



Technical Report

ISO/IEC TR 23888-3

Information technology — Artificial intelligence for multimedia —

Part 3:

Optimization of encoders and receiving systems for machine analysis of coded video content

*Technologies de l'information — Intelligence artificielle pour le
multimédia —*

*Partie 3: Optimisation des codeurs et des systèmes de réception
pour l'analyse automatique de contenus vidéo codés*

**First edition
2026-06**

Sample Document

get full document from standards.iteh.ai



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2026

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

	Page
Foreword	iv
1 Scope	1
2 Normative references	1
3 Terms and definitions	1
4 Abbreviated terms	2
5 Overview	2
5.1 General overview.....	2
5.2 Use cases and applications.....	3
6 Evaluation methodology	3
6.1 General.....	3
6.2 Bit rate.....	4
6.3 PSNR.....	4
6.4 mAP.....	4
6.5 MOTA.....	5
6.6 BD-rate.....	5
7 Pre-processing technologies	6
7.1 Region of interest-based methods.....	6
7.2 Foreground and background processing.....	7
7.3 Temporal subsampling.....	7
7.4 Spatial subsampling.....	8
7.5 Noise filtering.....	8
8 Encoding technologies	8
8.1 RoI-based quantization parameter adaption.....	8
8.2 Quantization step adjustment for temporal layers.....	9
8.3 Chroma QP offset setting.....	10
9 Post-processing technologies	10
9.1 Temporal resampling.....	10
9.2 Spatial resampling.....	10
9.3 Enhancement post-filtering.....	10
10 Metadata	11
10.1 General.....	11
10.2 Neural-network post-filter SEI message.....	11
10.3 Annotated regions SEI message.....	11
10.4 Object mask information SEI message.....	11
10.5 Encoder optimization information SEI message.....	12
10.6 Packed regions information SEI message.....	12
Annex A (informative) Software implementation examples	13
Annex B (informative) Combined software implementation examples	20
Bibliography	21

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives or www.iec.ch/members_experts/refdocs).

ISO and IEC draw attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO and IEC take no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO and IEC had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents and <https://patents.iec.ch>. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html. In the IEC, see www.iec.ch/understanding-standards.

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*, in collaboration with ITU-T as ITU-T H.Sup.MACVC).

A list of all parts in the ISO/IEC 23888 series can be found on the ISO and IEC websites.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html and www.iec.ch/national-committees.

Information technology — Artificial intelligence for multimedia —

Part 3:

Optimization of encoders and receiving systems for machine analysis of coded video content

1 Scope

This document provides information about optimizations for encoders and receiving systems for conducting machine analysis tasks on coded video content. It provides a concept-level overview of recent practices and provides comments on technical aspects and cautions to be taken when interpreting the results. This document describes technologies that have recently been studied and have demonstrated benefits to coding efficiency for some machine analysis tasks.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitute requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

Rec. ITU-T H.266 | ISO/IEC 23090-3, *Versatile video coding*

Rec. ITU-T H.265 | ISO/IEC 23008-2, *High efficiency video coding*

Rec. ITU-T H.264 | ISO/IEC 14496-10, *Advanced video coding*

Rec. ITU-T H.274 | ISO/IEC 23002-7, *Versatile supplemental enhancement information messages for coded video bitstreams*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in Rec. ITU-T H.266 | ISO/IEC 23090-3, Rec. ITU-T H.265 | ISO/IEC 23008-2, Rec. ITU-T H.264 | ISO/IEC 14496-10, Rec. ITU-T H.274 | ISO/IEC 23002-7 and the following apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

- ISO Online browsing platform: available at <https://www.iso.org/obp>
- IEC Electropedia: available at <https://www.electropedia.org/>

3.1

machine consumption

operation of a machine analysis task such as object detection, segmentation or object tracking

