flexible model.

# Proposed Color Difference predictor

In order to be able to address a wide range of Color Gamut Scalability applications, without any a-priori on the “Color Difference Predictor” model, we propose to use a 3D Look-Up Table (LUT) to convert/predict the tri-chromatic samples from the color space 1 (BL) to the color space 2 (EL). This 3D LUT approach is the one chosen in post-production facilities to process content from raw to targeted device color spaces. The principle of the 3D LUT is depicted in Figure 3: the 3D LUT can be considered as a sub-sampling of the 3D color space 1, where each vertex is associated with a color triplet corresponding to the color space 2 (predicted) values. For a given BL color sample (color space 1), the computation of its prediction in (EL) color space 2 is made using tri-linear interpolation as depicted in Figure 4 and equation (1).

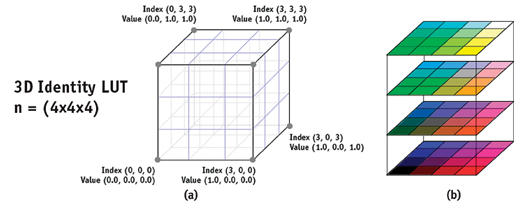


Figure : Principle of the 3D Color Look-Up Table (LUT).



Figure : Tri-linear interpolation with 3D LUT (courtesy of Nvidia).

(1)

Where:

S0(y) = y1-y and S1(y) = y-y0

y0 is the index of the nearest sub-sampled vertex inferior to y,

y1 is the index of the nearest sub-sampled vertex superior to y.

In order to encode the 3D LUT data efficiently, each color component of a vertex is encoded with previously encoded color components of neighboring vertices. We propose also to use an octree based description of the 3D LUT in such a way the unused (or less used) 3D color space regions are encoded with coarsely lattice size as depicted in Figure 5.

At the decoder side, all the not coded vertices inside one octant are interpolated to reconstruct the full definition 3D LUT.

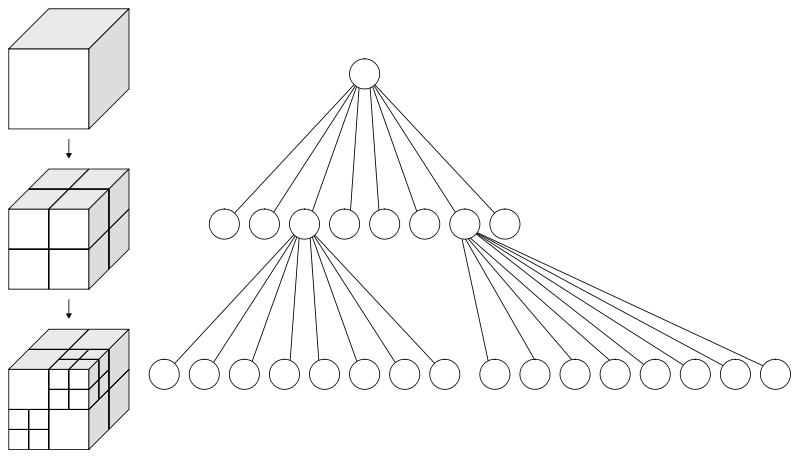


Figure : Octree based 3D LUT: each octant is encoded with 8 vertices at most.

This approach has several advantages:

* Many Color processing tools uses 3D LUTs to represent and save their intermediate and final color grading operations. In these cases, the 3D LUT information can be made available to the encoder easily.
* We propose that the size of the 3D LUT (number of vertices in one direction), is a parameter read in the bit-stream. In that way, the encoder may choose the best trade-off between “Color Difference Predictor” module accuracy and encoding cost.
* The 3D LUT using tri-linear interpolation has the following property: if its size is 1, ie one single octant with 8 vertices, and if the color components of the vertices are set appropriately (see 6-Annex), the 3D LUT is equivalent to the Gain-Offset model previously proposed in [4].
* At last, 3D color LUT interpolation module for color conversion is implemented in many STBs and display devices (graphics card).

# Results

## Use of Color grading LUTs

To validate the 3D LUT approach, we used a color grading tool (CineStyle [5]) to generate two 3D Color LUTs (Morning and FilmEmulation), corresponding to typical pre-defined color look styles (see Annex 7). Then we encoded the EL color graded and BL original Rec. 709 ParkScene class B sequence (8 bits). In the Table 1 we show the gains obtained with the modified SHM1.0 (SHM1.0-3DLUT) implementing the 3D LUT Color Difference Prediction mode and the SHM1.0 (qpEL = qpBL+2). The LUTs are loaded by the encoder at the initialization stage and encoded in VPS as described above.

