TUTORIAL

KEYWORDS SUPPLEMENTAL ENHANCEMENT INFORMATION (SEI) // VERSATILE SUPPLEMENTAL ENHANCEMENT INFORMATION (VSEI)

An Introduction to the Supplemental Enhancement Information Message for Coding Video with Constituent Rectangles

By Jill Boyce and Miska M. Hannuksela

The Constituent Rectangles SEI message aims to provide a standardized and interoperable way to code multiple synchronized videos within a single composite video. This will enable use with existing single-layer H.264/AVC, HEVC, and VVC video codec implementations. Use cases include VR/AR, 3D, free-viewpoint video, game streaming, and cloud gaming.

1545-0279/25©2025SMPTE

Abstract

The Constituent Rectangles Supplemental Enhancement Information (SEI) message is under development by the Joint Video Experts Team (JVET) for coding videos formed from compositing multiple videos as constituent rectangles. Many applications require multiple synchronized videos for multiview content or different types of content, such as alpha channels, depth maps, object masks, or image features. Use cases for the SEI message include VR/AR, 3D, free-viewpoint video, game streaming, and cloud gaming. Pictures containing frame-packed rectangular regions have previously been coded with existing single-layer video codecs/profiles in an ad hoc non-standardized manner, without clear description within the video bitstream. The Constituent Rectangles SEI message provides a standardized way to describe a video containing pictures formed by compositing multiple rectangular videos of one or more types using a single-layer video codec profile such as High-Efficiency Video Encoding (HEVC) Main or Versatile Video Coding (VVC) Main profile. The multiple videos otherwise could have been coded as separate pictures in layers of a multi-layer bitstream, requiring the use of a multi-layer video codec profile. The Constituent Rectangles SEI message is targeted for inclusion in a new version of the Versatile Supplemental Enhancement Information (VSEI) standard, a partner standard to the VVC standard, which defines SEI messages providing in-band metadata targeting a variety of use cases. SEI messages defined in VSEI may also be used with other video coding standards, including Advanced Video Coding (H.264/AVC) and HEVC.

The Constituent Rectangles (CR) Supplemental Enhancement Information (SEI) message is under development by the Joint Video Experts Team (JVET) for coding video formed by compositing multiple rectangular video regions. Many applications require the use of multiple synchronized videos for multi-view content or different types of content, such as alpha channels, depth maps, object masks, or image features.

Video coding standards like Advanced Video Coding (H.264/AVC),[1] High-Efficiency Video Coding (HEVC),[2] and Versatile Video Coding (VVC)[3] already provide support for multilayer video and auxiliary pictures, but not in the most widely deployed profiles (HEVC Main, HEVC Main 10, VVC Main 10). Product implementations require synchronization of layers to support the HEVC and VVC multilayer profiles, such as the HEVC Multiview Main, HEVC Scalable Main, and VVC Multilayer Main 10 profiles. The most popular hardware and software implementations of HEVC and VVC do not support the multilayer profiles. A variety of applications use non-standard, ad hoc proprietary methods of composition of multiple videos within a single video to take advantage of hardware and software video codec encoder and decoder implementations.

The Constituent Rectangles SEI message aims to provide a standardized and interoperable way to code multiple synchronized videos within a single composite video. This will enable use with existing single-layer H.264/AVC, HEVC, and VVC video codec implementations. Use cases include VR/AR, 3D, free-viewpoint video, game streaming, and cloud gaming.

Background

Supplemental Enhancement Information (SEI) Messages

The first video coding standard to include SEI messages was H.263, and they are also included in the H.264/AVC, HEVC, VVC, and VSEI standards. SEI messages are used to address a variety of use cases but follow a common design philosophy. SEI messages contain metadata inserted into a bitstream synchronized with the coded video. The information carried in SEI messages is intended to be utilized by the receiver/decoder system, but such usage is optional.

