

JVET-AE0090

AHG8/AHG9: On Machine Vision Indication


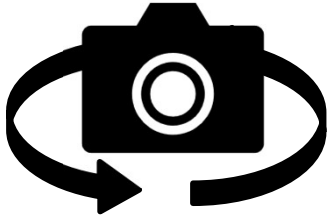
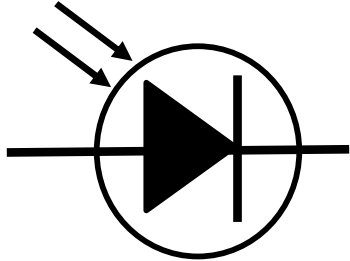
Jingying Gao, Han Boon Teo, Chong Soon Lim,  
Kiyofumi Abe

# Motivation

- With the rise of machine learning/neural network-based applications and the abundance of sensors, more and more video data are consumed by machines rather than humans.
- The Annotated Regions SEI message is specified in VSEI (ITU-T H.274 | ISO/IEC 23002-7) and carries parameters that identify annotated regions using bounding boxes representing the size and location of identified objects.
  - **ar\_not\_optimized\_for\_viewing\_flag** equal to 1 indicates that the decoded image is not optimized for human vision, but rather is optimized for some other purpose.
  - **ar\_not\_optimized\_for\_viewing\_flag** equal to 0 indicates that the decoded image may or may not be optimized for human vision.

**There is no indication that the decoded image is not optimized for machine vision**

# Motivation

	 Infrared Sensor	 Accelerometer Sensor	 Photodiode Sensor
<b>Not</b> for machine vision	no movement detected no object detected	presence of motion blur	presence of excessive light reflection

**The decoded image not optimized for machine vision may result in lower accuracy**

# Proposal: option 1

## Option 1: Adding machine vision indication SEI message in VSEI

Table 1 – Persistence scope of SEI messages (informative)

SEI message	Persistence scope
...	...
Machine vision indication	The picture associated with the SEI message

<b>machine_vision_indication</b> ( payloadSize ) {	<b>Descriptor</b>
<b>mvi_not_optimized_for_machine_vision_flag</b>	u(1)
}	

The machine vision indication SEI message carries information to identify the possible usage of decoded image associated with this SEI message.

**mvi\_not\_optimized\_for\_machine\_vision\_flag** equal to 1 indicates that the decoded image associated with machine vision indication SEI message is not optimized for machine vision, but rather is optimized for some other purpose such as human vision. **mvi\_not\_optimized\_for\_machine\_vision\_flag** equal to 0 indicates that the decoded image associated with machine vision indication SEI message may or may not be optimized for machine vision. When not present, the value of **mvi\_not\_optimized\_for\_machine\_vision\_flag** is inferred to be equal to 0.

# Proposal: option 2

## Option 2: Adding extension mechanism to Annotated Regions SEI message

**ar\_extension\_flag** equal to 0 specifies that no extension data is present in Annotated Regions SEI message. **ar\_extension\_flag** equal to 1 specifies that extension data might be present in Annotated Regions SEI.

**ar\_not\_optimized\_for\_machine\_vision\_flag** equal to 1 indicates that the decoded images that the Annotated Regions SEI message applies to are not optimized for machine vision, but rather are optimized for some other purpose such as human vision.

**ar\_not\_optimized\_for\_machine\_vision\_flag** equal to 0 indicates that the decoded images that the Annotated Regions SEI message applies to may or may not be optimized for machine vision.

annotated_regions( payloadSize ) {	Descriptor
<b>ar_cancel_flag</b>	u(1)
if( !ar_cancel_flag ) {	
<b>ar_not_optimized_for_viewing_flag</b>	u(1)
<b>ar_true_motion_flag</b>	u(1)
<b>ar_occluded_object_flag</b>	u(1)
<b>ar_partial_object_flag_present_flag</b>	u(1)
<b>ar_object_label_present_flag</b>	u(1)
...	
}	
<b>ar_extension_flag</b>	u(1)
if(ar_extension_flag && !ar_cancel_flag){	
<b>ar_not_optimized_for_machine_vision_flag</b>	u(1)
}	
}	

## Proposal: option 3

### *Option 3: modifying `ar_not_optimized_for_viewing_flag` in Annotated Regions SEI message*

*Change `ar_not_optimized_for_viewing_flag` to `ar_not_optimized_for_viewing_idx` in the Annotated Regions SEI.*

annotated_regions( payloadSize ) {	<b>Descriptor</b>
<b>ar_cancel_flag</b>	u(1)
if( !ar_cancel_flag ) {	
<del>ar_not_optimized_for_viewing_flag</del>	u(1)
<b>ar_not_optimized_for_viewing_idx</b>	u(2)
<b>ar_true_motion_flag</b>	u(1)
<b>ar_occluded_object_flag</b>	u(1)
<b>ar_partial_object_flag_present_flag</b>	u(1)
<b>ar_object_label_present_flag</b>	u(1)
...	
}	
...	
}	

## Proposal: option 3

### *Option 3: modifying `ar_not_optimized_for_viewing_flag` in Annotated Regions SEI message*

`ar_cancel_flag` equal to 1 indicates that the SEI message cancels the persistence of any previous Annotated Regions SEI message that is associated with one or more layers to which the Annotated Regions SEI message applies.

