|  |  |
| --- | --- |
| **Joint Collaborative Team on Video Coding (JCT-VC)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  11th Meeting: Shanghai, CN, 10–19 Oct. 2012 | Document: JCTVC-K0250r2 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **Coding interlaced video using sequence-adaptive field-frame** | | |
| *Status:* | Input Document to JCT-VC | | |
| *Purpose:* | Information | | |
| *Author(s) or Contact(s):* | Dzung Hoang and Jun Xin 3960 Freedom Circle, Suite 100 Santa Clara, CA 95054 U.S.A. | Tel: Email: | +1 408 350 5052 [dzung.hoang@zenverge.com](mailto:dzung.hoang@zenverge.com) [jun.xin@zenverge.com](mailto:jun.xin@zenverge.com) |
| *Source:* | Zenverge, Inc. | | |

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

# Abstract

This document presents results from encoding 1920x1080i interlaced video using a customized version of HM-8.0-dev using a sequence-adaptive field-frame (SAFF) scheme. With SAFF, the input video is divided into GOP-sized sequences and encoded both in field mode and frame mode. The GOP size is 64 in field mode and 32 in frame mode. In a post-processing step, the BD-rate is computed for each GOP comparing frame mode versus field mode. The mode that shows a BD-rate gain is chosen for each GOP. A bitstream-level BD-rate is then computed using the chosen modes. The HM SAFF encodings are compared to running JM 18.4 with HM-like settings in MBAFF mode. Experimental results suggest that coding efficiency gains of HEVC-SAFF over AVC-MBAFF is content-dependent, with the luma BD-rate gain ranging from 16.4% to 42.5%.

# Introduction

At the 10th meeting of the JCT-VC in Stockholm, Sweden, it was reported in [1] that the performance of HEVC for interlaced video is worse than AVC for certain video test clips when the video was encoded as fields using HM. For certain other video test clips, the performance of HEVC was reported to be worse than AVC when the video was encoded as frames. At the meeting, it was remarked that GOP-by-GOP selection of field versus frame coding is supported by HEVC, if each GOP is coded as a sequence, and that this would give better results than using a single coding mode.

In this document, we describe a sequence-adaptive field-frame (SAFF) scheme whereby the input video is divided into GOP-sized sequences and encoded in both field mode and frame mode. In a post-processing step, the BD-rate is computed for each GOP comparing frame mode versus field mode. The mode that shows a BD-rate gain is chosen for each GOP. A bitstream-level BD-rate is then computed using the chosen modes. The HM SAFF encodings are compared to running JM 18.4 with HM-like settings in MBAFF mode.

# Experimental Setup

## HEVC Sequence-Adaptive Field-Frame

HM-8.0-dev revision 2764 is used as a starting point with the following enhancements.

1. Added generation of field\_indication SEI message.
2. Automatic conversion of frame-based YUV file to fields for encoding.
3. Automatic conversion of field picture buffers to frame YUV for file output.
4. When field\_indication SEI message is enabled, compute PSNR in field mode even when coding in frame mode so that PSNR can be compared to field mode coding.
5. Added interpolation optimizations from JCTVC-I0142, which does not affect the output bitstream.

The RA-Main test condition [2] is used. Each input video clip is divided into sequences each containing a number of frames that is equivalent to the IntraPeriod specified in [2] for RA-Main. For field coding, the number of field pictures would be twice the IntraPeriod. Each video sequence is encoded once in field mode and again in frame mode. The same GOP prediction structure is used for both field and frame modes.

The BD-rate is computed for each GOP comparing frame mode versus field mode. The mode that shows a BD-rate gain is chosen for each GOP. A bitstream-level BD-rate is then computed using the chosen modes.

## AVC MBAFF

JM 18.4 is used to generate the AVC anchor bitstreams. The encoder\_JM\_RA\_B\_HE.cfg configuration file provided with JM 18.4 is used. To enable MBAFF mode, the following additional command-line options are used.

1. PicInterlace=0
2. MbInterlace=2
3. LevelIDC=41
4. BIdenticalList=0
5. ReferenceReorder=0
6. WeightedPrediction=0
7. WeightedBiprediction=0

The following set of QP values is used: “23, 28, 33, 38.” Thus QPJM = QPHM+1.

