

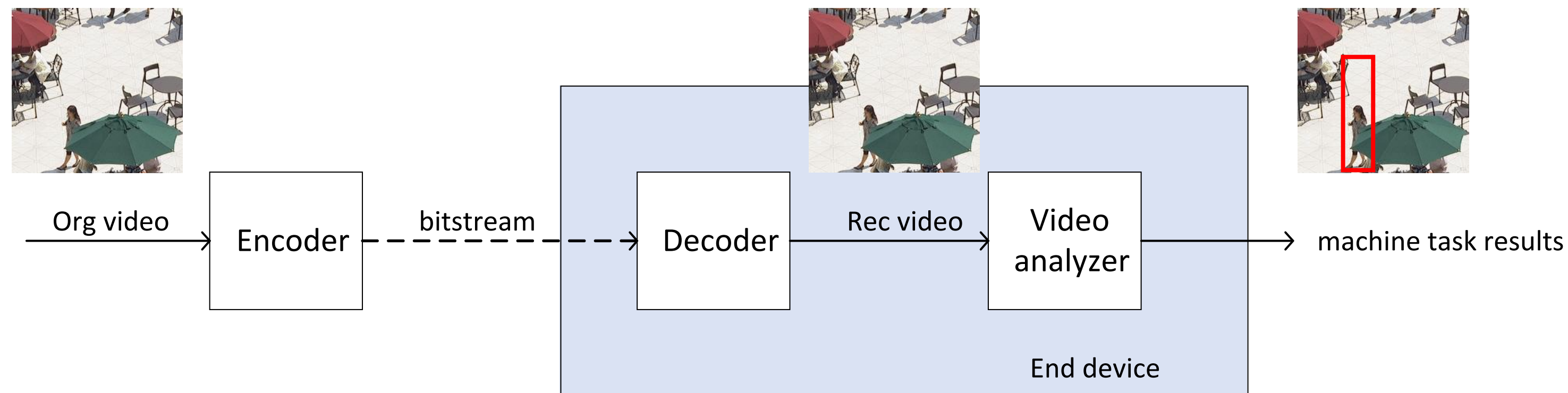
JVET-AE0095 AHG8/AHG9: Proposed changes to the candidate new object mask information SEI message

J. Chen, S. Wang, Y. Ye (Alibaba)

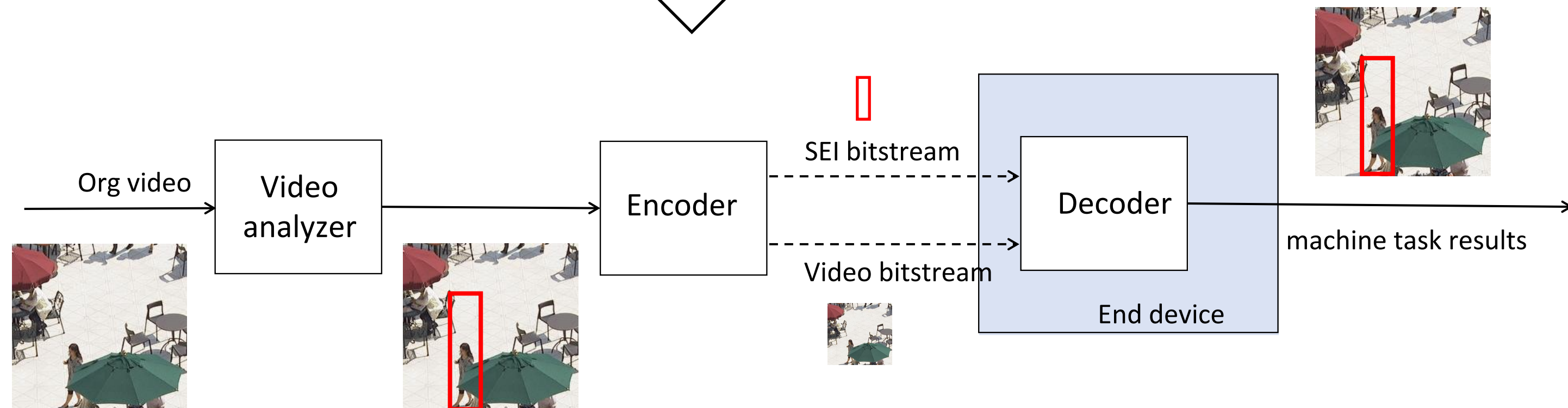
P. de Lagrange, E. François, D. Doyen (InterDigital)

