 ISO/IEC JTC 1/SC 29/WG 2 N248

**ISO/IEC JTC 1/SC 29/WG 2  
MPEG Technical requirements   
Convenorship: SFS (Finland)**

**Document type:** Output Document

**Title:** CfP 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:** 6 (with 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>

**INTERNATIONAL ORGANIZATION FOR STANDARDIZATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

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

**MPEG TECHNICAL REQUIREMENTS**

**ISO/IEC JTC 1/SC 29/WG 2 N248**

**Online – October 2022**

|  |  |
| --- | --- |
| **Title** | **CfP response report for Video Coding for Machines** |
| **Source** | **WG 2, MPEG Technical requirements** |
| **Status** | **Approved** |
| **Serial Number** | **22071** |

**Table of Contents**

[Abstract 2](#_Toc117875899)

[1. Introduction 3](#_Toc117875900)

[2. Status of the proposals 3](#_Toc117875901)

[3. Results 4](#_Toc117875902)

[4. Acknowledgement of responding organizations 4](#_Toc117875903)

[5. Conclusion 5](#_Toc117875904)

[6. References 5](#_Toc117875905)

[Appendix A: Questionnaire 5](#_Toc117875906)

# Abstract

This document presents the results of the VCM Track 2 Call for Proposals, providing a summary of the objective results obtained from the proposals and status of the received proposals. Based on the submitted results, it is recommended to proceed with a standardization project in the WG4.

# Introduction

The CfP on Video Coding for Machines [1], targeted at addressing requirements as detailed in [2], resulted in 17 proposals, which were studied at the 9th WG2 meeting. The proposals are categorized based on fundamental codec type as shown in Table 1.

Table 1. Proposal categorization.

|  |  |
| --- | --- |
| Number of proposals using end2end codecs based on neural networks only | 2 |
| Number of proposals using a video codec like VVC and added to it without changing the inner workings of the video codec | 13 |
| Number of proposals that extended a conventional video codec like VVC and required modification of the standardized video codec | 1 |
| Architectures not involving existing standardized video codec nor end2end codec | 1 |
| Total received proposals | 17 |

The use of region-of-interest methods in the encoder was seen in 10 proposals, as detailed below:

|  |  |
| --- | --- |
| RoI used in encoder side | 10 |
| * With video decoder output direct to task network | 4 |
| * Encoder affecting background (and adding side information to bitstream) | 3 |
| * Encoder altering QP only | 1 |

# Status of the proposals

The status of received proposals with respect to requirements (excluding feature-compression related, which is not the scope of this CfP) as per the CfP requirements questionnaire (see Appendix A) and completeness is recorded in Table 2.

Table 2. Status of received proposals.

|  |  |  |  |
| --- | --- | --- | --- |
| **Proposal** | **Valid decoder submitted** | **Valid bitstreams submitted** | **VCM requirements checklist submission** |
| P01 | Y | Y | Y |
| P02 | Y | Y | Y |
| P03 | Y | Y | N |
| P04 | Y | Y | N |
| P05 | Y | Y | N |
| P06 | Y | Y | Y |
| P07 | Y | Y | Y |
| P08 | Y | Y | Y |
| P11 | Y | Y (partial) | Y |
| P12 | Y | Y | Y |
| P14 | Y | Y (partial) | Y |
| P15 | Y | Y | Y |
| P16 | Y | Y | N |
| P17 | Y | Y | Y |
| P18 | Y | Y | Y |
| P20 | Y | Y | Y |
| P21 | Y | Y | Y |

# Results

Three tasks were used in the CfP evaluation, with the achieved performance as shown in Table 3 and a breakdown to the dataset level shown in Table 4.

Table 3. Task performance results (highest proposal).

|  |  |
| --- | --- |
| **Task** | **BD-rate change** |
| Object tracking | -57% |
| Instance segmentation | -45% |
| Object detection | -39% |

Table 4. Task performance results (highest proposal by dataset).

|  |  |  |
| --- | --- | --- |
| **Task** | **Dataset** | **BD-rate change** |
| Object tracking | TVD (videos) | -57% |
| Instance segmentation | OpenImages | -51% |
|  | TVD | -57% |
| Object detection | OpenImages | -47% |
|  | FLIR | -53% |
|  | TVD | -65% |
|  | SFU (videos) | -36% |

# Acknowledgement of responding organizations

The following organizations are thanked for responding to this CfP:

* Alibaba
* Institute of Computing Technology, Chinese Academy of Sciences (CAS-ICT)
* China Telecom
* City University of Hong Kong
* Ericsson
* Electronics and Telecommunications Research Institute (ETRI)
* Florida Atlantic University (FAU)
* Konkuk University
* Myongji University
* Nokia
* OP Solutions
* Poznan University of Technology (PUT)
* Tencent
* V-Nova
* Wuhan University
* Zhejiang University

# Conclusion

Based on assessment of evaluation of the proposals it is concluded that the submissions amount to technology that is capable of meeting the requirements as specified in the CfP. Based on these results, it is recommended to proceed with a standardization project in the WG4.

# References

1. “Call for Proposals on Video Coding for Machines,” ISO/IEC JTC 1/SC 29/WG 2, N00220, Jul. 2022.
2. “Use cases and requirements for Video Coding for Machines”, ISO/IEC JTC 1/SC 29/WG 2 N00190, Apr. 2022.

# Appendix A: Questionnaire

Proponents are requested to copy the following questionnaire to their submission containing the description of the proposed technology and fill it out according to the following instructions:

The following requirements for a VCM standard are defined in the Use cases and requirements document [2]. Please use the check boxes in the “Fulfillment” column to indicate which requirements your proposal fulfills and add a short reasoning in the “Reasoning” column as to why your proposal fulfills the requirement.

|  |  |  |
| --- | --- | --- |
| Requirement | Fulfillment | Reasoning |
| a) VCM shall support video coding for machine task consumption purposes. |  |  |
| b) VCM shall support feature coding. |  |  |
| c) VCM shall support a coding efficiency improvement for at least 30% BD-rate over the VVC standard on machine vision tasks. |  |  |
| d) VCM shall support a broad spectrum of encoding rates. |  |  |
| e) VCM shall support various degrees of delay configuration. |  |  |
| f) VCM shall be agnostic to network models. |  |  |
| g) VCM shall be agnostic to machine task types. |  |  |
| h) VCM shall provide description of the meaning or the recommended way of using the decoded data. |  |  |
| i) VCM should support the use and inclusion of information such as descriptors in its bitstream. |  |  |
| j) A single VCM bitstream shall support any number of instances of machine tasks. |  |  |
| k) VCM shall support at least the following colour formats; monochrome, RGB, and YUV (YCbCr). |  |  |
| l) VCM shall support at least the following input bit depths: 8-bit and 10-bit. |  |  |
| m) VCM complexity shall allow for feasible implementation within the constraints of the available technology at the expected time of usage. |  |  |
| n) VCM shall support rectangular picture format up to 7680x4320 pixels (8K). |  |  |
| o) VCM shall support fixed and variable rational frame rates for video inputs. |  |  |
| p) VCM shall support any input source from video or image. |  |  |
| q) VCM shall support privacy and security. |  |  |