## Test Video Clips

Because the common test conditions [2] lack any interlaced video content, an external source of interlace video was used. NBC Universal (NBCU) provided us a set of 1920x1080i@60Hz video clips for testing under an NDA. The NBCU test set consist of the following sequences. The sequences were converted from 4:2:2 chroma format to 4:2:0 using TM5-based chroma-conversion filters.

Table 1: NBCU Interlaced Test Video Sequences

| **Name** | **Frames** | **Dur.(s)** | **Content** | **Additional Comments** |
| --- | --- | --- | --- | --- |
| JayLenoShow\_Race | 1316 | 43.9 | outdoor scene of car racing around a track taken with an in-car camera, a dashboard camera, and an external camera | Interlaced, noisy, high background motion, bright clipped highlights, high camera motion, pans, action |
| MacysParade\_Cheerleaders1 | 4875 | 162.5 | Macy’s Thanksgiving Day Parade live coverage of dancing cheerleaders | interlaced, high camera motion, pans, zooms, action, saturated colors, high detail |
| MacysParade\_Cheerleaders2 | 5686 | 189.5 | Macy’s Thanksgiving Day Parade live coverage of a second group of dancing cheerleaders | interlaced, high camera motion, pans, zooms, action, saturated colors, high detail |
| MacysParade\_Flags | 1984 | 66.1 | Macy’s Thanksgiving Day Parade live coverage of a marching band with flag wavers | interlaced, high camera motion, pans, zooms, action, saturated colors, high detail |
| MacysParade\_Musical | 1421 | 47.4 | Macy’s Thanksgiving Day Parade live coverage of dancing musical theatre group | interlaced, high camera motion, pans, zooms, action, saturated colors, less high detail |
| MadFashion\_Cadence | 1801 | 60.0 | teaser and opening sequence for an episode of Mad Fashion television show | film content with 3:2 pulldown cadence, graphics, fades, still |

Due to the long duration and repetitive nature of these video sequences, we selected the following clips of duration 10 seconds or less. Because all the test sequences are 30fps, the resulting GOP size is 32 for frame coding and 64 for field coding.

Table 2: Selected Test Clips

| **Name** | **Frames** | **Dur.(s)** | **Source** | **Content** |
| --- | --- | --- | --- | --- |
| Cheer1 | 300 | 10 | MacysParade\_Cheerleaders1 frames 2781 to 3080 | group square dancing with lateral camera movement |
| Cheer2 | 299 | 9.97 | MacysParade\_Cheerleaders1 frames 3873 to 4171 | group square dancing with lateral camera movement |
| Cheer3 | 300 | 10 | MacysParade\_Cheerleaders2 frames 1480 to 1779 | cheerleaders with camera movement from front to overhead to front |
| Cheer4 | 300 | 10 | MacysParade\_Cheerleaders2 frames 3348 to 3647 | cheerleaders starts with no camera movement then zoom out and fades into news commentators |
| Mad1 | 300 | 10 | MadFashion\_Cadence frames 580 to 879 | teaser that ends with title |
| Music1 | 208 | 6.93 | MacysParade\_Musical frames 914 to 1121 | musical dance group with little camera motion |

# Experimental Results

## SAFF Field-Frame Selection

For each GOP coded with HM, BD-rate was computed comparing frame coding with field coding. The best mode (in BD-rate sense) is tabulated below for each GOP. GOPs with frame mode are highlighted. Detailed results can be found in the accompanying spreadsheet.

Table 3: Best Field-Frame Mode per GOP

| **GOP** | **Cheer1** | **Cheer2** | **Cheer3** | **Cheer4** | **Mad1** | **Music1** |
| --- | --- | --- | --- | --- | --- | --- |
| 0 | field | field | field | field | field | frame |
| 1 | field | field | field | field | field | frame |
| 2 | field | field | field | field | field | frame |
| 3 | field | field | field | frame | field | frame |
| 4 | field | field | field | frame | field | frame |
| 5 | field | field | field | field | field | frame |
| 6 | field | field | field | field | field | field |
| 7 | field | field | field | field | frame |  |
| 8 | field | field | field | field | frame |  |
| 9 | field | field | field | field | frame |  |

## Bitstream-Level BD-Rate

After the field-frame mode has been selected at the sequence (GOP) level, the composite rate and PSNR numbers for SAFF mode are computed for the entire bitstream by using weighted averaging where the weights are the length of each sequence (GOP).