Background (1/2)

- Annotated Region SEI message was adopted in VSEI, HEVC and AVC
 - Machine task (e.g., object detection, tracking) only needs to be performed in encoder-side
 - The task results are sent to the decoder through the SEI message



with AR SEI message



- Decoder power consumption is reduced
- More powerful network is used to perform the machine task in encoder side
- Original videos is used for machine task leading to more accurate results

Background (2/2)

- Annotated Region (AR) SEI message uses bounding box to indicate the region covered by an object of interest
 - The object shape information is lost
- However, the object shape information is essential to some applications
 - Intelligent traffic system: traffic accident recognition
 - Video conferencing system: virtual background switch
 -
- JCTVC-AF0024 proposed to add shape information in AR SEI message by sending mask itself in the SEI message
 - It was commented in the meeting, “this is basically sending coded bi-level pictures inside an SEI message and that this seems like it is not what we would ordinarily do with SEI messages”
 - Some experts suggest using auxiliary picture as the mask picture



What was proposed in JVET-AD0175

- It was proposed to add a new auxiliary picture type for object mask
 - The object mask is coded as an auxiliary picture and the sample values of the auxiliary picture represent the IDs of the mask (i.e., at most 1024 masks for a 10-bit auxiliary picture)
 - The samples with the same value form a mask, and the samples with different values belong to different masks
 - Multiple object mask auxiliary pictures can be used
 - To support masks overlapped (the samples with same position but in the different mask picture could belong to different masks overlapped with each other)
 - To support more than $1 \ll \text{BitDepthY}$ masks
- It was proposed to add a new SEI message to send the object mask auxiliary picture related information (OMI SEI message)
 - ID of each mask
 - Depth information: to determine which object is visible and which object is occluded in case of overlapping
 - Confidence: representing the degree of the confidence of the object mask
 - Label: to give any attribute of the object

What was the problem in JVET-AD0175

- It was asked whether the lossless coding is necessary for object mask auxiliary picture.
 - Lossless coding results in too much bit-overhead
 - When lossy coding is used for the proposed mask auxiliary pictures, the sample values inside the pictures are not guaranteed to perfectly match the object ID, as the object mask ID is represented by the sample value

The solutions in this contribution

- Change 1: add a threshold specifying a tolerance range around the object ID values.
 - lossy coding can be used as long as the distortion is within the tolerance range for mask region
- Change 2: send a bounding box parameters that delimit the masks' location
 - Encoder can do whatever for the regions out of bounding box to reduce the bitrate (e.g., using a extremely large QP, any mode)

An example

- Two examples of object mask auxiliary picture

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0
0	0	5	0	0	0	0	0	0	0	0	0	0	20	0	0	0
0	0	5	5	0	0	0	0	0	0	0	20	20	20	20	20	0
0	0	5	5	5	0	0	0	0	0	0	0	0	20	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Mask representation in JVET-AD0175

mask1: blue area, ID=5
mask2: yellow area, ID=20

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
18	19	20	21	22	23	24	25	26	27	28	15	35	23	29	2	30
1	2	5	2	1	7	8	9	10	11	12	6	7	20	26	10	22
3	4	4	3	8	13	14	15	16	17	18	20	18	21	22	20	23
5	6	6	6	7	19	20	21	22	23	24	19	20	17	24	15	24
7	8	9	10	11	12	13	14	15	16	17	18	18	20	14	16	25
10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
10	13	24	25	33	20	15	10	9	9	9	12	15	33	34	22	15

Mask representation in this contribution

mask1: blue area, ID=5, tolerance range [3, 7]
mask2: yellow are, ID=20 with tolerance range [17, 23]

