|  |  |
| --- | --- |
| **Joint Collaborative Team on Video Coding (JCT-VC)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  12th Meeting: Geneva, CH, 14–23 Jan. 2013 | Document: JCTVC-L0162 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **AHG7: Coding 4:2:2 chroma format with 4:2:0 and 4:4:4 format codecs** | | |
| *Status:* | Input Document | | |
| *Purpose:* | Proposal | | |
| *Author(s)* | Andrea Gabriellini Marta Mrak | andrea@rd.bbc.co.uk marta.mrak@bbc.co.uk |  |
| *Source:* | BBC Research & Development | | |

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

**Abstract**

This contribution presents an approach to coding content in 4:2:2 chroma format. The proposed approach assumes resampling of the input signal to the encoder to the 4:4:4 or 4:2:0 chroma format, i.e. it requires the horizontal or vertical up- or downsampling of the two chroma components by a factor of 2. All coding operations are then carried out at the new chroma format. The decoder is signalled the filter required to reverse the input resampling, i.e. to obtain a signal in the original chroma format (4:2:2). Tests are performed for the scenario where 4:2:2 chroma is upsampled to 4:4:4 at the encoder and the downsampling filters are sent to the decoder. Two different filters were used for this evaluation to demonstrate feasibility of this approach as well as the performance and implementation challenges. When 4:2:2 content is coded in 4:4:4 format, bit-rate increases of up to 5.1% for the luma component are observed, with run time increase of about 10% for encoding and about 25% for decoding, compared to coding directly in 4:2:2 format. Both applied filters are reversible filters which can support lossless format conversion.

# Introduction

The processing of content in 4:2:2 chroma format introduces some coding challenges in the form of non-square chroma component blocks. This shape requires adjusting a number of compression elements, compared to uniform chroma sampling formats (4:2:0 and 4:4:4), [1]. Moreover, when 4:2:0 videos are used in environments with 4:2:2 serial interfaces, additional challenges are introduced related to resampling chroma components, [2]. Generally, when codecs which support pixel-level coding of 4:2:2 format are not available, codecs with 4:2:0 and 4:4:4 format support for pixel-level coding may be used instead.

To enable this feature, the proposed concept enables resampling of chroma components at the end of decoding, providing output 4:2:2 format, where the resampling is defined by the encoder. It is assumed that the 4:2:2 coding on pixel-level, as being developed within AHG7 on Support for range extensions (e.g. methods to support coding of non-square CTB [3]), would still be supported in a different environment. Therefore the proposed method is another option for 4:2:2 format coding.

The main challenge related to resampling filter choice is the potential quality degradation of a signal, especially when cascaded coding is used. This problem is discussed in the next section, which also addresses possible filter designs. The results are presented for experiments where sequences in 4:2:2 format are coded with 4:4:4 format codec.

# Coding sequences resampled from 4:2:2 format

In order to support coding of sequences in 4:2:2 format using codecs for 4:2:0 or 4:4:4 formats the input chroma components have to be resampled before actual encoding. In case of coding using 4:2:0 codec, vertical downsampling is required which should incorporate compensation for chroma sample position. This problem has been studied in the past and efficient solutions are available ([4], [5]). In case of coding using 4:4:4 codec the input chroma components in 4:2:2 format, the horizontal upsampling is required, Figure 1.



Figure 1: Sequence in 4:2:2 format coded with 4:4:4 codec.   
For downsampling at the decoder a set of filter coefficients is provided in the bit-stream.

In both cases (4:2:0 or 4:4:4 coding), the resampling filter used at the decoder should be jointly designed with the corresponding encoder's filters. In order to enable selection of decoder's resampling filter it is desirable to enable signalling of the filter coefficients to the decoder.

Compared to coding in 4:2:0 format, the main advantage of coding in 4:4:4 format is in its capability to maintain higher quality of chroma components by avoiding signal distortion during encoder's downsampling. Ideally, up- and downsampling filters used in the 4:4:4 coding should be reversible to minimise distortion introduced during resampling. This document therefore focuses on 4:4:4 format challenges of the proposal.

While a degree of signal degradation introduced to the chroma signal by non-reversible filters in a single encoding/decoding cycle may be negligible, this phenomenon can seriously influence video quality where several stages of cascaded coding are used. In the cascaded decoding and encoding chain, Figure 2, a series of up- and downsampling can contribute to significant errors in chroma components. The selection of up- and downsampling filters has to take into account this possible scenario. Proposed signalling of suitable resampling filters to the decoder can prevent accumulation of resolution conversion errors.