The tables below show BD-rate comparison of HEVC Frame versus HEVC Field, HEVC Field versus HEVC SAFF, and HEVC Frame versus HEVC SAFF.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **HEVC Frame Coding vs HEVC Field Coding** | | |
|  | **BD-rate (piecewise cubic)** | | |
| **Clip** | **Y** | **U** | **V** |
| Cheer1 | 76.8% | 62.7% | 74.6% |
| Cheer2 | 98.7% | 115.9% | 117.3% |
| Cheer3 | 88.5% | 98.2% | 94.4% |
| Cheer4 | 14.2% | -5.1% | -5.9% |
| Mad1 | 48.9% | 33.9% | 15.6% |
| Music1 | -12.7% | -49.1% | -51.0% |
| Average | 52.4% | 42.8% | 40.8% |

|  |  |  |  |
| --- | --- | --- | --- |
|  | **HEVC Field Coding vs HEVC Adaptive** | | |
|  | **BD-rate (piecewise cubic)** | | |
| **Clip** | **Y** | **U** | **V** |
| Cheer1 | 0.0% | 0.0% | 0.0% |
| Cheer2 | 0.0% | 0.0% | 0.0% |
| Cheer3 | 0.0% | 0.0% | 0.0% |
| Cheer4 | 0.1% | 5.7% | 6.0% |
| Mad1 | 4.6% | 8.0% | 20.6% |
| Music1 | 15.7% | 92.2% | 99.8% |
| Average | 3.4% | 17.6% | 21.0% |

|  |  |  |  |
| --- | --- | --- | --- |
|  | **HEVC Frame Coding vs HEVC Adaptive** | | |
|  | **BD-rate (piecewise cubic)** | | |
| **Clip** | **Y** | **U** | **V** |
| Cheer1 | 76.8% | 62.7% | 74.6% |
| Cheer2 | 98.7% | 115.9% | 117.3% |
| Cheer3 | 88.5% | 98.2% | 94.4% |
| Cheer4 | 14.2% | 0.2% | -0.5% |
| Mad1 | 54.6% | 43.2% | 37.6% |
| Music1 | 1.0% | -2.4% | -2.4% |
| Average | 55.6% | 53.0% | 53.5% |

|  |  |  |  |
| --- | --- | --- | --- |
|  | **HEVC Field Coding vs JM 18.4 MBAFF** | | |
|  | **BD-rate (piecewise cubic)** | | |
| **Clip** | **Y** | **U** | **V** |
| Cheer1 | -40.3% | -45.8% | -43.2% |
| Cheer2 | -41.6% | -47.4% | -43.2% |
| Cheer3 | -42.5% | -41.8% | -41.8% |
| Cheer4 | -25.3% | -17.5% | -20.2% |
| Mad1 | -20.6% | -31.4% | -25.5% |
| Music1 | -5.2% | -2.1% | 9.4% |
| Average | -29.3% | -31.0% | -27.4% |

|  |  |  |  |
| --- | --- | --- | --- |
|  | **HEVC Frame Coding vs JM 18.4 MBAFF** | | |
|  | **BD-rate (piecewise cubic)** | | |
| **Clip** | **Y** | **U** | **V** |
| Cheers1 | 5.5% | -11.0% | -0.3% |
| Cheers2 | 15.9% | 14.2% | 24.0% |
| Cheers3 | 8.9% | 16.5% | 14.3% |
| Cheers4 | -14.1% | -20.6% | -23.7% |
| Mad1 | 20.5% | -3.0% | -7.7% |
| Music1 | -15.5% | -47.9% | -43.4% |
| Average | 3.5% | -8.6% | -6.1% |

|  |  |  |  |
| --- | --- | --- | --- |
|  | **HEVC Adaptive vs JM 18.4 MBAFF** | | |
|  | **BD-rate (piecewise cubic)** | | |
| **Clip** | **Y** | **U** | **V** |
| Cheers1 | -40.3% | -45.8% | -43.2% |
| Cheers2 | -41.6% | -47.4% | -43.2% |
| Cheers3 | -42.5% | -41.8% | -41.8% |
| Cheers4 | -25.3% | -21.6% | -24.3% |
| Mad1 | -23.2% | -34.7% | -35.4% |
| Music1 | -16.4% | -46.9% | -42.3% |
| Average | -31.5% | -39.7% | -38.4% |

