ISO/IEC JTC1/SC29/WG11
MPEG92/470
September 1992

CCITT SGXV
Working Party XV/1
Experts Group for ATM Video Coding

Document AVC-330
September 21, 1992

SOURCE : JAPAN
TITLE : BUFFERING FOR LOW DELAY MODE
Purpose : Proposal

----

1. Introduction

This contribution analyzes the buffering problem for the low delay mode where occasional picture dropping takes place to cope with scene changes or other forced update situations. Proposals are made for methods to handle the headers of dropped pictures and Temporal Reference (TR). An extension of the existing VBV specifications is also proposed.

Note: A picture can be a frame or field according to the coding structure.

2. Problems

During a low delay mode operation, there may come a picture which requires a large number of bits to encode it, thus accompanies a large buffering delay. In this situation, it is sensible to momentarily sacrifice the temporal reproduction by allowing picture dropping and to keep the low delay performance for stationary states.

Since MPEG-1 assumes all pictures are encoded and displayed, some modifications of its specifications are necessary taking into account the following items relevant to picture dropping;

- How should the decoder operate?
- How should the encoder control bit generation?
- Should the headers for dropped pictures be sent?
- How should TR be numbered?
- How should VBV specifications be modified?

Though the existing MPEG-1 scheme substantially allows picture dropping by assigning a minimum number of bits for a picture, another picture with a large number of bits causes transmission delay as given by

(Bm-Ba)/Ba x Tf

where
Bm is maximum bit generation per picture,
Ba is average bit generation per picture,
Tf is picture time.

It is essential for achieving low delay not to take time for displaying dropped pictures.

3. Decoding Operation

The following decoder is assumed to simplify the analysis;

1) The decoder and the encoder have the same video clock frequency as well as the same picture rate, and are operated synchronously [current VBV specification]. Some means are provided externally to achieve this synchronism, e.g. by using sampling clock information contained in the picture header, AAL, time stamp.

2) It has a receiving buffer of size B, which is given in the vbv_buffer_size field in the sequence header [current VBV specification].

3) It receives coded data at a constant rate and write in the buffer [current VBV specification].

4) The buffer is initially empty [current VBV specification].

5) Decoding starts after filling the buffer for the time specified by the vbv_delay field in the video stream [current VBV specification].

6) At decoding timing with the same interval as that of encoder, all of the data for the non-dropped picture which has been in the buffer longest is instantaneously removed, instantaneously decoded and starts to be displayed [current VBV specification]..

7) If there is no complete data for a coded picture, there takes place no decoding operation and the most recent decoded field is displayed repeatedly [TM2 specification].

4. Coding Control

4.1 Model for encoding

Two coding control examples, M=1 frame coding and M=2 field coding, are given to show that there exist methods to meet the above mentioned decoder requirements.

This encoder operates as follows;

1) It has a transmission buffer of size B. This means that the maximum number of bits per picture should be B.

2) The buffer is empty when encoding starts.

3) Just after a complete picture is input to the encoder, it instantaneously encodes the picture and instantaneously sends coded data to the buffer.

4) The buffer content is read out at a constant rate.

5) A buffer occupancy value Bl (less than B) is set corresponding to the required buffering delay time (Bl/R where R is transmission rate). For stationary signals, the coding operation is controlled so that the buffer occupancy stays between zero and B_l at any time.

6) If a scene change or forced updating takes place, the buffer occupancy exceeds Bl, and necessary number of pictures are dropped.

The encoder controls the bit generation for a picture by referring to the buffer occupancy just before the encoding, and the estimated buffer occupancy just before encoding the next picture. If there is no picture dropping, the low delay mode encoder controls so that no overflow nor underflow of the buffer with size Bl should take place.

4.2 Example for M=1 frame coding

Figure 1 shows an example for M=1 frame coding with Bl corresponding to 1.5 frame time delay. Until frame F, low delay operation is maintained and G is the first frame of a new scene. Three frames H,I,J are dropped and low delay operation is resumed from frame K. In order to do so, the number of bits for frame G should be regulated to fall in the zone between 0 and Bl-R/P (R: transmission rate, P: picture rate), filling dummy bits if necessary, just before starting to encode the next frame. Figure 2 shows such regulation.

4.3 Example for M=2 field coding

Figure 3 shows an example for M=2 field coding with Bl corresponding to 2 field time delay. In this case, the display order is different from the encoding order. Hence there may be two strategies; to drop fields in the encoding order or in the display order. Figure 3 illustrates the two methods, which will give different subjective impressions.