Encoders may optionally include one or more SEI messages in a bitstream. The standardized syntax and semantics must be followed if an SEI message contains a conforming bitstream.

Video coding standards like H.264/AVC, HEVC, and VVC normatively define decoder behavior. For example, a conforming decoder must output decoded pictures identical to those formed by the specified reference decoder. However, using an SEI message is optional for a receiver/decoder system. Video coding standards may provide a recommended post-process but do not define normative behavior required by the receiver/decoder system.

H.264/AVC, HEVC, and VVC profile definitions do not mention support for SEI messages because of the lack of normative decoder behavior. However, application specifi-

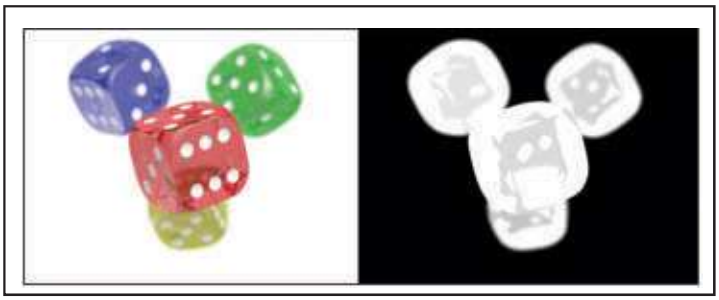

FIGURE 1. Exampled coded picture for texture plus alpha channel using the CR SEI message.

cations, such as those defined by Digital Video Broadcasting (DVB), Advanced Television Systems Committee (ATSC) and 3rd Generation Partnership Project (3GPP) etc., may require using SEI messages in receiver/decoder systems.

Commonly used SEI messages include the buffering period SEI, picture timing SEI, mastering display color volume SEI, and frame packing arrangement SEI.

Versatile Supplemental Enhancement Information (VSEI) Standard

The Versatile Supplemental Enhancement Information (VSEI)[4] standard is a partner to the VVC standard. SEI messages defined in VSEI may also be used with other video coding standards, including H.264/AVC and HEVC.

VVC includes a list of the supported SEI messages and their payload type. It also defines interface variables for some of the SEI messages defined in VSEI, allowing the VSEI standard to remain codec-independent without directly using specific syntax elements from VVC.

Because SEI messages do not define normative decoder behavior, newly defined SEI messages can be used in bitstreams conforming to VVC and HEVC profiles defined in earlier versions of the VVC and HEVC standards.

Frame Packing Arrangement SEI

A frame packing arrangement SEI message already supports packing two stereo video views in HEVC and VVC. This message specifies several packing options for arranging two constituent frames using single pictures, including top-bottom, side-by-side, and temporal interleaving.

FIGURE 3. Exampled coded picture for multi-view with three views using the CR SEI message.

FIGURE 2. Exampled coded picture for texture plus depth map using the CR SEI message.

1545-0279/25©2025SMPTE

FIGURE 4. Example coded picture for texture plus depth map for multiview with three views using the CR SEI message.

