|  |  |
| --- | --- |
| **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-N0070 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **Inter-layer prediction modes based on base layer sharpness filter** | | |
| *Status:* | Input Document to JCT-VC | | |
| *Purpose:* | Proposal | | |
| *Author(s) or Contact(s):* | Sychev Maxim  Anisimovskiy Valery  Ikonin Sergey | Tel: Email: | Sychev.Maxim@huawei.com Anisimovskiy.Valery@huawei.com Sergey.Ikonin@huawei.com | |
| *Source:* | Huawei | | |

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

# Abstract

Proposed contribution describes method of predicting higher resolution layer images from lower resolution layer images when scalable mode is used. Described algorithm based on sharpness filter applied to upsampling low resolution frame.

The simulation results show that it achieves 1.7% and 1.0% BD rate savings on average for AI-2x and AI-1.5x, respectively, compared with anchors. The Class A test sequences show 2,9% BD rate saving. Encoding times are 113,3% and 111,2%, and decoding times are 117,7% and 116,0%.

# Introduction

Modern inter-layer prediction algorithm for spatial scalability video codec should satisfy the following requirements:

* Algorithm should minimize prediction residual signal in order to provide better coding efficiency
* Algorithm should minimize the computational complexity and memory requirements of optimal prediction parameters search
* Algorithm should lead to little increase in decoding complexity
* Algorithm should allow easy and seamless integration into the existing scalable video codec architecture and infrastructure

The contribution proposes inter-layer prediction algorithms meeting the aforementioned requirements. According to the described algorithm additional inter-layer prediction mode is introduced into the scalable video codec. This mode is based on prediction using upscaled lower resolution layer image processed by sharpening filter

The new prediction mode is used for the cases it provide lower residual signal energy than other available prediction modes (i.e. intra-prediction, inter-prediction, current standard inter-layer prediction). This allows to achieve lower bit rate for higher resolution layer bitstream while keeping the same quality as for the current scalable video codec standards (or, alternatively, better quality at the same bit rate).

# Proposed technique

The contribution proposes inter-layer prediction modes for scalable coding. Figure 1 shows the scalable video encoder flow chart and the placement of the proposed inter-layer prediction modes within the scalable codec architecture (red blocks).



Figure 1. Scalable video encoder flow chart

The proposed mode uses the upsampled reconstructed frame of the lower resolution layer encoder. For each block of the higher resolution layer image it produces the candidate predictor block by applying sharpening filter to the collocated block in the upsampled lower resolution layer image (see Figure 2). The mode choice is performed based on cost function minimization. Cost function is Cost=D + λ R, where D is the distortion created during the encoding process, R is amount of bit required for sending the residual data and sharpening filter flag, and λ is some constant.



Figure 2. Inter-layer prediction modes based on sharpness filter

At the first step, the edge map is obtained by using Prewitt filter:

For each pixel the simplified gradient vector magnitude are estimated:

At the second step the blurring filter is applied to the edge map using filter to each rows and each columns to smooth vector gradient map.

At the last step the sharpening of upscaled frame if performed by warping source pixels to the edges with bilinear interpolation. The blurred edge map is used as parameters of warping:

Sharpened frame is obtained by using bilinear interpolation with displacement vector .

The flow chart of sharpening process you can see at figure 3.



Figure 3. Sharpness filtering flow chart

The chroma sharpening is performed in the same way, which needs to perform obtaining edge map for each color component (results is represented in Table 1). But there is one more implementation of color sharpening can be used. The second method is use the downscaled version of luma edge map for each color component for sharpening. Second approach shows the significantly better coding gain for color components with little decreasing coding gain of luma (results is represented in Table 3).

