|  |  |
| --- | --- |
| **Joint Collaborative Team on Video Coding (JCT-VC)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  35th Meeting: Geneva, CH, 22–27 March 2019 | Document: JCTVC-AI0023-v2 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **Encoder-only GOP-based temporal filter** | | |
| *Status:* | Input document to JCT-VC | | |
| *Purpose:* | Proposal | | |
| *Author(s) or Contact(s):* | P. Wennersten J. Östrand R. Sjöberg | Email: | per.wennersten@ericsson.com |
| *Source:* | Ericsson | | |

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

# Abstract

This contribution proposes an encoder-only temporal filter. The filtering is done at the encoder side as a pre-processing step. Source pictures before and after the selected picture to encode are read and a block based motion compensation method relative to the selected picture is applied on those source pictures. Samples in the selected picture are temporally filtered using sample values after motion compensation.

The overall filter strength is set depending on the temporal sub layer of the selected picture as well as the QP. Only pictures at temporal sub layers 0 and 1 are filtered and pictures of layer 0 are filter by a stronger filter than pictures of layer 1. The per sample filter strength is adjusted depending on the difference between the sample value in the selected picture and the co-located samples in motion compensated pictures so that small differences between a motion compensated picture and the selected picture are filtered more strongly than larger differences.

The method was reportedly tested under HM-16.20 RA, LDB and LDP common test configurations. The average Y/U/V BD-rates are reported to be −4.5%/−6.4%/−5.9% (RA), −3.3%/−4.5%/−5.3% (LDB) and −5.0%/−6.0%/−6.3% (LDP) when allowing a 2-picture look-ahead. For no look ahead, the RD-rates for LD are reported as −0.2%/−0.9%/−1.5% (LDB) and −1.4%/−1.8%/−2.2% (LDP). All BDR numbers were computed using unfiltered source sequences.

It is proposed to adopt the proposed method into the HM software.

# Introduction

The HM encoder minimizes the rate\*lambda + distortion metric for each block.

If the next picture was identical, the coded block would likely be used un-modified for that picture as well. In this case, rate\*lambda+2\*distortion should be minimized instead. Equivalently, lambda could be changed. HM approximates this by setting different QPs at different temporal layers.

If the next picture differs only slightly, the coded block will still likely be used un-modified for that picture as well. In this case, rate\*lambda + distortion1 + distortion2 should be minimized. Rather than calculating the two distortions separately, the distortion between the coded block and the average of the original versions of the two pictures can be calculated. This is the motivation for temporal filtering of the original pictures. The filter strength drops off rapidly if the original images are different, as it is no longer likely that the blocks coded for the first picture will be re-used for the second.

# Proposal

## GOP Based Temporal Filter

A temporal filter is introduced directly after reading picture and before encoding. Below are the steps described in more detail.

Step 1: Pictures are read by the encoder

Step 2: If a picture is low enough in the coding hierarchy, it is filtered before encoding. Otherwise the picture is encoded without filtering. RA pictures with POC % 8 == 0 are filtered as well as LD pictures with POC % 4 == 0. AI pictures are never filtered.

The overall filter strength, , is set according to the equation below for RA.

where is the number of pictures read.

For the LD case, is used.

Step 3: Two pictures before and/or after the selected picture (referred to as original picture further down) are read. In the edge cases e.g. if is the first picture or close to the last picture, only the available pictures are read.

Step 4: Motion of the read pictures before and after, relative to the original picture is estimated per 8x8 picture block.

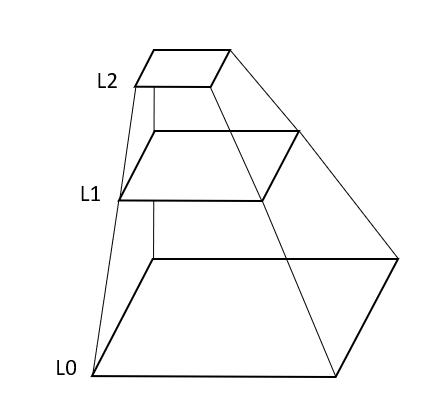


Figure 1 – The different layers of the hierarchical motion estimation. L0 is the original resolution. L1 is a subsampled version of L0. L2 is a subsampled version of L1.