We also compare our model with the Gain-Offset (GO) model proposed in [4]. For this purpose, for each tested sequence we computed the GO matrix parameters by least mean square minimization, with the first pictures of the sequences. Then we modified the SHM1.0 (SHM1.0-GO) to implement the GO model. The results are depicted in Table 1.

Table 1 show the 3D LUT model out-performs the GO model for these color graded sequences. The 3D LUT model allows to better model the non-linearity of the color difference predictor.

Table : BD rate performance gain of SHM1.0-3DLUT and SHM1.0-GO vs SHM1.0 .

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **AI\_SNR** | **BL+EL** |  |  | **EL** |  |  |
| *qpEL = qpBL+2* | Y | U | V | Y | U | V |
| **SHM1.0-3DLUT** |  |  |  |  |  |  |
| CineStyle\_Morning\_B\_ParkScene | -12.3% | -11.7% | -17.2% | -19.1% | -18.5% | -23.4% |
| CineStyle\_FilmEmulation\_B\_ParkScene | -20.5% | -33.4% | -23.9% | -29.6% | -39.9% | -32.6% |
| **SHM1.0-GO** |  |  |  |  |  |  |
| CineStyle\_Morning\_B\_ParkScene | -6.3% | -10.3% | -6.4% | -9.2% | -12.8% | -9.3% |
| CineStyle\_FilmEmulation\_B\_ParkScene | -6.2% | -4.6% | -9.9% | -9.1% | -7.5% | -12.6% |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **AI\_1.5x** | **BL+EL** |  |  | **EL** |  |  |
| *qpEL = qpBL+2* | Y | U | V | Y | U | V |
| **SHM1.0-3DLUT** |  |  |  |  |  |  |
| CineStyle\_Morning\_B\_ParkScene | -15.7% | -14.8% | -19.2% | -27.8% | -26.7% | -30.6% |
| CineStyle\_FilmEmulation\_B\_ParkScene | -21.5% | -31.7% | -23.8% | -32.3% | -41.0% | -34.2% |
| **SHM1.0-GO** |  |  |  |  |  |  |
| CineStyle\_Morning\_B\_ParkScene | -7.5% | -11.4% | -6.7% | -11.6% | -15.1% | -10.6% |
| CineStyle\_FilmEmulation\_B\_ParkScene | -10.2% | -8.8% | -13.0% | -17.6% | -16.1% | -20.0% |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **AI\_2x** | **BL+EL** |  |  | **EL** |  |  |
| *qpEL = qpBL+2* | Y | U | V | Y | U | V |
| **SHM1.0-3DLUT** |  |  |  |  |  |  |
| CineStyle\_Morning\_B\_ParkScene | -11.2% | -10.4% | -12.2% | -15.8% | -15.0% | -16.7% |
| CineStyle\_FilmEmulation\_B\_ParkScene | -13.6% | -20.0% | -16.5% | -18.3% | -24.3% | -21.0% |
| **SHM1.0-GO** |  |  |  |  |  |  |
| CineStyle\_Morning\_B\_ParkScene | -7.6% | -7.1% | -8.2% | -10.6% | -10.1% | -11.0% |
| CineStyle\_FilmEmulation\_B\_ParkScene | -4.7% | -5.9% | -4.7% | -6.4% | -7.5% | -6.3% |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **RA\_SNR** | **BL+EL** |  |  | **EL** |  |  |
| *qpEL = qpBL+2* | Y | U | V | Y | U | V |
| **SHM1.0-3DLUT** |  |  |  |  |  |  |
| CineStyle\_Morning\_B\_ParkScene | -9.3% | -8.4% | -13.1% | -13.4% | -12.5% | -17.1% |
| CineStyle\_FilmEmulation\_B\_ParkScene | -13.0% | -27.8% | -14.6% | -17.4% | -31.5% | -18.9% |
| **SHM1.0-GO** |  |  |  |  |  |  |
| CineStyle\_Morning\_B\_ParkScene | -4.8% | -3.6% | -8.0% | -6.6% | -5.4% | -9.8% |
| CineStyle\_FilmEmulation\_B\_ParkScene | -3.3% | -6.3% | -2.6% | -4.5% | -7.4% | -3.7% |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **RA\_1.5x** | **BL+EL** |  |  | **EL** |  |  |
| *qpEL = qpBL+2* | Y | U | V | Y | U | V |
| **SHM1.0-3DLUT** |  |  |  |  |  |  |
| CineStyle\_Morning\_B\_ParkScene | -13.4% | -11.1% | -15.6% | -22.7% | -20.2% | -24.4% |
| CineStyle\_FilmEmulation\_B\_ParkScene | -14.7% | -26.5% | -14.9% | -21.4% | -32.3% | -21.5% |
| **SHM1.0-GO** |  |  |  |  |  |  |
| CineStyle\_Morning\_B\_ParkScene | -8.6% | -6.0% | -10.7% | -14.1% | -11.3% | -15.9% |
| CineStyle\_FilmEmulation\_B\_ParkScene | -4.4% | -7.4% | -2.0% | -6.6% | -9.4% | -4.1% |

