**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 4**

**MPEG VIDEO CODING**

**ISO/IEC JTC 1/SC 29/WG 4 m65098**

**October 2023, Hannover**

**Title: [VCM][CE4-Related] The simplification of temporal up-sampling method**

**Source: Huei-Jiun Yang, Kuan-Ru Huang, Sheng-Po Wang, Ching-Chieh Lin, Chun-Lung Lin (ITRI)**

**Abstract**

This document presents a modification related to the temporal up-sampling mothed to simplify the interpolation procedure by skipping SSIM calculation and subsequent condition selection. This proposed method shows no substantial performance change in the average BD-rate under the All-Intra, Random Access, and Low Delay test conditions.

# Description of the proposed method

The temporal up-sampling tool [1] was adopted for VCM-RS and included it in the Common Test Conditions (CTC) [2] during the 142nd MPEG meeting. The major goal of the temporal up-sampling is to achieve smooth transitions. In the adopted temporal up-sampling tool in CE4, prior to the temporal up-sampling, SSIM between frames I0 and I1 is utilized to determine which method is selected to generate the intermediate frames between frames I0 and I1. If SSIM is larger than 0.996, frame I0 and the following frame I2 are utilized to estimate a new frame I1‘ using IFNET. Then the new frame generated I1‘ is used to replace the original frame I1. In contrast, if SSIM is less than 0.2, the intermediate frames are generated by duplicating frame I0. If SSIM falls within the range of 0.2 and 0.996, IFNET interpolation is then carried out to estimate the intermediate frames between frames I0 and I1.

SSIM is a metric used to compare the similarity between two decoded frames. However, the SSIM evaluation may be affected by static background regions when the total area of the regions of interest is relatively small. This document presents to simplify the temporal up-sampling procedure by removing the additional candidate methods determined by SSIM evaluation and directly use IFNET.

# Experimental Results

The performance of the proposed method is conducted on VCM-RS v0.6 according to the Common Test Condition (CTC) document [3]. SFU-HW and TVD dataset are used as the test sequences for the AI, RA, and LD modes.

Table 1 and 2 show the experimental results of the AI mode for the video object detection and tracking tasks, respectively. In comparison to the anchors, the BD-rate for SFU-HW (-0.30%) and TVD datasets (0.15%) without SSIM evaluation show no substantial performance loss.

Table 1. Summarized results for video object detection task in the AI mode

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Object**  **Detection** | **End-to-End BD-Rate [%]** | | | **End-to-End BD-mAP** |
| **mAP** | **mAP (low4)** | **Pareto mAP** |
| **Class A** | ############ | 0.00% | 0.00% | 0.00 |
| **Class B** | ############ | ############ | 0.00% | 0.00 |
| **Class C** | -0.97% | -0.69% | -0.97% | 0.12 |
| **Class D** | 0.00% | 0.00% | 0.00% | 0.00 |
| **Average** | #NUM! | ############ | **-0.30%** | 0.04 |
| **Class A+B+C+D** | -0.27% | -0.19% | -0.27% | 0.03 |

Table 2. Summarized results for video object tracking task in the AI mode

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Object Tracking** | **End-to-End BD-Rate [%]** | | | **End-to-End BD-MOTA** |
| **MOTA** | **MOTA (low4)** | **Pareto MOTA** |
| **TVD-01-1** | -2.49% | -1.80% | -2.49% | 0.23 |
| **TVD-01-2** | 3.26% | 1.67% | 3.26% | -0.19 |
| **TVD-01-3** | -0.89% | 0.24% | -0.89% | -0.02 |
| **TVD-02-1** | 1.14% | 0.00% | 1.14% | -0.12 |
| **TVD-03-1** | 0.00% | 0.00% | 0.00% | 0.00 |
| **TVD-03-2** | 0.00% | 0.00% | 0.00% | 0.00 |
| **TVD-03-3** | 0.00% | 0.00% | 0.00% | 0.00 |
| **Average** | 0.15% | 0.02% | **0.15%** | -0.01 |

Table 3 and 4 show the experimental results of the RA mode for the video object detection and tracking tasks, respectively. The BD-rate for SFU-HW (0.00%) and TVD datasets (-0.73%) without SSIM evaluation show no substantial performance loss when compared to the anchors.