## Analysis of Results

We observe that the experimental results are content-dependent. Cheer1, Cheer2, and Cheer3 have high camera motion and action and favour field coding. Music1 has little camera motion and favours frame coding. Cheer4 has little camera motion and favours field for GOPs where the action is high and frame where the action is lower. Even though Mad1 was originally shot in film, a 3:2 cadence was later added. This cadence caused the encoding to favour field coding when there is motion or transitions. When there is little or no motion, the cadence is not visible and thus frame coding is favoured.

## Additional Open-GOP Results

The HEVC Field and HEVC Frame results presented above are for the case where each IntraPeriod number of pictures is a separate coded video sequence, and can thus be characterized as “closed-GOP.” The reason for doing this initially is that the closed-GOP results can be combined in a spreadsheet to derive the SAFF results. Certainly, we expect that open-GOP field-based and frame-based encoding would yield better coding results.

We ran additional experiments to determine the coding efficiency impact of closed-GOP versus open-GOP for both field-based and frame-based encodings. The results in the table below show that on average, closed-GOP encodings use 4.2% more bits for fields and 5.2% for frames.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Closed GOP HEVC Field vs Open GOP HEVC Field** | | | | **Closed GOP HEVC Frame vs Open GOP HEVC Frame** | | | |
| **BD-rate (piecewise cubic)** | | | | **BD-rate (piecewise cubic)** | | | |
| **Clip** | **Y** | **U** | **V** | **Clip** | **Y** | **U** | **V** |
| Cheer1 | 2.9% | 5.3% | 4.6% | Cheer1 | 3.2% | 5.9% | 4.4% |
| Cheer2 | 2.5% | 4.7% | 4.1% | Cheer2 | 2.7% | 4.6% | 3.7% |
| Cheer3 | 2.1% | 2.9% | 2.8% | Cheer3 | 2.1% | 2.8% | 2.8% |
| Cheer4 | 3.1% | 4.4% | 4.8% | Cheer4 | 3.5% | 5.2% | 5.6% |
| Mad1 | 7.3% | 10.5% | 9.8% | Mad1 | 10.2% | 13.7% | 14.5% |
| Music1 | 7.2% | 11.6% | 10.8% | Music1 | 9.6% | 17.6% | 17.4% |
| Average | 4.2% | 6.6% | 6.2% | Average | 5.2% | 8.3% | 8.1% |

More detailed results can be found in the accompanying spreadsheet.

# Summary and Recommendations

In summary, we implemented SAFF on top of HM and found that the coding efficiency gains of HEVC-SAFF over AVC-MBAFF is content-dependent. The luma BD-rate gain varies from 16.4% to 42.5%. This wide range suggests that there is room for improving the coding efficiency of HEVC for interlaced content with additional coding tools that target interlaced specifically.

One area that we did not investigate was GOP structure optimizations for field coding. Preliminary results presented in [3] show gains for field coding that closes the gap for certain interlaced test sequences that are better coded as frames. It would be instructive to repeat the SAFF experiments with the optimized field-based GOP structure. Therefore, we recommend that this topic be further studied in future JCT-VC meetings.

# References

1. A. Luthra and D. Baylon, “Performance of HEVC for Interlaced Video,” JCT-VC Document JCTVC-J0466, Stockholm, SE, July 2012.
2. F. Bossen, “Common test conditions and software reference configurations,” JCT-VC Document JCTVC-J1100, Stockholm, SE, July 2012.
3. C. Auyeung, “GOP structures for Interlaced Video,” JCT-VC Document JCTVC-K0153, Shanghai, CN, October 2012.

# Patent rights declaration(s)

**Zenverge, Inc. 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).**