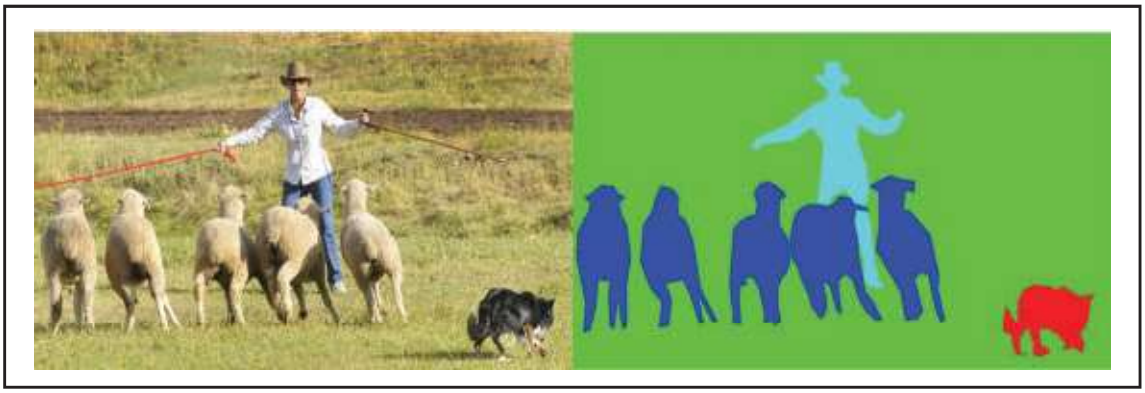

FIGURE 5. Example coded picture for texture plus object mask using the CR SEI message.

Constituent Rectangles SEI Message

The Constituent Rectangles (CR) SEI message provides syntax that allows an encoder to encode a video formed by compositing any number of rectangular videos of any video type with any non-overlapping layout. The information provided in the SEI message enables a receiver/decoder to interpret the received video correctly.

Use Case Examples

The Constituent Rectangles SEI message can be used in a variety of cases in which multiple synchronized videos are required.

A picture coded with the Constituent Rectangles (CR) SEI message may contain two or more views of a 3D scene, each represented by a constituent rectangle. For example, a picture may contain a video and an alpha channel constituent rectangle. Another example of usage is a picture containing constituent rectangles representing video and depth for one or more views.

Figure 1 shows an example coded picture using the CR SEI containing two CRs: one rectangle representing a normal video texture picture and one rectangle representing an alpha channel picture. Without the CR SEI message, the alpha channel could be coded as an auxiliary picture in an

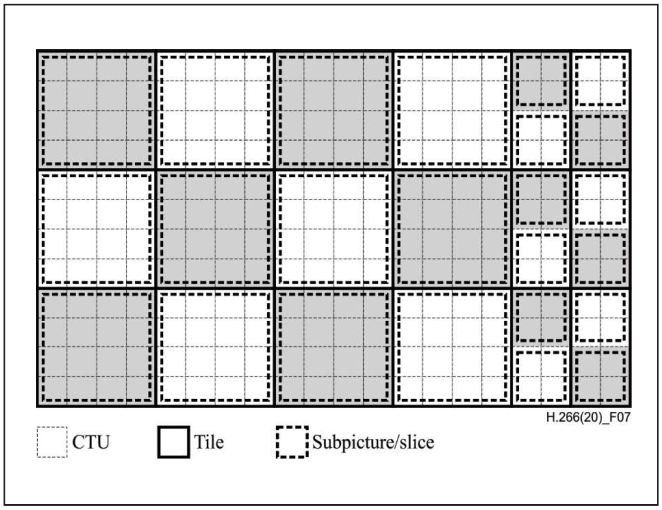

FIGURE 6. Subpicture partitioning example in VVC.

additional layer using the VVC Multilayer Main 10 profile. However, the VVC Main 10 profile can be used with the CR SEI.

Figure 2 shows an example coded picture using the CR SEI containing two CRs, with one rectangle representing a normal video texture picture and one rectangle representing a depth picture, from the MPEG Immersive Video (MIV) Painter sequence, Dziembowski.[5]

Without the CR SEI message, the depth map could be coded as an auxiliary picture in an additional layer using the VVC Multilayer main profile. The VVC Main 10 profile can be used with the CR SEI, and the bitstream can be encoded in a single layer, as shown in **Fig. 2.**

Figure 3 shows an example coded picture using the CR SEI with three rectangles, representing three texture views of a 3D multiview scene containing three views.

Figure 4 shows an example coded picture using the CR SEI. Six rectangles represent three texture views and three corresponding depth views of a 3D multiview scene.

Figure 5 shows an example coded picture using the CR SEI. Two rectangles represent a texture view and a corresponding object mask.

