**** **ISO/IEC JTC 1/SC 29/WG 2 N00247**

**ISO/IEC JTC 1/SC 29/WG 2**

**MPEG Technical requirements   
Convenorship: SFS (Finland)**

**Document type:** Output Document

**Title: CfE response report for Video Coding for Machines**

**Status:** Approved

**Date of document:** 2022-10-28

**Source:** ISO/IEC JTC 1/SC 29/WG 2

**Expected action:** none

**Action due date:** none

**No. of pages:** 10 (Including the cover page)

**Email of Convenor:** igor.curcio@nokia.com

**Committee URL:** [**https://sd.iso.org/documents/ui/#!/browse/iso/iso-iec-jtc-1/iso-iec-jtc-1-sc-29/iso-iec-jtc-1-sc-29-wg-2**](https://sd.iso.org/documents/ui/#!/browse/iso/iso-iec-jtc-1/iso-iec-jtc-1-sc-29/iso-iec-jtc-1-sc-29-wg-2)

**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 2**

**MPEG TECHNICAL REQUIREMENTS**

**ISO/IEC JTC 1/SC 29/WG 2 N00247**

**Online – October 2022**

**Title: CfE response report for Video Coding for Machines**

**Source: WG 2 MPEG Technical requirements**

**Status: Approved**

**Serial number: 22070**

**Editor(s): Yuan Zhang (China Telecom)**

**Chris Rosewarne (Canon)**

**Shan Liu (Tencent)**

**Christopher Hollmann (Ericsson)**

Table of Content

[1. Abstract 2](#_Toc117803577)

[2. Introduction 3](#_Toc117803578)

[3. Results 3](#_Toc117803579)

[4. Conclusion 9](#_Toc117803580)

[5. References 9](#_Toc117803581)

# Abstract

This document presents the VCM Track 1 Call for Evidence evaluation report, providing a summary-level report of objective results received for the CfE responses. This document reports overall BD-rate performance and rate distortion curves across different proposals and tasks. Detailed results are shown in Section 2. A set of RD curves for the task and dataset combination for each proposal is provided. 8 CfE responses were received, 6 responses are valid according to the completeness. Base on the evaluation, all responses outperformed the feature anchor. For the video task, 4 out of 6 responses outperformed the video anchor, for the image task, all responses outperformed both the feature and image anchor. One contribution uses same codec for both CfE for feature compression and CfP for video compression, and outperforms both feature and image anchor, although video result is not included. It is asserted that the evidence shown is sufficient to start planning a standardization activity and it is recommended that WG2 to start drafting a Call for Proposal.

# Introduction

VCM issued a Call for Evidence [1] in the 139th MPEG meeting in July. Before the document submission deadline, VCM has received 8 responses (m60742 [2], m60752 [3], m60761 [4], m60788 [5], m60799 [6], m60802/m60803 [7][8], m60821 [9], m60925 [10]) submitted by experts from different organizations. 6 responses are valid (m60761 [4], m60788 [5], m60799 [6], m60802/m60803 [7][8], m60821[9], m60925[10]). This document reports overall BD-rate performance and rate distortion curves across different proposals and tasks. Detailed results are shown in Section 2. Contribution m60742 [2] uses same codec for both CfE for feature compression and CfP for video compression, and outperforms both feature and image anchor, although video result is not included.

# Results

This section provides results comparison of all received proposals. Note that the metric computation failed for some proposals because of the non-monotonic issue.

Table 1 to 4 show BD-rate performance of all proposals on object tracking task. According to the overall performance, all received proposals outperformed the feature anchor, and 4 proposals outperformed the video anchor. The largest gain achieved is -87.44% and -97.58% comparing to video anchor and feature anchor respectively. Figure 1 to 4 shows results of object tracking task on TVD-OVERALL, TVD-01, TVD-02, and TVD-03, respectively.

|  |  |  |
| --- | --- | --- |
| Overall | BD-rate over Video | BD-rate over Feature |
| m60761 | -87.44% | -97.58% |
| m60788 | 63.69% | -74.43% |
| m60799 | -80.18% | -97.09% |
| m60803 | 218.93% | -33.01% |
| m60821 | -77.40% | -95.84% |
| m60925 | -64.94% | -92.17% |

Table 1. Proposal summary results for object tracking on TVD - OVERALL.

Chart, diagram

Description automatically generated

Figure 1. RD curve for object tracking on TVD-OVERALL.

|  |  |  |
| --- | --- | --- |
| TVD-01 | BD-rate over Video | BD-rate over Feature |
| m60761 | -71.75% | -97.05% |
| m60788 | 261.97% | -64.72% |
| m60799 | -71.07% | -96.93% |
| m60803 | 447.59% | -33.92% |
| m60821 | -51.66% | -94.64% |
| m60925 | -40.90% | -92.61% |

Table 2. Proposal summary results for object tracking on TVD - 01.

Chart

Description automatically generated

Figure 2. RD curve for object tracking on TVD-01.

|  |  |  |
| --- | --- | --- |
| TVD-02 | BD-rate over Video | BD-Rate over Feature |
| m60761 | -23.27% | -97.28% |
| m60788 | 619.65% | -79.93% |
| m60799 | -19.20% | -96.58% |
| m60803 | 2000.40% | -11.23% |
| m60821 | 56.22% | -94.23% |
| m60925 |  |  |

Table 3. Proposal summary results for object tracking on TVD - 02.