## Color Space conversion Rec.709 to Rec.2020

In [4], the authors considered another use case where color space of BL is Rec.709 and color space of EL is Rec.2020. For this purpose, we created Rec.2020 EL sequences from regular HEVC sequences with conversion to and from the CIE-XYZ linear color space, as depicted in Figure 6.



Figure : Y’CbCr Rec.709 to Rec.2020 conversion workflow.

The same workflow is used to build the Rec.709-to-Rec.2020 3D LUT, of size 9x9x9, by sub-sampling the color space 1 uniformly. To compare with Gain-Offset (GO) model proposed in [4], for each tested sequence we computed the GO matrix parameters by least mean square minimization. While the SHM1.0-3DLUT uses the same 3D LUT for all the contents. The BD-rate gains vs SHM1.0 are depicted in Table 2 and Table 3 for GO model and for the proposed 3D-LUT model respectively. The figures show equivalent performances.

Table : BD rate performance gain of SHM1.0-GO vs SHM1.0 (8 bits).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Average Class B | **BL+EL** |  |  | **EL** |  |  |
|  | Y | U | V | Y | U | V |
| **AI\_SNR** | -3.8% | 0.3% | -14.1% | -7.0% | -3.0% | -17.5% |
| **AI\_15x** | -3.2% | -1.1% | -23.4% | -7.3% | -4.8% | -30.9% |
| **RA\_SNR** | -3.2% | 3.8% | -21.6% | -5.8% | 1.7% | -25.0% |
| **RA\_15x** | -3.3% | -0.9% | -33.1% | -7.7% | -4.9% | -39.0% |

Table : BD rate performance gain of SHM1.0-3DLUT vs SHM1.0 (8 bits).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Average Class B | **BL+EL** |  |  | **EL** |  |  |
|  | Y | U | V | Y | U | V |
| **AI\_SNR** | -4.0% | -4.0% | -16.1% | -7.4% | -7.3% | -19.6% |
| **AI\_15x** | -2.9% | -4.6% | -24.5% | -5.2% | -7.0% | -30.9% |
| **RA\_SNR** | -3.4% | -1.6% | -23.3% | -6.1% | -4.1% | -26.8% |
| **RA\_15x** | -3.2% | -3.7% | -35.2% | -7.3% | -7.5% | -41.0% |

# Proposed syntax

It is proposed to encode the 3D Color LUT data into the PPS or the VPS.

As in [4], it is proposed to signal the Color Gamut prediction mode by adding an additional syntax elements to PPS when **pps\_extension\_flag** is set. The syntax element, use\_color\_prediction, is used to indicate the use of color prediction in the decoding of coded pictures as shown in Table 4.

Table : signaling of the prediction parameters.

|  |
| --- |
| **pps\_extension\_flag** |
| if( pps\_extension\_flag ) |
| **use\_color\_prediction** |
| if(use\_color\_prediction) |
| **3D\_ LUT\_ color\_data ( )** |
| rbsp\_trailing\_bits( ) |

If the **use\_color\_prediction** flag is equal to ‘1’, the **3D\_ LUT\_ color\_data** function is called tosignal 3D color LUT data as shown inTable 5.