A hierarchical motion estimation scheme is used and the layers L0, L1 and L2, are illustrated in Figure 1. Subsampled pictures are generated by averaging each 2x2 block for all read pictures and the original picture, i.e. L1 in Figure 1. L2 is derived from L1 using the same subsampling method.

First, motion estimation is done for each 16x16 block in L2. The squared difference is calculated for each selected motion vector and the motion vector corresponding to the smallest difference is selected. The selected motion vector is then used as initial value when estimating the motion in L1. Then the same is done for estimating motion in L0. As a final step, subpixel motion is estimated for each 8x8 block by using an interpolation filter on L0.

The VTM 6-tap interpolation filter was used:

0: 0, 0, 64, 0, 0, 0

1: 1, -3, 64, 4, -2, 0

2: 1, -6, 62, 9, -3, 1

3: 2, -8, 60, 14, -5, 1

4: 2, -9, 57, 19, -7, 2

5: 3, -10, 53, 24, -8, 2

6: 3, -11, 50, 29, -9, 2

7: 3, -11, 44, 35, -10, 3

8: 1, -7, 38, 38, -7, 1

9: 3, -10, 35, 44, -11, 3

10: 2, -9, 29, 50, -11, 3

11: 2, -8, 24, 53, -10, 3

12: 2, -7, 19, 57, -9, 2

13: 1, -5, 14, 60, -8, 2

14: 1, -3, 9, 62, -6, 1

15: 0, -2, 4, 64, -3, 1

Step 5: Motion compensation is applied on the pictures before and after the original picture according to the best matching motion for each block. I.e. so that the sample coordinates of the original picture in each block have the best matching coordinates in the referenced pictures.

Step 6: The samples of the processed one by one for the luma and chroma channels as described in the following steps.

Step 7: The new sample value, , is calculated using the following formula.

Where is the sample value of the original sample, is the intensity of the corresponding sample of motion compensated picture and is the weight of motion compensated picture when the number of available motion compensated pictures is .

In the luma channel, the weights, , is defined as follows:

Where

For all other cases of , and :

For the chroma channels, the weights, , is defined as follows:

Where and

Step 8: The filter is applied for the current sample. The resulting sample value is stored separately.

Step 9: The filtered picture is encoded.

## GOP configuration change for random access

In addition to above, a GOP configuration change is proposed since it improves the BD-rate in this context.

In the file cfg/encoder\_randomaccess\_main10.cfg, line 28 is changed to:

Frame1: B 16 -1 0.0 0.0 0 0 1.0 0 0 0 2 3 -16 -24 -32 0

I.e. the QP is decreased by 2 for the lowest layer of B-frames.

