Text

Description automatically generated ISO/IEC JTC 1/SC 29/WG 2 N107

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

**Document type:** Output Document

**Title:** **Common Test Conditions and Evaluation Methodology for Video Coding for Machines**

**Status:** Approved

**Date of document:** 2021-07-16

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

# Expected action: None

# Action due date: None

**No. of pages:** 4 (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 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 N107**

**Online – July 2021**

**Title: Common Test Conditions and Evaluation Methodology for Video Coding for Machines**

**Source: WG2 MPEG Technical Requirements**

**Status: Approved**

**Serial number: 20709**

# Abstract

This document summarizes the discussions and decisions made during MPEG 135th meeting with regards to the common test conditions, evaluation methodology and reporting template for Video Coding for Machines [1].

# Introduction

During the VCM AHG and BOG meeting discussions before and through the MPEG 135th meeting, the VCM group decided to adopt some new anchors including:

* New anchor results for instance segmentation using new input dataset TVD [2];
* New anchor results for object tracking using new input dataset TVD [3];
* New anchor results for object detection using OpenImageV6 dataset [4];
* New anchor results for instance segmentation using OpenImageV6 dataset [5];
* New anchor results for video object detection using SFU-HW-Objects-v1 dataset [6][7];
* Updated anchor results for object detection using TVD dataset [8];
* Updated anchor results for object detection using FLIR dataset [9].

It was also decided that newly proposed anchors during and after this meeting should be generated using VTM-12.0, whereas the previous anchors were generated using VTM-8.2.

# Common Test Conditions

In total, five datasets are used for testing six machine tasks. Specifically, three datasets are used for testing object tracking, two are used for instance segmentation, two are used for object tracking, one is used for pose estimation, one is used for action recognition, and one is used for video object detection. They are summarized in Table 1.

Table 1. Machine tasks and corresponding evaluation datasets

|  |  |
| --- | --- |
| Machine Task | Evaluation Dataset |
| Object detection | OpenImageV6  FLIR (IR)  TVD (image) |
| Instance segmentation | OpenImageV6  TVD (image) |
| Object tracking | HiEve-10  TVD (video) |
| Pose Estimation | HiEve-10 |
| Action Recognition | HiEve-10 |
| Video Object Detection | SFU-HW-Objects-v1 |

# Evaluation Methodology

BD-rate-mAP, BD-rate-fmAP and BD-rate-MOTA are used to evaluate the proposed solution against the anchor. Specifically, BD-rate-MOTA is used for object tracking, BD-rate-fmAP is used for action recognition, and BD-rate-mAP is used for all other machine tasks listed in Table 1. Here the rate is measured in BPP, i.e., bits per pixel. R-D curves are also used to demonstrate the R-D performance of the proposed solution compared with the anchor.

# Reporting Template

The reporting template in [1] is updated by adding object detection results for newly adopted TVD image dataset [2] to object tracking, and SFU-HW-Objects-v1 video dataset [3] to object tracking. Note that the “Video\_Object\_Detection” task is different to image-based object detection and its anchor results are per video sequence results as opposed to per image results. The results for object detection and object segmentation using OpenImageV6 are updated by the results provided in [4]. Some error uncompressed mAP values for the FLIR dataset are fixed with the corrected values in [5]. In addition, the existing FLIR anchor results are replaced with the updated results using 10-bit process in [6]. The RD curves and Pareto front points for the two newly added datasets have been added to the reporting template. The summary page is modified as well to reflect the changes above.

# Conclusions

We recommend the VCM group to use the updated test conditions and reporting template in and attached to this document for future activities.

# References

1. S. Liu, S M. Rafie, Y. Zhang, “Common Test Conditions and Evaluation Methodology for Video Coding for Machines”, ISO/IEC JTCre 1/SC 29/WG2 output document, N0080, April 2021, Online.
2. W. Gao, X. Xu, S. Liu, M. Qin, “[VCM] TVD dataset for object tracking”, MPEG-m57470, Online, July 2021.
3. W. Gao, X. Xu, S. Liu, M. Qin, “[VCM] TVD dataset for object segmentation”, MPEG-m57471, Online, July 2021.
4. H. Zhang, N. Le, R. Ghaznavi-Youvalari, F. Cricri, H. R. Tavakoli, E. Aksu, M. Hannuksela, T. Ji, K. Misra, P. Cowan, A. Segall, “[VCM] VCM Anchors on Open Images Dataset," MPEG-m57343, Online, July 2020
5. C. Hollmann, “[VCM] Anchor results for OpenImages with VTM-12.0”, MPEG-m57626, Online, July 2020.
6. K. Misra, T. Ji, P. Cowan, A. Segall, “[VCM] Proposed object detection anchors for SFU-HW-Objects-v1 video dataset”, MPEG-m56868, Online, April 2021.
7. C. Hollmann, Jacob Ström, Lukasz Litwic, “[VCM] Supplemental anchor results for SFH-HW”, MPEG-m57353, July 2021
8. W. Gao, X. Xu, S. Liu, “[VCM] Updated Anchor results for object detection using TVD dataset”, MPEG-m57617, Online, July 2021.
9. Y. Lee, S. Kim, K. Yoon, H. Lim, H. Choo, W. Cheong, J. Seo, “Updated FLIR Anchor results for object detection”, MPEG-m57375, Online, April 2021.