Table : coding of the 3D LUT color data.

|  |  |
| --- | --- |
| 3D\_ LUT\_ color\_data ( ) { | **Descriptor** |
| **nbp\_code** | u(3) |
| coding\_octant(0, 0, 0, 0) |  |
| } |  |

**nbp\_code** indicates the 3D LUT size as listed in Table 6 for the given value of nbp\_code.

Table : Interpretation of nbp\_code.

|  |  |
| --- | --- |
| **nbp\_code** | **3D LUT size** |
| 0 | 2 |
| 1 | 3 |
| 2 | 5 |
| 3 | 9 |
| 4 | 17 |
| 5 | 33 |

The decoding of the octant( layer, y,u,v) is a recursive function. Each octant is composed of 8 vertices associated with a flag (encoded\_vertex\_flag[i]) indicating whether the residual components values are encoded or all inferred to be zero. The component values are reconstructed by adding the residuals to the prediction of the components values. The prediction of the components values is computed using tri-linear interpolation of the 8 neighboring vertices of layer-1.

Table : syntax elements for coding\_octant().

|  |  |
| --- | --- |
| coding\_octant ( layer, y,u,v) { | **Descriptor** |
| for( i = 0; i < 8 ; i++ ) { |  |
| **encoded\_vertex\_flag**[i] | u(1) |
| if ( encoded\_vertex\_flag[i] ) { |  |
| **resY**[j] | ue(v) |
| **resU**[j] | ue(v) |
| **resV**[j] | ue(v) |
| **}** |  |
| } |  |
| **split\_octant\_flag** | u(1) |
| if ( split\_octant\_flag ) { |  |
| for( i = 0; i < 8 ; i++ ) { |  |
| coding\_octant ( layer+1, y+dy[i],u+du[i],v+dv[i]) |  |
| } |  |
| } |  |
| } |  |

Where (y+dy[i]), (u+du[i]) and (v+dv[i]) are the 8 children octants coordiantes (coordinates of the first 3D color vertex) of the current octant (having (y,u,v) as first vertex coordinates).

# Conclusion

A scalable system for supporting different color gamuts and/or different color grading in different layers of a bit-stream is proposed. The experiments focused on the case of a 8-bit Rec.709 base layer, and 8-bit generic color graded enhancement layer and on a 8-bit Rec.709 base layer and 8-bit Rec.2020 enhancement layer, but this is not a restriction.

The proposed model is scalable in complexity and hence accuracy: it allows addressing low complexity use cases equivalent to Gain Offset prediction model previously proposed in JCTVC-L0334, but also to address more general use cases where the Color mapping function in between the Base and Enhancement layers is more complex than a simple linear transformation.

The model is based on 3D color LUT that is a tool the graphists and colorists are familiar with, hence can adapt to existing video workflows naturally.

We suggest JCT-VC to create a HEVC Scalable Extensions Core Experiment in order to assess Color Gamut Scalability inter-layer prediction related methods.

# Annex: equivalence between GO and 3D LUT models

In case 3D LUT size is 1 (one single octant with 8 vertices), one can choose the three color components of the vertices in order to the 3D LUT is equivalent to the Gain-Offset model (2) proposed in [4].

(2)

To do so, one has to set the 3 color components (yX,uX,vX)X=A,…H of the 8 vertices (Figure 7) with the following values:

Then, the tri-linear interpolation is equivalent to the Gain-Offset model (2).



Figure : The 8 vertices (A,B,C,D,E,F,G,H) of the first 3D LUT octant.

# Annex: Test sequences

Figure 8: Original rec709 (left) and converted rec2020 (right) ParkScene (class B) sequence.

Figure 9: Color graded with CineStyle, using FilmEmulation (left) and Morning (right) “looks”.

# References

1. Ajay Luthra, Jens-Rainer Ohm, Jörn Ostermann, “ Requirements of the scalable enhancement of HEVC,” WG11 Requirements and Video, ISO/IEC JTC1/SC29/WG11 N12783, May 2012, Geneva, Switzerland.
2. ITU-R Recommendation BT.709 “Parameter values for the HDTV standards for production and international programme exchange” Dec. 2010.
3. ITU-R Recommendation BT.2020 “Parameter values for UHDTV systems for production and international programme exchange” April 2012.
4. L.Kerofsky, A.Segall, SH.Kim, K.Misra, “Color Gamut Scalable Video Coding: New Results,” JCTVC-L0334, 12th JCTVC Meeting, Geneva, CH, 14–23 Jan. 2013.
5. <https://www.technicolorcinestyle.com/cinestyle-color-assist/>

# Patent rights declaration(s)

**Technicolor may have current or pending patent rights 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).**