

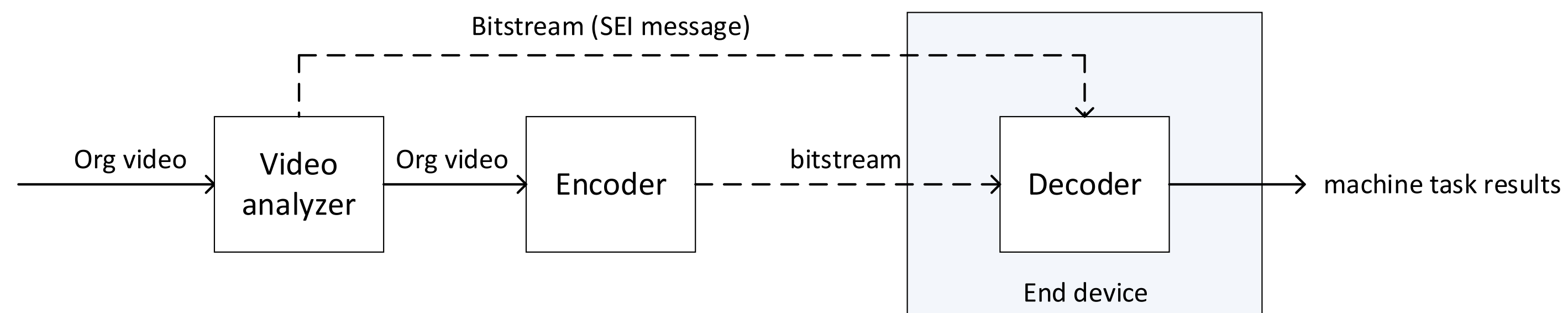
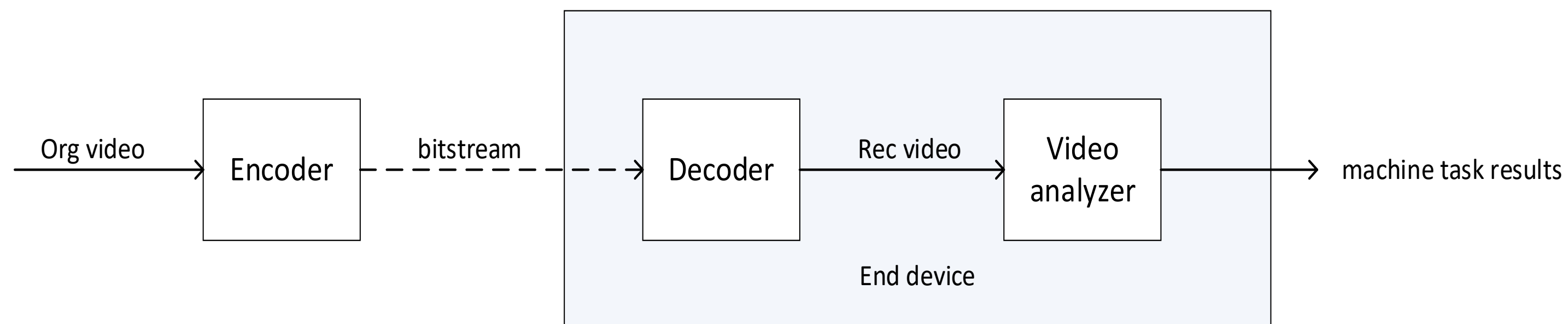
AHG8/AHG9: On object mask auxiliary picture and object mask information SEI message for VSEI and HEVC



Jie Chen, Shurun Wang, Yan Ye
Alibaba

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



- Reduce decoder power consumption
- Use more powerful network to perform the machine task in encoder side
- Perform machine task in uncompressed video which has higher quality

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
 -
- JVET-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
- As more video is being coded for machine analysis purposes, it is appropriate to revisit carrying mask information as part of the video bitstream

Proposal (1/3)

- It is proposed to add a new auxiliary picture type for object mask in VSEI and HEVC
 - object mask is coded as an auxiliary picture
 - the proposed mask auxiliary picture is a multi-value picture instead binary-value picture
 - the sample values of the auxiliary picture represents 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