Figure 2: Cascaded coding. Repeated resampling may cause additional signal degradation.

***Reversible filters and coding considerations***

In order to avoid information loss caused by resampling, conversion filters have to be carefully selected. When an input signal in 4:2:2 format is coded using 4:2:0 format codec, losses are inevitable but can be minimised by selection of a suitable filter pair. Additionally, resampling filters can support other requirements, such as shifting of chroma sampling positions in 4:2:2 and 4:2:0 formats (e.g. filters defined in SMPTE RP2050 [4]).

For coding of an input signal in 4:2:2 format with a 4:4:4 format codec, conversion losses are not inevitable. To achieve a lossless format conversion a reversible integer FIR filter would be an obvious choice. However, such filters require increased precision. If conversion losses have to be prevented, it is desirable to increase the precision of upsampled samples. On the other hand, non-linear filters can be lossless without the requirement of increased precision.

**S-filter**

The evaluated non-linear filter is here called S-filter. At the encoder, chroma components are upsampled in horizontal direction. Chroma samples *y* in 4:4:4 format are obtained from chroma samples *x* in 4:2:2 format using line doubling:

, (1)

where *n* = 0, ..., *N* - 1 and 2 ⋅ *N* is the number of luma pixels in each row.

At the decoder, reconstructed chroma samples in 4:4:4 format are used to compute final samples in 4:2:2 format as:

, (2)

where ⎣⋅⎦ represents rounding to the lower integer. In the case of lossy coding, the term added to *y*(2 ⋅ *n*) in (2) reduces aliasing.

**F-filter**

Application of FIR filters can be beneficial because of possibility to achieve additional required effects, such as better anti-aliasing properties. In order to achieve reversibility with limited computational precision, considered filters taps are in format *a* / 2*k* where both *a* and *k* are integers. A possible reversible filter that satisfies this condition is listed in Table 1.

Table 1: Evaluated F-filter coefficients

|  |  |  |
| --- | --- | --- |
| **Sampling** | **Filter taps** | ***k*** |
| Up | 1, 1 | *k*u = 1 |
| Down | -1, 2, 6, 2, -1 | *k*d = 3 |

For the example from Table 1 and for upsampling from 4:2:2 to 4:4:4, bilinear interpolation is used. For downsampling from 4:4:4 to 4:2:2, a five-tap filter is used. Note that this process can be reversible if the precision of upsampled signal increases by *k*u and if the downsampling is performed in processing precision required by *k*d. In practice this means that if the input signal in 4:2:2 format is *b*-bit (e.g. 8-bit), and if reversibility is sought, the coding precision should be at least *b* + *k*u (e.g. 9-bit). This condition guarantees that the original 4:2:2 samples can be obtained at the decoder, in the case of the lossless coding.

In case that an increase of codec's internal processing precision is not feasible, rounding during downsampling can be used. In that case the perfect reconstruction of 4:2:2 signal is not possible. However, filters can be designed to minimise such effects. It should be also noted that, while focusing on reversible filters, this proposal also supports non-reversible filters, which may be beneficial for some other specific requirements.

***Implementation***

The proposed method requires all internal coding computations to be carried out in 4:2:0 or 4:4:4 chroma format. A new section has been added to the SPS to enable the delivery to the decoder of the information required to reconstruct the signal to the original, intended chroma format. This information consists of a flag specifying the use of the proposed technique, a second flag specifying what family of filters to use (two options available, non-linear and linear filters), and details of the filter implementation, such as the associated bit-shift, the length of the filter and the associated taps.

The encoder used for the experiments described in this contribution has two important modifications affecting RDO and RDOQ:

* RDO  
  The reference software uses the sum of square errors (SSE) as a measure of distortion in its RDO operations. When the distortion for the chroma components is computed in the tested modification, only pixels of interest from the signal in 4:4:4 format are used. Therefore the calculation of SSE for chroma components has been modified and takes into account samples on even horizontal positions only.
* RDOQ  
  Similarly to the RDO change, RDOQ has been modified to give the chroma components a relative weight which is closer to what would be obtained for a 4:2:2 signal going through the reference codec that operates in 4:2:2 mode. This is obtained by doubling the value of lambda used for the chroma components.

Without the two modifications to the encoder described above the proposed approach tends to generate higher quality and higher bit-rate chroma components than the reference software, making the comparison results harder to interpret.

# Experiments

In this experiment only coding of 4:2:2 signals using 4:4:4 codecs is considered. Two different filters specified in Section 2.1 are used. While F-filter can be arbitrarily selected, the one from Table 1 has been used in the experiment.

