

**Joint Video Experts Team (JVET)
of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC
29/WG 11**

Document: JVET-AE0095

MPEG: m63801

31st Meeting, Geneva, CH, 11-19 July 2023

Title: AHG8/AHG9: Proposed changes to the candidate new object mask information SEI message

Status: Input document to JVET

Purpose: Proposal

Author(s) or Contact(s): [Philippe de Lagrange](#) | philippe.delagrange@interdigital.com (InterDigital)
[Edouard François](#) | edouard.francois@interdigital.com | submitter (InterDigital)
[D. Doyen](#) | didier.doyen@interdigital.com (InterDigital)
[Jie Chen](#) | jiechen.cj@alibaba-inc.com (Alibaba)
[Shurun Wang](#) | shurun.wsr@alibaba-inc.com (Alibaba)
[Y. Ye](#) | yan.ye@alibaba-inc.com (Alibaba)

Source: InterDigital, Alibaba

HEADER EXCERPT FROM SOURCE DOCUMENT

Abstract

This contribution proposes an update to the candidate new object mask information (OMI) SEI message proposed in contribution JVET-AD0175. The OMI SEI message is used in conjunction with auxiliary pictures that code object masks. Objects can be detected from the auxiliary pictures thanks to their ID, signaled in the SEI message. One potential issue is when lossy coding is used for those pictures, in which case the sample values inside the mask auxiliary pictures are not guaranteed to perfectly match the object IDs. The proposal aims at addressing this issue, by adding in the SEI a parameter specifying a tolerance range around the object ID values. In addition, bounding box parameters that delimit the masks' location are also proposed to be included.

Motivation of OMI SEI message

The Annotated region (AR) SEI message adopted in VSEI, HEVC and AVC is used for the object detection and tracking applications, which are a common workload for end-to-end video systems. With the AR SEI message, the object detection and tracking task only needs to be performed in encoder-side and the results are sent to the decoder through the SEI message so that the decoder doesn't need to perform the video analysis again. It has many advantages compared with performing those video tasks in decoder-side. First, the power consumption of the decoder can be reduced, and second the object detecting and tracking results will be more accurate as usually the encoder is more computationally powerful than the decoder, and the encoder can perform the video analysis on the original video instead of the compressed one.

LINKS

The screenshot displays the JvET Search web application. At the top, the 'JvET Search' logo is on the left, and a search bar with the text 'AHG8/AHG9: Proposed changes to the candidate new object mask information SEI message' is on the right. Below the search bar, a list of authors is shown: Philippe de Lagnan, R. Edouard François, D. Doyen, Jie Chen, Shunan Wang, and Y. Ye. The main content area features a section titled 'ABSTRACT' with a blue horizontal bar, followed by a 'MEETING NOTES' section with three blue horizontal bars. A QR code is located in the bottom right corner of the interface.

The screenshot displays the Jvett Document Management System interface. At the top, a dark blue header contains the text "Jvett Document Management System" and "Jvett Video Experts Team". Below the header, a light blue sidebar on the left lists navigation options: "User", "Login", "Create new user", "Claim your password", and "Contact". The main content area has a white background and features a red header "Preview document JVET-AE0095". Below this, a red section titled "Document Information" contains details: "Submitted by" (Jens-Rainer Oehm), "Title" (H265/H266 Proposed changes to the candidate new object mask information), "Author" (V. Seeliger), and "Document" (JVET-AE0095 upload document). To the right of the "Document Information" section, the text "Associated Resources" is visible. At the bottom of the main content area, a red section titled "Comments" is shown, but it is empty. A QR code is located in the bottom right corner of the interface.

Archivio

PDF JVET-AE0095-v1.pdf

DOC JVET-AE0095_v2.docx

```
graph TD; Archivio --> PDF["PDF JVET-AE0095-v1.pdf"]; Archivio --> DOC["DOC JVET-AE0095_v2.docx"];
```

	Joint Video Experts Team (JVET) of ITU-T SG21 WP3/21 and ISO/IEC JTC 1/SC 29 31st Meeting: Geneva, CH, July 2023	Document: JVET-AE0095
Title:	Meeting Report of the 41st Meeting of the Joint Video Experts Team	
Status:	Report document from the chair of JVET	
Purpose:	Report	

1 Summary



MEETING NOTES

[JVET-AE0095](#) AHG8/AHG9: Proposed changes to the candidate new object mask information SEI message [P. de Lagrange, E. François, D. Doyen (InterDigital), J. Chen, S. Wang, Y. Ye (Alibaba)] [late]

AHG10: Encoding algorithm optimization (4)

Contributions in this area were discussed at 1240–1300 and 1430–1530 on Sunday 16 July 2023 (chaired by JRO).

[JVET-AE0095](#) AHG8/AHG9: proposed changes to the candidate new object mask information SEI message [P. de Lagrange, E. François, D. Doyen (InterDigital), J. Chen, S. Wang, Y. Ye (Alibaba)] [late]

This contribution proposes an update to the candidate new object mask information (OMI) SEI message proposed in contribution JVET-AD0175. The OMI SEI message is used in conjunction with auxiliary pictures that code object masks. Objects can be detected from the auxiliary pictures thanks to their ID, signalled in the SEI message. One potential issue is when lossy coding is used for those pictures, in which case the sample values inside the mask auxiliary pictures are not guaranteed to perfectly match the object IDs. The proposal aims at addressing this issue, by adding in the SEI a parameter specifying a tolerance range around the object ID values. In addition, bounding box parameters that delimit the masks' location are also proposed to be included.

The purpose of using an alpha map to reconstruct binary object shapes (rather than “grey-level shapes”) via an auxiliary picture is well recognized and important. An open question is whether the method of defining a threshold in case of lossy coding of the map is more efficient (and practical from the encoder view point) than right away doing lossless coding (e.g. of a picture that just contains 0/1 values). For complicated shapes, it might be difficult to determine at which QP lossy coding needs to be performed, to guarantee that no points of the object are lost, and no points outside of the object are interpreted as object points.

To some extent, perfect lossless coding may not be necessary, and post processing of the lossy decoded shape could also be performed.

Further study was recommended, bringing examples/results with real-world examples rather than hand-crafted examples.

This would have potential applications in machine consumption tasks, but even more for general purposes of segmented video, such as video conferencing.

It was pointed out that V-PCC has some similar concepts of coding the occupancy map, using either lossy or lossless compression. Comparison would be beneficial.

Decision: Adopt JVET-AE0095 into TuC for VSEI extension, without the thresholding.

CITATION

P. de Lagrange, E. François, D. Doyen, J. Chen, S. Wang, Y. Ye, "AHG8/AHG9: Proposed changes to the candidate new object mask information SEI message", JVET-AE0095, 31st Meeting: Geneva, CH, July 2023.