**** **ISO/IEC JTC 1/SC 29/WG 2 N42**

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

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

**Document type:** Output Document

**Title:** Call for Evidence for Video Coding for Machines

**Status:** Approved

**Date of document:** 2021-01-15

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

**Expected action:** none

**Action due date:**  none

**No. of pages:** 12 (Including the cover page)

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

**Committee URL:** <https://isotc.iso.org/livelink/livelink/open/jtc1sc29wg2>

**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 N42**

**Online – January 2021**

**Title: Call for Evidence for Video Coding for Machines**

**Source: WG 2 MPEG Technical requirements**

**Status: Approved**

**Serial number: 20126**

**Editor(s): M. Rafie (GTI),**

**Yuan Zhang (China Telecom),**

**Shan Liu (Tencent)**

**Table of Contents**

[**Expected action:** none 1](#_Toc61525426)

[1 Introduction 4](#_Toc61525427)

[2 Timeline 5](#_Toc61525428)

[3 System Overview 6](#_Toc61525429)

[4 Test Conditions 6](#_Toc61525430)

[5 Evaluation Methods and Procedures 7](#_Toc61525431)

[5.1 Anchors and Datasets Definition and Requirements 7](#_Toc61525432)

[5.2 Machine-vision tasks 8](#_Toc61525433)

[5.2.1 Single-task test cases 8](#_Toc61525434)

[5.2.2 Multi-task test cases 8](#_Toc61525435)

[5.2.3 Machine-vision performance metrics 8](#_Toc61525436)

[5.3 Hybrid machine/human-vision tasks 9](#_Toc61525437)

[5.3.1 Hybrid machine/human-vision performance metrics 9](#_Toc61525438)

[6 Submission Requirements 9](#_Toc61525439)

[7 Participation fee 10](#_Toc61525440)

[8 Logistics 10](#_Toc61525441)

[9 References 12](#_Toc61525442)

# Introduction

The MPEG Technical Requirements Working Group (WG 2) is studying the standardization of coding technologies of unprocessed or processed video for machine intelligence use cases. Those technologies are expected to be with a compression capability that exceeds that of state-of-the-art video coding standard to fulfill a single or multiple machine intelligent tasks, or optionally to support hybrid machine/human vision at sufficient quality.

To better coordinate this study, WG 2 created a Video Coding for Machines (VCM) Ad-hoc Group (AhG) to investigate the use cases and requirements [1], test conditions, evaluation methodologies [2], and potential coding technologies. Two specific technologies require evidence:

* Efficient compression of processed or unprocessed video
* The shared backbone of feature extraction

These technologies can be used in (a) a single machine task, (b) multiple machine tasks, or (c) hybrid machine/human-vision tasks. The VCM AhG is considering to issue multiple rounds of Call for Evidence (CfE) to cover the full scope of VCM as defined in the draft requirements and the use cases document [1].

The first round of CfE focuses mainly on the “compression efficiency”.

This CfE requests information regarding video compression technology that has compression performance beyond that of Versatile Video Coding (VVC) for machine-only consumption as well as hybrid machine/human consumption. Evaluation of the submissions in response to the CfE will be performed at the 134th MPEG, as further described below. Depending on the result of the evaluation, a formal Call for Proposals (CfP) is likely to be issued in preparation for starting a formal standardization project.

Companies and organizations that have developed compression technologies (for machine-vision or machine/human-vision consumptions) with compression capability better than that of the anchors provided in the MPEG-VCM evaluation framework document [2] are kindly invited to bring such information to WG 2 in response to this CfE.

# Timeline

|  |  |
| --- | --- |
| 2021-01-15 | Release of Call for Evidence document |
| 2021-01-31 | Availability of neural networks, test data, and description for the respective use cases. Pointers for the datasets for training and evaluation are described in [1] and [2].  The URL link for having access to the above material: https://mpegfs.int-evry.fr/mpegcontent/ws-mpegcontent/Explorations/VCM/ |
| 2021-03-21 | Registration deadline for intend to response |
| 2021-04-12 | Deadline for electronic submission of binaries, bitstream results |
| 2021-04-19 | Deadline for submission of descriptions (MPEG input contribution) of approaches and evaluation results (for both the evaluation process of feature compression and a shared backbone) |
| 2021-04-22/30 | Evaluation of responses - the CfE will be evaluated at the 134th MPEG |

# System Overview

The generic system architecture that contains a pair of VCM encoder and decoder is shown in Fig 1.

|  |
| --- |
| Fig1. Pipeline for VCM |
| Fig 2. An example of potential VCM architecture |

Fig 2 shows an example of a potential VCM architecture. The VCM codec could be a video codec, a feature codec, or both. Regarding the detailed examples of VCM pipelines, refer to the use cases and draft requirements for Video Coding for Machines [1].