These examples illustrate some of the flexibility of the CR SEI message. The number of rectangles and the layout of the rectangles within a composite picture are selected by the encoder to address the particular use case.

Standardization Status

The Constituent Rectangles SEI message was first proposed in JVET-AH0162, Boyce,[6] at the April 2024 meeting of the Joint Video Experts Team (JVET). It was adopted into the "Technologies under consideration for future extensions of VSEI," McCarthy.[7] A software implementation of the CR SEI message within the VTM reference software for VVC was made available.[8]

Several modifications to the CR SEI message were proposed in JVET-AI0178, Boyce,[9] which were adopted into the updated version of the "Technologies under consideration for future extensions of VSEI," McCarthy,[10] which is updated at each meeting of the JVET. Contributions included in the technologies under consideration document will be considered in a working draft of the new version of VSEI.

A working draft of the fourth version of VSEI is available in Boyce[11] and is planned to be issued as a Committee Draft (CD) in 2024. It is expected to be standardized in 2025.

Syntax Parameters

The full syntax of the CR SEI message can be found.[10]

The CR SEI syntax includes several syntax elements, many of which are optional, as listed below.

- Number of CRs

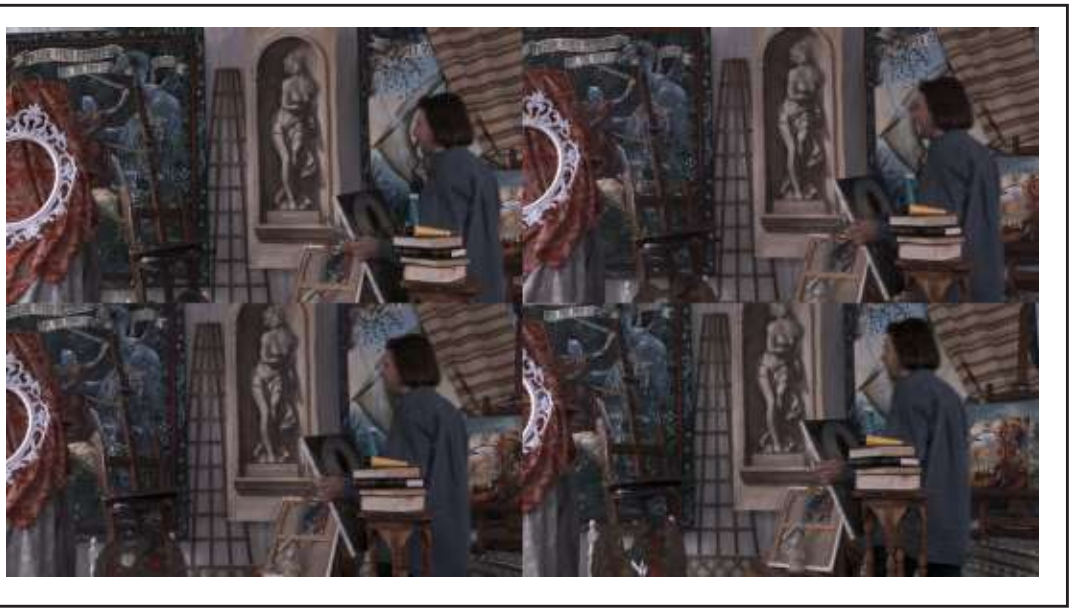

FIGURE 7. 2-row x 2-column "same size" signaling grid for four rectangles using the CR SEI message.