HM-range-extensions is the HEVC implementation that has been used as the basis for the experiments. In particular, revision 3055 has been used as the reference for coding the selected 4:2:2 content (six sequences, as described in [6]). A modified version of this codec has been used to generate the results reported for the proposed approach. Note that the reference operates directly on sequences that are in 4:2:2 format (also without encoder-side changes defined in Section 2.2). Both tested versions operate in 4:4:4 mode of HM-range-extensions software and include changes proposed in this document.

All test sequences are 10-bit and the processing precision used for tests with both filters is 10-bit. In order to obtain 10-bit 4:4:4 sequence from 10-bit 4:2:2 sequence in the test where the F-filter is evaluated, a rounding is used as:

.

This step obviously introduces losses. However, it can be prevented by increasing codec's internal processing bit-depth.

Table 2: Comparison between 4:2:2 coding and proposed method with non-linear filter (S-filter).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **All Intra HE Main-tier** | | | **All Intra HE High-tier** | | |
|  | Y | U | V | Y | U | V |
| YCbCr 4:2:2 | 3.1% | -1.6% | -3.9% | 5.1% | 1.0% | -1.9% |
| Enc Time[%] | 116% | | | 115% | | |
| Dec Time[%] | 126% | | | 126% | | |
|  |  |  |  |  |  |  |
|  | **Random Access HE Main-tier** | | | **Random Access HE High-tier** | | |
|  | Y | U | V | Y | U | V |
| YCbCr 4:2:2 | 2.1% | -4.4% | -5.5% | 3.9% | -2.2% | -5.9% |
| Enc Time[%] | 110% | | | 110% | | |
| Dec Time[%] | 125% | | | 122% | | |
|  |  |  |  |  |  |  |
|  | **Low delay B HE Main-tier** | | | **Low delay B HE High-tier** | | |
|  | Y | U | V | Y | U | V |
| YCbCr 4:2:2 | 0.4% | 0.9% | -0.2% | 1.4% | 3.8% | 1.3% |
| Enc Time[%] | 109% | | | 109% | | |
| Dec Time[%] | 121% | | | 119% | | |

Table 3: Comparison between 4:2:2 coding and proposed method with linear filter (F-filter).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **All Intra HE Main-tier** | | | **All Intra HE High-tier** | | |
|  | Y | U | V | Y | U | V |
| YCbCr 4:2:2 | 1.2% | 1.8% | -0.2% | 1.6% | 5.5% | 2.0% |
| Enc Time[%] | 115% | | | 114% | | |
| Dec Time[%] | 127% | | | 123% | | |
|  |  |  |  |  |  |  |
|  | **Random Access HE Main-tier** | | | **Random Access HE High-tier** | | |
|  | Y | U | V | Y | U | V |
| YCbCr 4:2:2 | 0.3% | -0.8% | -2.4% | 0.5% | 1.6% | -2.5% |
| Enc Time[%] | 109% | | | 108% | | |
| Dec Time[%] | 129% | | | 124% | | |
|  |  |  |  |  |  |  |
|  | **Low delay B HE Main-tier** | | | **Low delay B HE High-tier** | | |
|  | Y | U | V | Y | U | V |
| YCbCr 4:2:2 | -1.0% | 3.3% | 1.3% | -1.4% | 6.6% | 2.2% |
| Enc Time[%] | 108% | | | 107% | | |
| Dec Time[%] | 125% | | | 121% | | |

The results of the experiments are reported in Table 2 and Table 3 for S-filter and F-filter, respectively. Both sets of results show the comparison between the HM-range-extensions codec (revision 3055, using coding of 4:2:2 format) and the proposed approach for the six 4:2:2 sequences defined in [6]. Note that for the tested solutions the PSNR is computed on 4:2:2 signals, i.e. between the original signal before the upsampling, and after reconstruction and the downsampling.

In general the compression performance of the proposed approach is lower than the straight 4:2:2 coding with intra coding seemingly paying a slighter higher penalty. Running times are higher for the proposed approach, reflecting the fact that 50% more data needs to be processed once the signal is upsampled to 4:4:4. The increase in run times is higher for the decoder, where it is between 21% and 29%.

Decoded sequences were informally inspected for visual artefacts with concern of possibly visible aliasing artefacts for chroma components. Such problems were not observed for viewed sequences and observed artefacts were not different in nature from those observed in the reference sequences.