Proposal (2/3)

- 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	20	0	0
0	0	5	0	0	0	0	10	10	10	0	0	0	20	0	0
0	0	5	5	0	0	0	10	10	10	0	20	20	20	20	20
0	0	5	5	5	0	0	10	10	10	0	0	0	20	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

Example 1: a 16x8 object mask auxiliary picture

- mask0: white area, ID=0
- mask1: blue area, ID=5
- mask2: green area, ID=10
- mask3: yellow area, ID=20

5	5	0	0
5	5	0	0
0	0	0	0
0	0	0	0

object mask auxiliary picture 0

0	0	0	0
0	10	10	0
0	10	10	0
0	0	0	0

object mask auxiliary picture 1

Example 2: two 4x4 object mask auxiliary picture

The mask with ID 5 is overlapped with the mask with ID 10 at position (1,1)

Proposal (3/3)

- It is proposed to add a new SEI message to send the object mask auxiliary picture related information
 - It is called object mask information (OMI) SEI message
 - It uses persistence mechanism and only needs to be re-sent when there are any changes of the object masks.
 - When the OMI message is signaled, only the updated part of mask information needs to be sent (by signaling update flag to control the presence of following syntax elements)
 - The main information sent in OMI SEI message for each mask includes
 - 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
 - Multiple primary pictures are supported and all the information is included in an SEI message

Proposed Syntax for VSEI (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)
}	
}	

Proposed Syntax for VSEI (2/2)

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)
for(k=0; k<omi_num_mask_in_pic_update[i][j]; k++) {	
omi_mask_id[i][j][k]	u(v)
if(maskIdExist[i][j][on_mask_id[i][j][k]]) {	
omi_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)	
omi_mask_confidence[i][j][k]	u(v)
if(omi_mask_depth_info_present_flag)	
omi_mask_depth[i][j][k]	u(v)
while(!byte_aligned())	
omi_bit_equal_to_zero	f(1)
if(omi_mask_label_info_present_flag)	
omi_mask_label[i][j][k]	st(v)
}	
}	
}	
}	

Proposed Syntax for HEVC (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
- The number of object mask auxiliary pictures associated with the SEI message is indicated

object_mask_info(payloadSize) {	Descriptor
omi_cancel_flag	u(1)
if(!om_cancel_flag) {	
omi_aux_id_minus128	ue(v)
omi_num_mask_pic_minus1	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)
}	
}	
}	
}	

Proposed Syntax for HEVC (2/2)

for (i=0; i<=om_num_mask_pic_minus1; i++) {	
om_mask_pic_update_flag [i]	f(1)
if(om_mask_pic_update_flag[i]) {	
omi_mask_pic_layer_id [i]	u(6)
om_num_mask_in_pic_update [i]	ue(v)
for(j=0; j<on_num_mask_in_pic_update[i]; j++) {	
om_mask_id [i] [j]	u(v)
if(maskIdExist[i][on_mask_id[i] [j]]) {	
om_mask_cancel [i] [j]	u(1)
maskIdExist[i][on_mask_id[i] [j]]= !on_mask_cancel[i] [j]	
}	
else {	
maskIdExist[i][on_mask_id[i] [j]]=1	
}	
if(maskIdExist[i][on_mask_id[i] [j]]) {	
if(omi_mask_confidence_info_present_flag)	
omi_mask_confidence [i] [j]	u(v)
if(omi_mask_depth_info_present_flag)	
omi_mask_depth [i] [j]	u(v)
if(omi_mask_label_info_present_flag){	
while(!byte_aligned())	
omi_bit_equal_to_zero	f(1)
omi_mask_label [i] [j]	st(v)
}	
}	
}	
}	
}	

Summary

- A new type of auxiliary picture is proposed to represent the object masks associated with the primary picture
- A corresponding new SEI message is also proposed to signal the information on the object mask auxiliary picture
- Based on different design of the auxiliary picture in VSEI and HEVC, different syntax is proposed for VSEI and HEVC

Thanks