1545-0279/25©2025SMPTE

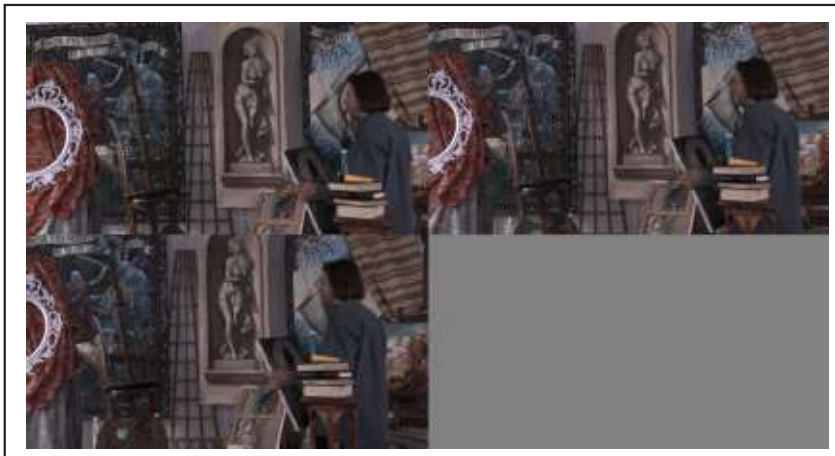

FIGURE 8. 2-row x 2-column "same size" signaling grid for three active rectangles and one empty rectangle using the CR SEI message.

FIGURE 9. Explicit size and position signaling of four rectangles using a unit size using the CR SEI message.

- Size and position of CRs
- Optional CR type
- Optional CR ID
- Optional Associated CR ID
- Optional CR group ID
- Optional CR description

The SEI message syntax provides several options for signaling the size and position of the CRs, as described in the next section. These options minimize the bitrate used in the SEI message for the particular use case.

Types and ID values may be signaled or inferred to save bitrate when signaling is not required.

The CR type may optionally be signaled for each rectangle, which can be used to indicate that it is texture (e.g., normal video), alpha channel, depth map, object mask, etc.

A CR ID may be explicitly signaled or inferred to reduce the bitrate.

An optional Associated CR ID may be signaled, with auxiliary types such as alpha, depth, and object masks associated with a primary rectangle. An optional Group ID may be signaled to flexibly indicate groups of CRs, which can indicate the view ID of a CR SEI containing one or more rectangles from multiple views. Usage of the Associated CR ID and Group ID are described in Ref. 9.

Options for signaling of size and position

Several options are supported for signaling the CR size and position, allowing flexibility for a variety of use cases, some of which may include large numbers of CRs in a picture. **Figure 6** shows an example of subpicture partitioning from the VVC specification.

The VVC standard includes a subpicture feature, in which a picture may be partitioned into subpictures, for which the coding tools may treat the subpicture picture boundaries like picture boundaries. If subpictures are present, rectangle

sizes and positions may be derived from the subpicture sizes and positions by mapping CRs to subpictures, avoiding the need for duplicate signaling.

When all rectangles are the same size, their size and position can be efficiently signaled based on their position in a grid of rectangles, rows, and columns. **Figure 7** shows an example of four rectangles of the same size in a 2-row x 2-column grid for a multiview use case.

Because even when "same size" signaling is used, the number of rectangles might not map into a convenient grid, it is possible to indicate a CR type as empty. **Figure 8** shows an example of "same size" signaling with an empty rectangle set to gray.

Finally, a third option is to explicitly signal the size and position of each rectangle using a specified unit size.

Figure 9 shows an example of explicit signaling in a multiview use case. The use case involves two texture views with associated depth maps, with the depth maps at 1/4 the resolution of the texture views. The unit size grid for this example is also illustrated.

Conclusion

JVET is developing the Constituent Rectangles SEI message for potential inclusion in a future version of the VSEI standard. The message's flexibility enables a wide range of current and future use cases requiring the synchronization of multiple videos in a standardized and interoperable manner using widely deployed single-layer video codecs.

Upon including the CR SEI message in the VSEI standard, the VVC and HEVC standards can define a payload type to reference it, which can be used in VVC or HEVC bitstreams conforming to any profiles defined in those standards.