## Syntax elements

To transmit the parameters of the sharpness filter to the decoder, 4 new syntax elements are added to SPS structure:

|  |  |
| --- | --- |
| **sps\_strong\_intra\_smoothing\_enable\_flag** | **u(1)** |
| **sps\_sharpening\_depth** | **ue(v)** |
| **sps\_sharpening\_depth\_chroma** | **ue(v)** |
| **sps\_sharpening\_trunc** | **ue(v)** |
| **sps\_sharpening\_p0** | **ue(v)** |
| **vui\_parameters\_present\_flag** | **u(1)** |

The non-zero value of **sps\_sharpening\_depth** signals the usage of sharpness filter.

At CU-level one bit is added as syntax element to the bitstream.

|  |  |
| --- | --- |
| **base\_mode\_flag** | **u(1)** |
| If( base\_mode\_flag ) |  |
| **sharpness\_mode\_flag** | **u(1)** |

Semantics

**sps\_sharpening\_depth** correspond to the value of *ShD* coefficient

**sps\_sharpening\_depth\_chroma** correspond to the value of *ShD* coefficient for color component sharpening

**sps\_sharpening\_trunc** correspond to the value of *Tr*

**sps\_sharpening\_p0** correspond to the value of *K*

**sharpness\_mode\_flag** equal to 1 specifies that the macroblock prediction mode equal to sharpness interlayer intra prediction, otherwise, if **sharpness\_mode\_flag** equal to 0, the standard interlayer intra prediction is used.

## RefIdx framework

The RefIdx implementation adds support for the sharpened inter-layer prediction mode using the reference picture buffer: the sharpened version of the upsampled reconstructed base layer picture is added to the reference picture buffer so that it may be used for prediction of the current frame PUs along with temporal reference pictures and the upsampled reconstructed base layer picture. No modifications of bit stream syntax at CU-level are required for RefIdx implementation.

The described coding tool has the following benefits:

* Reduction of residual signal when predicting image blocks containing distinct edges smeared by downsampling/upsampling transform thereby improving coding efficiency
* Easy and seamless integration into the existing scalable video codec architecture and infrastructure: the proposed prediction mode augments the inter-layer prediction with sharpening enhancement of upsampled lower resolution layer image without changing any of the encoder modules or encoder architecture

The results of RefIdx framework is represented in Table 2.

# Results

The proposed method has been implemented based on SHM2.0 and simulated under the common test conditions. Table 1 summarizes the experimental results. It reportedly shows 1.7% and 1.0% BD rate savings on average for AI-2x and AI-1.5x, respectively, compared with anchors. The Class A test sequences shows -2,9 of BD rate savings. Encoding times are 113,3% and 111,2%, and decoding times are 117,7% and 116,0%, respectively.

Table 1. Experimental results on AI configuration

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **AI HEVC 2x** | | | **AI HEVC 1.5x** | | |
|  | Y | U | V | Y | U | V |
| Class A | -2,9% | -2,1% | -2,3% |  |  |  |
| Class B | -1,1% | -0,6% | -0,8% | -1,0% | -0,3% | -0,7% |
| **Overall (Test vs Ref)** | -1,7% | -1,0% | -1,3% | -1,0% | -0,3% | -0,7% |
| **Overall (Test vs single layer)** | 10,4% | 12,0% | 11,4% | 9,2% | 9,7% | 8,7% |
| **Overall (Ref vs single layer)** | 12,3% | 13,2% | 12,8% | 10,3% | 10,1% | 9,5% |
| **EL only (Test vs Ref)** | -2,7% | -1,9% | -2,2% | -1,2% | -0,4% | -0,9% |
| Enc Time[%] | 113,3% | | | 111,2% | | |
| Dec Time[%] | 117,7% | | | 116,0% | | |
| BL Match | Matched | | | Matched | | |

Table 2. Experimental results for AI, RA and LDP configuration in REF\_IDX\_FRAMEWORK