# Syntax changes

In Section 7.3.2.2

|  |  |
| --- | --- |
| **num\_short\_term\_ref\_pic\_sets** | ue(v) |
| for( i = 0; i < num\_short\_term\_ref\_pic\_sets; i++) |  |
| short\_term\_ref\_pic\_set( i ) |  |
| **long\_term\_ref\_pics\_present\_flag** | u(1) |
| if( long\_term\_ref\_pics\_present\_flag ) { |  |
| **num\_long\_term\_ref\_pics\_sps** | ue(v) |
| for( i = 0; i < num\_long\_term\_ref\_pics\_sps; i++ ) { |  |
| **lt\_ref\_pic\_poc\_lsb\_sps**[ i ] | u(v) |
| **used\_by\_curr\_pic\_lt\_sps\_flag**[ i ] | u(1) |
| } |  |
| } |  |
| if( chroma\_format\_idc = = 1 | | chroma\_format\_idc = = 3 ) { |  |
| **chroma\_resampling\_filter\_flag** | u(1) |
| if( chroma\_resampling\_filter\_flag ) { |  |
| if( chroma\_format\_idc = = 3 ) |  |
| **chroma\_resampling\_non\_linear\_filter\_flag** | u(1) |
| if( !chroma\_resampling\_non\_linear\_flag ) { |  |
| **chroma\_resampling\_filter\_bit\_shift** | ue(v) |
| **chroma\_resampling\_filter\_taps\_number** | ue(v) |
| for( i = 0; i < chroma\_resampling\_filter\_taps\_number; i++ ) { |  |
| **chroma\_resampling\_filter\_tap[i]** | sev(v) |
| } |  |
| } |  |
| } |  |
| } |  |
| **sps\_temporal\_mvp\_enable\_flag** | u(1) |
| **strong\_intra\_smoothing\_enable\_flag** | u(1) |
| **vui\_parameters\_present\_flag** | u(1) |
| if( vui\_parameters\_present\_flag ) |  |
| vui\_parameters( ) |  |
| **sps\_extension\_flag** | u(1) |
| if( sps\_extension\_flag ) |  |
| while( more\_rbsp\_data( ) ) |  |
| **sps\_extension\_data\_flag** | u(1) |
| rbsp\_trailing\_bits( ) |  |
| } |  |

In Section 7.4.2.2

**chroma\_resampling\_filter\_flag** equal to 1 specifies that the signal to be decoded should be resampled to 4:2:2 chroma format.

**chroma\_resampling\_non\_linear\_filter\_flag** equal to 1 specifies that the signal to be decoded should be downsampled to the 4:2:2 chroma format by using a non-liner filter. chroma\_resampling\_non\_linear\_filter\_flag equal to 0 specifies that the signal to be decoded should be resampled to the 4:2:2 chroma format by using a linear filter. If chroma\_format\_idc is equal to 1 (4:2:0) chroma\_resampling\_non\_linear\_filter\_flag is equal to 0.

**chroma\_resampling\_bit\_shift** specifies the extent of the bit down-shift required to implement the linear downsampling filter.

**chroma\_resampling\_filter\_taps\_number** specifies the number of taps of the linear resampling filter to be applied to the chroma components to obtain the 4:2:2 format.

**chroma\_resampling\_filter\_tap[i]** specifies the value of the i-th tap coefficient.

# References

1. K. Sharman, N. Saunders and J. Gamei, "AHG7: Options present in Extended Chroma Format model", JCTVC-K0181, 11th meeting, Shanghai, CN, October 2012.
2. T. Chujoh, "AHG7: Chroma sampling filter hint SEI", JCTVC-K0152, 11th meeting, Shanghai, CN, October 2012.
3. J. Sole, "HEVC Range Extensions Core Experiment 1: Square and non-square transforms for 4:2:2 chroma format coding", JCTVC-K1121, 11th meeting, Shanghai, CN, October 2012.
4. “4:2:2 / 4:2:0 Format Conversion Minimizing Color Difference Signal Degradation in Concatenated Operations – Filtering,” SMPTE RP2050-1:2012, January 2012.
5. A. Nakagawa, K. Kazui, and S. Kobayashi, “Proposal on standardization of a filter set for converting between 4:2:2/4:2:0 chrominance sampling,” ISO/IEC JCT1/SC29/WG11, MPEG2008/M15569, Hannover, Germany, July 2008.
6. D. Flynn, “Common Test Conditions and software reference configurations for HEVC range extensions”, JCTVC-K1006, 11th meeting, Shanghai, CN, October 2012.

**Patent rights declaration(s)**

**The British Broadcasting Corporation (BBC) 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).**