**Table 3. Summarized results for video object detection task in the RA mode**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Object**  **Detection** | **End-to-End BD-Rate [%]** | | | **End-to-End BD-mAP** |
| **mAP** | **mAP (low4)** | **Pareto mAP** |
| **Class A** | ############ | 0.00% | 0.00% | 0.00 |
| **Class B** | #NUM! | #NUM! | 0.00% | 0.00 |
| **Class C** | 0.01% | 0.00% | 0.01% | 0.00 |
| **Class D** | ############ | ############ | 0.00% | 0.00 |
| **Average** | #NUM! | #NUM! | **0.00%** | 0.00 |
| **Class A+B+C+D** | 0.00% | 0.00% | 0.00% | 0.00 |

Table 4. Summarized results for video object tracking task in the RA mode

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Object Tracking** | **End-to-End BD-Rate [%]** | | | **End-to-End BD-MOTA** |
| **MOTA** | **MOTA (low4)** | **Pareto MOTA** |
| **TVD-01-1** | -2.39% | -1.36% | -2.39% | 0.34 |
| **TVD-01-2** | ############ | 0.03% | 0.08% | -0.07 |
| **TVD-01-3** | -1.55% | -2.51% | -1.55% | 0.15 |
| **TVD-02-1** | ############ | 4.09% | -1.61% | -0.23 |
| **TVD-03-1** | 0.00% | 0.00% | 0.00% | 0.00 |
| **TVD-03-2** | 0.00% | 0.00% | 0.00% | 0.00 |
| **TVD-03-3** | 0.35% | 0.32% | 0.35% | -0.03 |
| **Average** | #NUM! | 0.08% | **-0.73%** | 0.02 |

Table 5 and 6 show the experimental results of the LD mode for the video object detection and tracking tasks, respectively. When compared to the anchors, the BD-rate for SFU-HW (0.00%) and TVD datasets (0.12%) without SSIM evaluation show no substantial performance loss.

**Table 5. Summarized results for video object detection task in the LD mode**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Object**  **Detection** | **End-to-End BD-Rate [%]** | | | **End-to-End BD-mAP** |
| **mAP** | **mAP (low4)** | **Pareto mAP** |
| **Class A** | 0.00% | 0.00% | 0.00% | 0.00 |
| **Class B** | ############ | 0.00% | 0.00% | 0.00 |
| **Class C** | 0.00% | 0.00% | 0.00% | 0.00 |
| **Class D** | 0.00% | 0.00% | 0.00% | 0.00 |
| **Average** | ############ | 0.00% | **0.00%** | 0.00 |
| **Class A+B+C+D** | 0.00% | 0.00% | 0.00% | 0.00 |

Table 6. Summarized results for video object tracking task in the LD mode

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Object Tracking** | **End-to-End BD-Rate [%]** | | | **End-to-End BD-MOTA** |
| **MOTA** | **MOTA (low4)** | **Pareto MOTA** |
| **TVD-01-1** | ############ | -1.48% | -1.45% | 0.24 |
| **TVD-01-2** | ############ | 0.14% | -0.20% | 0.03 |
| **TVD-01-3** | 0.07% | 0.56% | 0.07% | 0.02 |
| **TVD-02-1** | 2.26% | 2.89% | 2.26% | -0.20 |
| **TVD-03-1** | 0.00% | 0.00% | 0.00% | 0.00 |
| **TVD-03-2** | 0.00% | 0.00% | 0.00% | 0.00 |
| **TVD-03-3** | 0.16% | -0.08% | 0.16% | -0.01 |
| **Average** | #NUM! | 0.29% | **0.12%** | 0.01 |

# Conclusion

This document reports the experimental results of simplifying the interpolation procedure by skipping SSIM calculation and subsequent condition selection. The results show that the simplification of temporal up-sampling results in no significant performance loss from the perspective of average BD-rate under AI, RA and LD test conditions. However, the interpolation algorithm still need to be enhanced in order to eliminate losses happened in the results with some specific test sequences. We recommend further investigation of interpolation algorithm to achieve better performance.

# References

[1] m63086, “[VCM] CE4.2 - Temporal Resample Coding from China Telecom”, ISO/IEC JTC 1/SC 29/WG 4, Apr. 2023.

[2] N332, “Common test conditions for video coding for machines”, ISO/IEC JTC 1/SC 29/WG 04, May 2023.

[3] N379, “Common test conditions for video coding for machines”, ISO/IEC JTC 1/SC 29/WG 04, Aug. 2023.

# Patent rights declaration(s)

ITRI may have patent claims relating to the technology described in this contribution and, conditioned on reciprocity, is prepared to grant licenses under reasonable and non-discriminatory terms for such patent claims required to implement 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).