The syntax changes (1/2)

- Not assign a particular value of sdi_aux_id to object mask auxiliary picture, but use unspecified values which is indicated by omi_aux_id_minus128
- Multiple primary pictures can be supported and layer_id of the primary is indicated
- The number of auxiliary picture associated with each primary picture is already indicated by SDI SEI message

object_mask_info(payloadSize) {	Descriptor
omi_cancel_flag	u(1)
if(!om_cancel_flag) {	
omi_aux_id_minus128	ue(v)
omi_num_primary_pic_layer_minus1	ue(v)
for(i=0; i<=omi_num_primary_pic_layer_minus1;i++)	
omi_primary_pic_layer_id[i]	ue(v)
omi_mask_id_length_minus8	ue(v)
omi_mask_confidence_info_present_flag	u(1)
if(omi_mask_confidence_info_present_flag)	
omi_mask_confidence_length_minus1	u(4)
omi_mask_depth_info_present_flag	u(1)
if(omi_mask_depth_info_present_flag)	
omi_mask_depth_length_minus1	u(4)
omi_mask_label_info_present_flag	u(1)
if(omi_mask_label_info_present_flag) {	
omi_mask_label_language_present_flag	u(1)
if(omi_mask_label_language_present_flag) {	
while(!byte_aligned())	
omi_bit_equal_to_zero	f(1)
omi_mask_lable_language	st(v)
}	
}	
}	

The syntax changes (2/2)

- Change 1: add a tolerance parameter around the object ID value
 - option 1: omi_mask_id_delta[i][j][k] is signalled for each omi_mask_id[i][j][k].
 - option 2: omi_mask_id_delta[i][j] is signalled for each auxiliary picture.
- Change 2: add bounding box parameters to delimit the object masks in the auxiliary picture

for (i=0; i<= omi_num_primary_pic_layer; i++) {	
for(j=0;j< numAuxLayer[omi_primary_pic_layer_id[i]]; j++){	u(1)
om_mask_pic_update_flag[i][j]	f(1)
if(omi_mask_pic_update_flag[i][j]) {	
om_num_mask_in_pic_update[i][j]	ue(v)
omi_mask_id_delta[i][j]	u(v)
for(k=0; k<omi_num_mask_in_pic_update[i][j]; k++) {	
omi_mask_id[i][j][k]	u(v)
omi_mask_id_delta[i][j][k]	u(v)
omi_mask_top[i][j][k]	u(v)
omi_mask_left[i][j][k]	u(v)
omi_mask_width[i][j][k]	u(v)
omi_mask_height[i][j][k]	u(v)
if(maskIdExist[i][j][on_mask_id[i][j][k]]) {	
om_mask_cancel[i][j][k]	u(1)
maskIdExist[i] [j] [on_mask_id[i] [j] [k]]=	
!on_mask_cancel[i][j][k]	
}	
else {	
maskIdExist[i] [j] [on_mask_id[i][j][k]]=1	
}	
if(maskIdExist[i] [j] [on_mask_id[i][j][k]]) {	
if(omi_mask_confidence_info_present_flag)	
om_mask_confidence[i][j][k]	u(v)
if(omi_mask_depth_info_present_flag)	
om_mask_depth[i][j][k]	u(v)
while(!byte_aligned())	
om_bit_equal_to_zero	f(1)
if(omi_mask_label_info_present_flag)	
om_mask_label[i][j][k]	st(v)
}	

Summary

- The object mask auxiliary picture and OMI SEI message were proposed in JVET-AD0175 to enhance VVC's capability of handling information from machine analysis tasks.
- To solve the potential issues when lossy coding is used, a tolerance range around the object ID values as well as the bounding box parameters delimiting the masks' location are added in this contribution.
- It is asserted that with the increasing importance of machine analysis and artificial intelligence applications, JVET should support more non-normative ways of signaling information obtained from machine analysis tasks at the encoder side.
- Thus, it is suggested to adopt the proposed SEI message to the next version of VSEI.

Thanks