# Test Conditions

A set of relevant use cases, tasks, related requirements, as well as reference test cases (anchors) are described in [1] and [2] for machine-vision, and hybrid machine/human-vision consumptions. This document contains information on how to provide evidence for these test cases.

MPEG VCM has identified the following test cases:

* Machine-vision tasks
  1. Single-task test cases
  2. Multi-task test cases
* Hybrid machine/human-vision tasks
  1. Multi-task test cases

Submitters are encouraged (but not required) to submit their results for all test cases. However, submitters are required to provide their results for at least one dataset for the given test case.

# Evaluation Methods and Procedures

## Anchors and Datasets Definition and Requirements

VVC/H.266 codec with software version VTM-8.2 is used as the reference (anchor) for the performance evaluation of the MPEG-VCM codec. Decoded video/feature shall be tested for one or more key tasks per specific test cases and compare the performance results against the provided anchors. At the current stage of this CfE, the following key tasks are identified for the MPEG-VCM performance evaluation:

* Still Images
  + Object Detection
  + Object (Instance) Segmentation
* Video
  + Object Tracking
  + Action Recognition
  + Pose Estimation

Table 1 shows the VCM tasks along with their metrics, datasets, benchmarks, training, and test condition testing description.

Table 1 Test conditions, key tasks, metrics, datasets, benchmarks for various tasks

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Task** | **Metrics** | **Datasets** | **Benchmarks** | **Training/Testing** |
| Object Detection | [mAP](https://mc.ai/the-confusing-metrics-of-ap-and-map-for-object-detection/)  vs  BPP/Rate  PSNR | Open Images  Compressed  (Image) | <https://storage.googleapis.com/openimages/web/index.html> | Version 6. |
| FLIR Thermal Dataset  [Compressed]  (Image) | <https://www.flir.com/oem/adas/adas-dataset-form/> | Ability to sense thermal infrared radiation or heat |
| Object (Instance) Segmentation | [mAP](https://mc.ai/the-confusing-metrics-of-ap-and-map-for-object-detection/)  vs  BPP/Rate | Open Images  Compressed  (Image) | https://storage.googleapis.com/openimages/web/index.html | Version 6. |
| Object Tracking | MOTA/PSNR  vs  Rate/BPP | HiEve-10  [Uncompressed]  (video) | <http://humaninevents.org/> | SJTU - Human in Event  HiEve-10 has 7 training videos and 3 testing videos |
| Action Recognition | Frame-mAP/PSNR  Vs  Rate/BPP | HiEve-10  [Uncompressed]  (video) | <http://humaninevents.org/> | SJTU - Human in Event  HiEve-10 has 7 training videos and 3 testing videos |
| Pose Estimation | AP/PSNR  vs  Rate/BPP | HiEve-10  [Uncompressed]  (video) | <http://humaninevents.org/> | SJTU - Human in Event  HiEve-10 has 7 training videos and 3 testing videos |
| Hybrid Machine / Human Vision | BD-PerVal  (mAP/MOTA, SSIM, PSNR vs. BPP/Rate) | Any combination of listed Datasets  (video) | See above | See above |

## Machine-vision tasks

The anchors for the test cases are provided in the MPEG-VCM evaluation framework document [2]. Both a single task and multi-task are considered for machine-only test cases.

### Single-task test cases

For machine-vision single-task test cases, the datasets and tasks listed in Table 1 are encouraged to be used. Submitters are required to provide at least one machine task result(s). Submissions of test cases for machine-vision single task

* shall use the same compression technologies for all tasks listed in the test case,
* may generate different bitstreams for different tasks,
* shall indicate if any parameters of compression are adjusted per task or per dataset.

### Multi-task test cases

For machine-vision multi-task test cases, any combination of datasets and tasks listed in Table 1 is encouraged to be used. The provision of test results for machine-vision multi-task test cases is desirable (but not mandatory) for this CfE. Submissions of test cases for machine-vision multi-task shall

* use the same compression technologies and generate one bitstream for multiple tasks listed in the test case,
* indicate if any parameters of compression are adjusted per data

### Machine-vision performance metrics

For machine-vision test cases, submitters shall report BD-rate (mAP/MOTA vs. BPP/bitrate). mAP/MOTA is used to measure the codec performance, while bits per pixel (BPP) is used to measure the cost associated with the storage/transmission of the generated VCM bitstream. The performance range shall cover the provided anchors with no less than 4 test points in the range. The performance curves shall be compared to the anchors using the BD-PerVal (BD-Rate, BD-mAP, or BD-MOTA) functions as specified in the reporting MPEG-VCM template document [3]. For multi-task test cases, the results of each test case should be provided separately. In addition to specifying the test performance metrics for each task distinctively; optionally, the submitter can provide a joint performance metric for machine multi-task test cases. The joint performance metric for multiple tasks can be specified as a function of one or more task metrics. For example, a joint performance metric can be defined as a function of a single task (priority-based) metric or a weighted-average of multiple task metrics.