References

1. International Telecommunication Union-Telecommunication (ITU-T) Recommendation H.264 | International Organization for Standardization /International Electrotechnical Commission (ISO/IEC) 23008-2, "Advanced video coding." [Online]. Available: https://www.itu.int/rec/T-REC-H.264
2. International Telecommunication Union-Telecommunication (ITU-T) Recommendation H.265 | International Organization for Standardization /International Electrotechnical Commission (ISO/IEC) 23008-2, High-efficiency video coding." [Online]. Available: https://www.itu.int/rec/T-REC-H.265
3. International Telecommunication Union-Telecommunication (ITU-T) Recommendation H.266 | International Organization for Standardization /International Electrotechnical Commission (ISO/IEC) 23090-3, "Versatile Video Coding." [Online]. Available: https://www.itu.int/rec/T-REC-H.266
4. International Telecommunication Union-Telecommunication (ITU-T) Recommendation H.274 | International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) 23002-7, "Versatile supplemental enhancement information messages for coded video bitstreams." [Online]. Available: https://www.itu.int/rec/T-REC-H.274
5. A. Dziembowski, B. Kroon, and J. Jung (Eds.), *Common test conditions for MPEG immersive video,* ISO/IEC JTC 1/SC 29/WG 04 N 0372, Jul. 2023, Geneva. [Online]. Available: https://mpeg-miv.org/index.php/content-database-2/
6. J. Boyce, M. M. Hannuksela, "AHG9: Constituent rectangles SEI," document JVET-AH0161, 34th JVET Meeting. Apr. 2024. [Online]. Available: https://jvet-experts.org/doc_end_user/documents/34_Rennes/wg11/JVET-AH0162-v1.zip
7. S. McCarthy, J. Boyce, J. Chen, S. Deshpande, M. M. Hannuksela, H. Tan, and Y.-K. Wang, "Technologies under consideration for future extensions of VSEI (version 4)," document JVET-AH2006, 34th JVET Meeting, April 2024. [Online]. Available: https://jvet-experts.org/doc_end_user/documents/34_Rennes/wg11/JVET-AH2032-v3.zip
8. Joint Video Experts Team (JVET)-AH0162 SW patch Merge branch 'JVET_AH0162' into 'jvet-tuc' (6fd78072) Commits jvet-tuc /VVCSoftware VTM GitLab (fraunhofer.de)
9. J. Boyce and M. M. Hannuksela, "AHG9: On Constituent rectangles SEI," document JVET-AI0178, 35th JVET Meeting, July 2024.[Online]. Available: https://jvet-experts.org/doc_end_user/documents/35_Sapporo/wg11/JVET-AI0178-v1.zip
10. S. McCarthy, J. Boyce, J. Chen, S. Deshpande, M. M. Hannuksela, H. Tan, and Y.-K. Wang, "Technologies under consideration for future extensions of VSEI (version 5)," document JVET-AI2006, 35th JVET Meeting, July 2024. [Online]. Available: https://jvet-experts.org/doc_end_user/documents/35_Sapporo/wg11/JVET-AI2006-v2.zip
11. J. Boyce, J. Chen, S. Deshpande, M. M. Hannuksela, S. McCarthy, G. J. Sullivan, H. Tan, and Y.-K. Wang, "Additional SEI messages for VSEI version 4 (Draft 3)," document JVET-AI2006, 35th JVET Meeting, July 2024. [Online]. Available: https://jvet-experts.org/doc_end_user/documents/35_Sapporo/wg11/JVET-AI2032-v3.zip

About the Authors

Jill Boyce is a Distinguished Engineer in Video coding at Nokia Technologies and is an IEEE Fellow recognized for her contributions to the field of video coding.

Miska M. Hannuksela is the head of video research at Nokia Technologies and a Nokia Bell Labs Fellow.

Presented at the SMPTE 2024 Media Technology Summit, Hollywood, CA, 21-24 Oct. 2024. *Copyright © 2025 by SMPTE.*

DOI: 10.5594/JMI.2025/LZES6606
Date of publication: 24 January 2025