

SOURCE: Japan

TITLE: Backward compatibility with H.261 using NEXT profile

PURPOSE: Discussion

1. Introduction

Spatial scalability is an efficient tool for layered coding from the coding efficiency point of view, especially when the bit rate of the upper layer and the lower layer are close. From this point, it seems necessary to keep H.261 compatibility in the NEXT profile as a tool to be used in the future.

A video coding tool for layered coding was prepared in the NEXT profile frozen at the previous meeting in N.Y., but in addition to that, the help of H.22X and H.24X? (Communication procedure) seems to be necessary for realizing H.261 compatibility.

The purpose of this document is to show a supposed way to realize H.261 compatibility and to discuss some issues to be clarified, referring to the current Video and System WD, in order to confirm if there are any misunderstandings or problems or insufficiencies in the current regulation.

2. Video (H.26X)

A video bit stream consists of two (or more) independent streams, i.e. the upper layer and the lower layer, and multiplexed in H.22X. The corresponding syntax and parameter setting for the upper layer will be as in 2.1. There is no special syntactic constraint in the lower layer.

2.1 Syntax and parameter setting for the upper layer

(1) Sequence_scalable_extension

lower_level_prediction_horizontal_size	352
lower_level_prediction_vertical_size	288
scalable_mode	"01" (spatial scalability)

(2) Picture_scalable_extension

lower_layer_temporal_reference	TR (in 10 bits)
horizontal_subsampling_ratio_m	1
horizontal_subsampling_ratio_n	2
vertical_subsampling_ratio_m	3
vertical_subsampling_ratio_n	5
lower_level_horizontal_offset	0
lower_level_vertical_offset	0

2.2 Issues to be clarified

(1) Will the VBV of the upper layer work in its own bit rate or in the total bit rate? If we suppose a first transmission of the lower layer and the next transmission of the upper layer in a frame-by-frame basis, the VBV cannot work as it does in the one layer case. This is more serious when a picture contains a large amount of bits and generate many skipped pictures in the lower layer.

If we don't use prediction from the lower layer when a picture skipping occurs in the lower layer, will the problem be settled?

(2) The temporal reference of H.261 has only 5 bits. It will lead to a possibility that the temporal reference will go to another round and cannot be identified when the bit rate of the lower layer is low and many skipped pictures are generated.

cf) The maximum number of bits for 1 frame in H.261 is 256 K bits.

(The efficiency gain by layering is small when the bit rate of the lower layer is low. So, this may not be a practical problem.)

From (1) and (2), it seems to be desirable to avoid generating many skipped pictures.

(3) The description of the up sampling filter seems to be wrong.

In video WD p.83,l.19, $\text{phase}=(i*p)\%q*8//q$ seems to be correct.

The same for P.84,l.5, $\text{phase}=(i*m)\%n*16//n$ seems to be correct.

(4) There is no way prepared to show whether the picture in the lower layer is progressive or interlaced in the H.26X syntax. Is it correct to understand that this information will be provided by H.22X or H.24X?

(5) The same for the chroma signal of the lower layer.

3. System (H.22X)

The individual independent streams for the upper layer and the lower layer multiplexed in H.22X have to have indicators to show to which layer the data corresponds in the transmission unit basis.

The MPEG system group seems to be preparing corresponding tools PSI(for TS) and PSM(for PS) ("hierarchy descriptor" in the system WD P.52 seems to be the corresponding part).

The recent tendency for H.22X seems that, in a usual case, H.22X transmits PES packets by ATM cells, so PSM will be used. Then, the data will be identified by the stream_id.

Anyway, the current status of the semantics for a hierarchy descriptor in the system WD is empty and it needs full description.

It is desirable to transmit the lower layer first and the upper layer next for delay and memory size points of view.

4. Communication procedure

Generally, many kinds of combinations of the lower layer and the upper layer will be allowed for the NEXT profile and then H.32X terminals will have to exchange their decoding capability for each other especially for their lower layer when communication starts.

The exception seems to be the case when only H.261 is allowed for the lower layer of the H.32X terminals supporting layered coding.

Which will be the case of the above two?

5. Conclusion

A definite way of realizing backward compatibility with H.261 has been discussed. If we are going to realize the compatibility of H.32X terminals with H.320 terminals using spatial scalability in the NEXT profile, it seems that at least the followings need to be taken care of;

- some constraints or rules should be introduced for picture skipping in the lower layer and VBV regulation in the upper layer for VIDEO.

- a complete description should be prepared for the way of identifying the individual layers for SYSTEM.

END