4 Abbreviated terms

AVC	Advanced Video Coding (Rec. ITU-T H.264 ISO/IEC 14496-10)
BD-rate	Bjontegaard delta bit rate
CTU	coding tree unit
HEVC	High Efficiency Video Coding (Rec. ITU-T H.265 ISO/IEC 23008-2)
IoU	intersection over union
mAP	mean average precision
MOTA	multiple object tracking accuracy
NNPF	neural-network post-filter
NNPFA	neural-network post-filter activation
NNPFC	neural-network post-filter characteristics
OMI	object mask information
PRI	packed regions information
PSNR	peak signal-to-noise ratio
QP	quantization parameter
RoI	region of interest
RPR	reference picture resampling
SEI	supplemental enhancement information
TID	temporal identifier
URI	uniform resource identifier
VSEI	Versatile Supplemental Enhancement Information Messages for Coded Video Bitstreams (Rec. ITU-T H.274 ISO/IEC 23002-7)
VTM	Reference software for versatile video coding (Rec. ITU-T H.266.2 ISO/IEC 23090-16:2025)
VVC	Versatile Video Coding (Rec. ITU-T H.266 ISO/IEC 23090-3)
$Y' C_B C_R$	colour space representation commonly used for video/image distribution, also written as YUV
YUV	colour space representation commonly used for video/image distribution, also written as $Y' C_B C_R$

5 Overview

5.1 General overview

Most video processing systems consist of four main processing steps, as shown in [Figure 1](#). This document describes technologies for optimization of encoders and receiving systems, such as pre-processing, encoding and post-processing for machine consumption. The decoding process, on the other hand, is fully specified in the respective Rec. ITU-T H.266 | ISO/IEC 23090-3 Versatile Video Coding (VVC), Rec. ITU-T H.265 | ISO/IEC 23008-2 High Efficiency Video Coding (HEVC) and Rec. ITU-T H.264 | ISO/IEC 14496-10 Advanced

Video Coding (AVC) video coding standards, amongst others. Hence, the samples of the decoded video are fully specified by the given input bitstream.

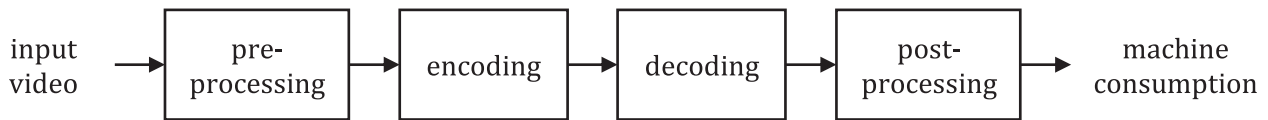


Figure 1 — General video coding and processing pipeline

An overview of the commonly used practices for evaluating encoder optimization technologies for machine consumption can be found in [Clause 6](#). Descriptions of pre-processing technologies can be found in [Clause 7](#). Encoder optimization technologies are described in [Clause 8](#) and post-processing technologies are described in [Clause 9](#). Metadata that is useful for machine consumption is described in [Clause 10](#).

It is noted that depending on specific use cases, the technologies outlined in this document can be implemented individually or in combination to optimize the machine consumption performance within the constraints of the system capabilities. When employing multiple technologies simultaneously, it is important to consider that certain combinations can be impractical or infeasible due to inherent methodological constraints. Tested technologies and their combinations are listed in [Annex A](#) and [Annex B](#), respectively.

5.2 Use cases and applications

There are various use cases and applications using encoded video that benefit from optimizing both encoders and receiving systems for machine consumption. Some of them are highlighted below:

- Surveillance: A considerable amount of bandwidth is needed to transmit a high volume of data generated by a large number of sensors. The number of sensors also has an impact on the computational load on the server side, as having to analyse the input from many sensors can become a huge burden. This can be eased by distributing the computation to the front-end devices.
- Intelligent transportation: A key aspect for vehicular applications is interoperability between not only vehicles from different vendors, but also the infrastructures of various locations. Connected vehicles are expected to play a significant role in future transport systems and the tremendous number of vehicles emphasizes the need of reducing the amount of data being transmitted between them to avoid overloading the network.
- Intelligent industry: One example in this area is visual content analysis, checking and screening. Machine automation is desirable for increasing efficiency.

A more detailed description of use cases can be found in ISO/IEC TR 23888-1^[1].

6 Evaluation methodology

6.1 General

A set of assessment metrics are used for the evaluation of encoder and receiving systems optimization technologies for machine consumption. An overview evaluation framework is shown in [Figure 2](#). Here the input video is encoded to generate a bitstream. This bitstream is then decoded, and the decoded video is used for machine consumption. In this diagram, the “encoder” includes both pre-processing and encoding steps, and the “decoder” includes both decoding and post-processing steps, as shown in [Figure 1](#).