## Hybrid machine/human-vision tasks

For hybrid machine/human-vision test cases, any combination of datasets and tasks listed in Table 1 is encouraged to be used. The provision of test results for hybrid machine/human-vision test cases is desirable (but not mandatory) for this CfE. Submissions of test cases for hybrid machine/human consumption

* shall use the same compression technologies and generate one bitstream for multiple tasks listed in the machine test case as well as human vision,
* may have additional bitstream for human vision use cases,
* indicate if any parameters of compression are adjusted per dataset.

### Hybrid machine/human-vision performance metrics

For hybrid machine/human-vision test cases, submitters shall report BD-PerVal (mAP/MOTA, SSIM, PSNR vs. BPP/bitrate). BD-PerVal should be calculated in the way as other standardization groups, e.g. JVET. The performance range shall cover the provided anchors with no less than 4 test points in the range. The performance curves shall be compared to the anchor performance curves using the BD-rate functions as specified in the MPEG-VCM BD-PerVal reporting template [3]. For hybrid machine/human multi-task test cases, the results of each (machine or human) test case should be provided separately. In addition to specifying the test performance metrics for each task distinctively; optionally, the submitter can provide a joint performance metric for machine/humane multi-task test cases. The joint performance metric for multiple tasks can be specified as a function of one or more task metrics. For example, a joint performance metric can be defined as a function of a single task (priority-based) or a weighted-average of multiple task metrics.

# Submission Requirements

The submitter shall submit an input contribution to the 134th MPEG meeting with documentation of the compression technology. The following material is to be submitted electronically no later than 2021-04-12. The material shall also be presented to the 134th MPEG meeting for discussion.

The submission shall contain but is not limited to:

* compressed video bitstreams of at least one test case,
* executable decoders and task-specific model to fulfil the tasks, if applicable,
* filled-in reporting template [3] with compression efficiency (BD-rate with mAP/MOTA), HW/SW configurations, runtime complexity, and memory consumption measurements,
* comparison of compression efficiency versus to the given anchor,
* a description of the pipeline and the compression approach.

In case of any modification of the network, the submitter shall provide the description of their modified network, the reference datasets if this network is retrained.

For the case of one bitstream supporting multiple tasks, additional information that the submission may contain is:

* preferably a description of the shared backbone of feature extraction,
* indication of whether joint training occurred and on which dataset the trained feature extractor weights were provided.

For the hybrid machine/human vision case, additional information that the submission may contain is:

* filled-in reporting template [3] with compression efficiency (BD-rate with PSNR/MS-SSIM), runtime complexity, and memory consumption measurements.

# Participation fee

Participation in the call will not be associated with any fee.

# Logistics

Prospective contributors of responses to the Call for Evidence should contact the following people:

Igor Curcio (WG2 Convenor)

Nokia Technologies

Tel. +x xxxxxxxxxx, email [igor.curcio@nokia.com](mailto:igor.curcio@nokia.com)

M. Rafie

Gyrfalcon Technology Inc (GTI)

Tel. +1 4082059400, email [m.rafie@gyrfalcontech.com](mailto:m.rafie@gyrfalcontech.com)

Yuan Zhang

China Telecom

Tel. +86 18918588990, email [zhangy666@chinatelecom.cn](mailto:zhangy666@chinatelecom.cn)

Shan Liu

Tencent

Tel. +1 4082427048, email [shanl@tencent.com](mailto:shanl@tencent.com)

* Expressions of interest to submit a response shall be made by contacting the people above on or before 2021-03-21. Interested parties are kindly invited to express their intent as early as possible.
* Details on how to format and submit documents, bitstreams, and other required data will be communicated directly to those who express interest in participation.

Details for access to the test data and tools for evaluation can be found in [2], for further questions contact one of the above individuals.

# References

|  |  |
| --- | --- |
| [1] | m56227, "Use cases and draft requirements for Video Coding for Machines," Online, Jan 2021. |
| [2] | m56228, "Evaluation Framework for Video Coding for Machines," Online, Jan 2021. |
| [3] | C. Hollmann, (Ericsson), S. Liu, W. Gao, X. Xu and (Tencent), "m56252\_[VCM] On VCM Reporting Template," Jan 2021. |
| [4] | K. Andersson, F. Bossen, J.-R. Ohm, A. Segall, R. Sjöberg and J. S. a. G. J. Sullivan, "Summary information on BD-rate experiment evaluation practices. JVET-Q2016 (w19168)," Jan 2020. |