# Results

Proposed method was implemented in and compared to HM-16.20. The method is applied for RA and LD coding, and not to AI coding.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Random Access Main 10 (CTC)** | | |
|  | **Over HM-16.20** | | |
|  | Y | U | V |
| Class A1 | -3.1% | -5.4% | -4.7% |
| Class A2 | -6.0% | -8.5% | -8.4% |
| Class B | -6.8% | -8.4% | -7.5% |
| Class C | -1.4% | -2.9% | -2.8% |
| Class E |  |  |  |
| **Overall** | -4.5% | -6.3% | -5.9% |
|  | -4.7% | -6.5% | -5.9% |
| Class D | 0.6% | -2.1% | -2.8% |
| Class F | -0.7% | -0.8% | -1.0% |
| Enc Time[%] | 104% | | |
| Dec Time[%] | 100% | | |
|  |  |  |  |
|  | **Low delay B Main 10**  **2 picture look-ahead** | | |
|  | **Over HM-16.20** | | |
|  | Y | U | V |
| Class A1 |  |  |  |
| Class A2 |  |  |  |
| Class B | -4.2% | -5.1% | -6.1% |
| Class C | -1.3% | -2.3% | -3.3% |
| Class E | -4.5% | -6.3% | -6.7% |
| **Overall** | -3.3% | -4.5% | -5.3% |
|  | -3.4% | -4.5% | -5.4% |
| Class D | -0.1% | -3.0% | -3.6% |
| Class F | -0.7% | -1.1% | -1.4% |
| Enc Time[%] | 104% | | |
| Dec Time[%] | 99% | | |
|  |  |  |  |
|  | **Low delay P Main 10**  **2 picture look-ahead** | | |
|  | **Over HM-16.20** | | |
|  | Y | U | V |
| Class A1 |  |  |  |
| Class A2 |  |  |  |
| Class B | -6.6% | -7.1% | -7.8% |
| Class C | -2.1% | -2.9% | -3.4% |
| Class E | -6.2% | -8.1% | -7.8% |
| **Overall** | -5.0% | -6.0% | -6.3% |
|  | -5.2% | -5.9% | -6.3% |
| Class D | -1.0% | -3.2% | -4.2% |
| Class F | -1.0% | -1.4% | -1.1% |
| Enc Time[%] | 106% | | |
| Dec Time[%] | 96% | | |

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Low delay B Main 10 (CTC)**  **No look ahead** | | |
|  | **Over HM-16.20** | | |
|  | Y | U | V |
| Class A1 |  |  |  |
| Class A2 |  |  |  |
| Class B | -0.7% | -0.9% | -1.9% |
| Class C | 1.5% | 1.1% | 0.5% |
| Class E | -1.7% | -3.5% | -3.3% |
| **Overall** | -0.2% | -0.9% | -1.5% |
|  | -0.3% | -0.9% | -1.4% |
| Class D | 3.1% | 1.0% | 0.5% |
| Class F | 1.2% | 0.8% | 1.5% |
| Enc Time[%] | 102% | | |
| Dec Time[%] | 102% | | |
|  |  |  |  |
|  | **Low delay P Main 10 (CTC)**  **No look-ahead** | | |
|  | **Over HM-16.20** | | |
|  | Y | U | V |
| Class A1 |  |  |  |
| Class A2 |  |  |  |
| Class B | -2.5% | -2.2% | -2.9% |
| Class C | 1.0% | 0.6% | 0.4% |
| Class E | -2.9% | -4.3% | -4.4% |
| **Overall** | -1.4% | -1.8% | -2.2% |
|  | -1.7% | -1.8% | -2.1% |
| Class D | 2.4% | 0.7% | 0.4% |
| Class F | 1.0% | 1.4% | 1.1% |
| Enc Time[%] | 104% | | |
| Dec Time[%] | 102% | | |

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Random Access Main 10 (with GOP configuration change)** | | |
|  | **Over HM-16.20** | | |
|  | Y | U | V |
| Class A1 | -3.4% | -9.3% | -8.2% |
| Class A2 | -6.7% | -11.4% | -11.2% |
| Class B | -7.9% | -12.6% | -11.2% |
| Class C | -1.9% | -5.6% | -5.3% |
| Class E |  |  |  |
| **Overall** | -5.1% | -9.8% | -9.0% |
|  | -5.3% | -9.8% | -8.9% |
| Class D | 0.1% | -4.3% | -4.3% |
| Class F | -1.1% | -1.8% | -1.8% |
| Enc Time[%] | 102% | | |
| Dec Time[%] | 96% | | |

# Conclusion

In this document a temporal filter was introduced providing improved compression efficiency. We recommend adopting the proposed method into the HM software.

# References

[1]. HM-16.20 reference software, <https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.20>

# Patent rights declaration(s)

**Ericsson 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).**