`ar_cancel_flag` equal to 0 indicates that Annotated Regions information follows.

`ar_not_optimized_for_viewing_idx` identifies the intended usage of the image as specified in Table 2.

Table 2: Definition of `ar_not_optimized_for_viewing_idx`

Value	Interpretation
0	Indicates that the decoded pictures that the annotated regions SEI message applies to may or may not be optimized for user viewing.
1	Indicates that the decoded pictures that the annotated regions SEI message applies to are not optimized for user viewing, but rather are optimized for some other purpose such as algorithmic object classification performance.
2	Indicates that the decoded pictures that the annotated regions SEI message applies to are not optimized for machine viewing, but rather are optimized for some other purpose such as user viewing.
3	Reserved for future use.

## Example of Use Case

Machine vision inference speed can be slow.

Table 3: Examples of average inference speeds of machine vision tasks and models adopted in VCM evaluation framework.

Machine Vision Task	Model	GPU	Input resolution	Inference Speed (FPS)
Object Detection	Faster RCNN X101-FPN	Nvidia V100 GPU	640x480	<b>10.2</b> [1]
Object Tracking	JDE-1088x608	Nvidia Titan Xp GPU	1088x608	<b>22.2</b> [2]

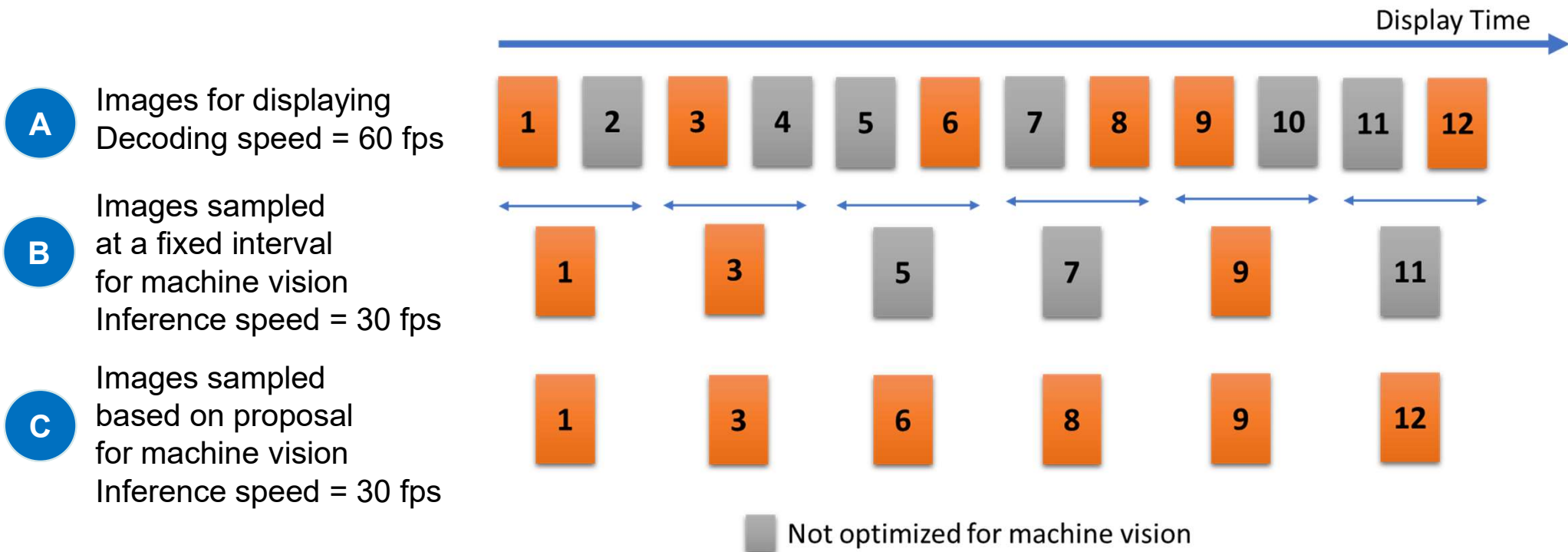
**It's not feasible to perform inference on every decoded image.**

[1] Zhongdao, Z., Shi, J., Wang, L., Li, Y., & Yeung, D. Y. (2021). Towards Realtime MOT: Detection, Tracking, and Recognition of Multiple Objects in Video. GitHub. Retrieved from <https://github.com/Zhongdao/Towards-Realtime-MOT>

[2] Facebook Research. (n.d.). Detectron2 Model Zoo. GitHub. Retrieved from [https://github.com/facebookresearch/detectron2/blob/main/MODEL\\_ZOO.md](https://github.com/facebookresearch/detectron2/blob/main/MODEL_ZOO.md)



## Example of Use Case



**Based on our proposal, machine vision tasks may have better input images. This may result in higher accuracy.**

## Conclusion

- It is proposed to add machine vision indication in VSEI by implementing one of the following options:
  - ❑ *Option 1*: add machine vision indication SEI message in VSEI.
  - ❑ *Option 2*: add extension mechanism to Annotated Regions SEI message.
  - ❑ *Option 3*: modify *ar\_not\_optimized\_for\_viewing\_flag* in the Annotated Regions SEI message.

**It is proposed to adopt one of the proposed changes to VSEI.**