Diagram

Description automatically generated

Figure 3. RD curve for object tracking on TVD-02.

|  |  |  |
| --- | --- | --- |
| TVD-03 | BD-rate over Video | BD-rate over Feature |
| m60761 | -94.03% | -97.90% |
| m60788 | -49.41% | -84.12% |
| m60799 | -92.05% | -97.55% |
| m60803 | 93.98% | -39.65% |
| m60821 | -91.59% | -97.18% |
| m60925 |  |  |

Table 4. Proposal summary results for object tracking on TVD - 03.

Chart

Description automatically generated

Figure 4. RD curve for object tracking on TVD-03.

Table 5 shows BD-rate performance of all proposals on instance segmentation task. All received proposals outperformed the image anchor. The largest gain achieved is -93.04% and -98.60% comparing to image anchor and feature anchor respectively. Figure 5 shows detailed rate distortion curve for instance segmentation task on OpenImages dataset.

|  |  |  |
| --- | --- | --- |
|  | BD-Rate over Image | BD-Rate over Feature |
| m60761 | -79.21% | -95.56% |
| m60788 | -47.46% | -89.48% |
| m60799 | -93.04% | -98.60% |
| m60802 | -19.35 | -83.38 |
| m60821 | -78.11% | -95.84 |
| m60925 | -69.08% | -92.30% |

Table 5. Proposal summary results for Instance segmentation.

Chart, diagram

Description automatically generated

Figure 5. RD curve for instance segmentation on OpenImages.

Table 6 shows BD-rate performance of all proposals on object detection task. Note that since this is an optional task, not all responses submit their results on this task. All 4 proposals outperformed the image anchor. The largest gain achieved is -94.46% and -98.34% comparing to image anchor and feature anchor respectively. Figure 6 shows detailed rate distortion curve for object detection task on OpenImages dataset.

|  |  |  |
| --- | --- | --- |
|  | BD-Rate over Image | BD-Rate over Feature |
| 60761 | -81.11% | -94.15% |
| 60788 | -54.51% | -85.06% |
| 60799 | -94.46% | -98.34% |
| 60821 | -70.39% | -91.14% |

Table 6. Proposal summary results for Object Detection.

Chart, diagram

Description automatically generated Figure 6. RD curve for object detection on OpenImages.

# Conclusion

The 6 valid responses outperformed the feature anchor. For the video task, 4 out of 6 responses outperformed the video anchor, for the image task, all responses outperformed both the feature and image anchor. Based on the evaluation results of the CfE responses, it is asserted that the evidence shown is sufficient to start planning a standardization activity and it is recommended that WG2 to start drafting a Call for Proposal.

# References

1. “Call for Evidence on Video Coding for Machines,” ISO/IEC JTC 1/SC 29/WG 2, N00215, Jul. 2022.

[2] Ke Jia, Lu Yu, Hualong Yu, Ao Liu, Jiawang(ZJU), “Response to the CfE on Video Coding for Machine from Zhejiang University”, m60742, online, October 2022

[3] Wang Lichuan, Zhang Yuan (China Telecom), “[VCM] dct based cfe response”, m60752, online, October 2022

[4] Heeji Han, Minseok Choi, Haechul Choi (HNU), Soon-heung Jung, Jin Young Lee, Sangwoon Kwak, Hyon-Gon Choo, Won-Sik Cheong, Jeongil Seo (ETRI), “[VCM] Response from Hanbat National University and ETRI to CfE on Video Coding for Machines”, m60761, online, October 2022

[5] Minhun Lee, Seungjin Park, Seoung-Jun Oh, Donggyu Sim (Kwangwoon University), Younhee Kim, Jooyoung Lee, Se Yoon Jeong (ETRI), “[VCM Track 1] Response to CfE: A transformation-based feature map compression method”, m60788, online, October 2022

[6] Yong-Uk Yoon, Dong-Ha Kim, Jae-Gon Kim (Korea Aerospace University), Jooyoung Lee, Seyoon Jeong, Yeonhee Kim (ETRI), “[VCM] Response to VCM CfE: Multi-scale feature compression with QP-adaptive feature channel truncation”, m60799, online, October 2022

[7] Hanming Wang, Zijun Wu, Tao Han, Yuan Zhang (China Telecom), “[VCM][Response to CfE] An End-to-End Image Feature Compressing Method with Feature Fusion Module”, m60802, online, October 2022

[8] Hanming Wang, Zijun Wu, Tao Han, Yuan Zhang (China Telecom), “[VCM][Response to CfE] An End-to-End Video Feature Compressing Method with Feature Fusion Modules”, m60803, online, October 2022

[9] C. Rosewarne, R. Nguyen, “[VCM Track 1] Response to CfE on Video Coding for Machine from Canon”, m60821, online, October 2022

[10] Yong Zhang, Zizheng Liu, Han Zhu, Xiaozhong Xu, Shan Liu (Tencent), Junxi Zhang, Jianghao Jia, Zhenzhong Chen (Wuhan Univ.), “[VCM] Response to VCM Call for Evidence from Tencent and Wuhan University - a Learning-based Feature Compression Framework”, m60925, online, October 2022