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **AI HEVC 2x** | | | **AI HEVC 1.5x** | | |
|  | Y | U | V | Y | U | V |
| Class A | -2.9% | -2.1% | -2.3% |  |  |  |
| Class B | -1.1% | -0.6% | -0.9% | -1.0% | -0.3% | -0.7% |
| **Overall (Test vs Ref)** | -1.6% | -1.0% | -1.3% | -1.0% | -0.3% | -0.7% |
| **Overall (Test vs single layer)** | 10.8% | 13.5% | 12.7% | 9.4% | 9.4% | 8.5% |
| **Overall (Ref vs single layer)** | 12.6% | 14.7% | 14.2% | 10.5% | 9.8% | 9.3% |
| **EL only (Test vs Ref)** | -2.7% | -2.0% | -2.3% | -1.2% | -0.4% | -0.9% |
| Enc Time[%] | 114.8% | | | 110.7% | | |
| Dec Time[%] | 133.2% | | | 124.5% | | |
| BL Match | Matched | | | Matched | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **RA HEVC 2x** | | | **RA HEVC 1.5x** | | |
|  | Y | U | V | Y | U | V |
| Class A | -2,5% | -0,2% | -0,3% |  |  |  |
| Class B | -0,7% | 0,2% | -0,1% | -0,8% | -0,1% | -0,6% |
| **Overall (Test vs Ref)** | -1,2% | 0,1% | -0,2% | -0,8% | -0,1% | -0,6% |
| **Overall (Test vs single layer)** | 18,0% | 33,4% | 31,5% | 15,2% | 28,8% | 28,3% |
| **Overall (Ref vs single layer)** | 19,4% | 33,2% | 31,7% | 16,2% | 28,8% | 29,1% |
| **EL only (Test vs Ref)** | -1,8% | -0,4% | -0,7% | -0,8% | 0,0% | -0,6% |
| Enc Time[%] | 106,8% | | | 106,2% | | |
| Dec Time[%] | 153,4% | | | 154,5% | | |
| BL Match | Matched | | | Matched | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **LD-P HEVC 2x** | | | **LD-P HEVC 1.5x** | | |
|  | Y | U | V | Y | U | V |
| Class A | -1,8% | -0,2% | -0,5% |  |  |  |
| Class B | -0,5% | 0,1% | -0,1% | -0,6% | -0,1% | -0,5% |
| **Overall (Test vs Ref)** | -0,9% | 0,0% | -0,2% | -0,6% | -0,1% | -0,5% |
| **Overall (Test vs single layer)** | 38,8% | 48,5% | 47,3% | 33,3% | 39,5% | 39,6% |
| **Overall (Ref vs single layer)** | 40,0% | 48,4% | 47,5% | 34,1% | 39,7% | 40,3% |
| **EL only (Test vs Ref)** | -1,3% | -0,4% | -0,6% | -0,6% | 0,0% | -0,6% |
| Enc Time[%] | 105,5% | | | 104,6% | | |
| Dec Time[%] | 153,2% | | | 147,4% | | |
| BL Match | Matched | | | Matched | | |

Table 3. Experimental results on AI configuration with second approach of color sharpening

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **AI HEVC 2x** | | | **AI HEVC 1.5x** | | |
|  | Y | U | V | Y | U | V |
| Class A | -2,9% | -2,9% | -3,4% |  |  |  |
| Class B | -1,0% | -1,2% | -1,3% | -0,8% | -1,1% | -1,2% |
| **Overall (Test vs Ref)** | -1,6% | -1,7% | -1,9% | -0,8% | -1,1% | -1,2% |
| **Overall (Test vs single layer)** | 10,5% | 11,2% | 10,7% | 9,4% | 8,8% | 8,2% |
| **Overall (Ref vs single layer)** | 12,3% | 13,2% | 12,8% | 10,3% | 10,1% | 9,5% |
| **EL only (Test vs Ref)** | -2,5% | -2,7% | -2,9% | -0,7% | -1,2% | -1,3% |
| Enc Time[%] | 113,3% | | | 111,2% | | |
| BL Match | Matched | | | Matched | | |

# Patent rights declaration(s)

**Huawei Corporation 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).**