When fields are dropped in the encoding order, such as p6,b3,b4,p9, repetition of the most previously decoded field happens at maximum three places. Encoding of b7 and b8 should use only p5 and p10, or p5 only in the worst case where p10 is also dropped, thus the reproduced pictures will degrade. When fields are dropped in the display order, such as b3,b4,p6,b7,b8, repetition of the previously coded field happens at maximum two places. In this case, encoding of p9,p10 is carried out by anticipating that b7,b8 are to be dropped.

In either case, low delay operation is resumed from field p13.

Since subjective impression is involved, the two methods should be experimented. If there is no significant difference in impression, field dropping in the encoding order is more straightforward.

5. Header for dropped picture

In order to keep low delay except for transient periods, the decoder should identify dropped pictures so that it does not take time to display them.

There may be two alternative solutions for dropped pictures;

- to send only their picture headers
- to send no data at all

If we send all the picture headers regardless of being dropped or not, the header of dropped picture should contain some explicit indication to ease distinction between the case of picture dropping and the case where a picture is coded with almost no coded data.

Our preference is the first choice.

6. Numbering of Temporal Reference (TR)

H.261 defines TR as indicating the CIF source picture number while MPEG1 defines it as indicating the display order. Use of this TR information is not specified in H.261, leaving it to implementation. It may be used for exactly reproducing the interval between coded pictures. Since all pictures are transmitted in MPEG-1, the display order becomes equivalent to the source picture order.

The following three alternatives need consideration for the case illustrated in Figure 1;

Source picture	A	B	C	D	E	F	G	H	I	J	K	L	M	N
Coded picture	A	B	C	D	E	F	G	*	*	*	K	L	M	N
Source picture TR	1	2	3	4	5	6	7				11	12	13	14
Display timing TR	1	2	3	4	5	6	10				11	12	13	14
Display order TR	1	2	3	4	5	6	7				8	9	10	11

The third method has been suggested in the TM2 experiment description.

Since the decoder described in Section 3 does not need TR for its operation, the choice should be from other considerations. It is proposed to use the straightforward source picture TR method.

7. Modification of VBV specifications

Reflecting the above discussion, we propose to modify the VBV specifications as in Annex.

8. Conclusion

Impacts of picture dropping in the low delay mode have been analyzed. Since it will contribute to achieve low delay during the steady state, picture dropping should be allowed in the standard. As conclusion, proposals have been made for handling of header for dropped pictures, TR numbering and VBV specifications.

The current discussion has been for the CBR case. Picture dropping in VBR environments and the VBV specifications need further study.

END

Annex Proposed modifications to the VBV specifications

Preamble

It is a requirement of the encoder (or editor) that the bit stream it produces will not cause the VBV to either overflow of underflow.

==> It is a requirement of the encoder (or editor) that the bit stream it produces will not cause the VBV to either overflow of underflow. *If low buffering delay is intended with allowing picture dropping, the buffer occupancy just after decoding a picture shall further conform to Item 5 below.*

Item 2

{For low delay operation, the buffer size B is interpreted as corresponding to the maximum number of bits per picture generated at the encoder when picture dropping takes place}

Item 4

==> add at the beginning "*This item applies to cases that all pictures are coded and transmitted.*"

Item 5 new item

This item applies to cases that some pictures are not coded nor transmitted.

Encoder may wish to realize low buffering delay with allowing occasional picture droppings. It will regulate its information generation by setting a virtual buffer size Bl smaller than B for stationary pictures. The VBV operates as follows;

All of the data for the non-dropped picture which has been in the buffer longest is instantaneously removed. Then after each interval all data of the non-dropped picture which (at that time) has been in the buffer longest is instantaneously removed. Sequence header, group of picture layer data elements and headers of dropped pictures which immediately precede a picture are removed at the same time as that picture. At some decoding timing where picture dropping takes place in the coder, there will be no sufficient data to remove. In that case, no data removing takes place.

During the stationary state with low buffering delay, the VBV occupancy immediately after each picture is removed shall lie between zero and (Bl - R/P).

To meet this requirement the number of bits dc for the coded picture just before the steady state (including any preceding header and group of picture layer data elements) must satisfy;

no dropped picture	$Bp - Bl + R/P < dc < Bp$
one dropped picture	$Bp - Bl + 2R/P < dc < Bp + R/P$
two dropped pictures	$Bp - Bl + 3R/P < dc < Bp + 2R/P$
n dropped pictures	$Bp - Bl + (n+1)R/P < dc < Bp + nR/P$

where:

Bp: VBV buffer occupancy just before removing the data
Bl: VBV buffer corresponding to low delay operation
R : bit rate
P : picture rate

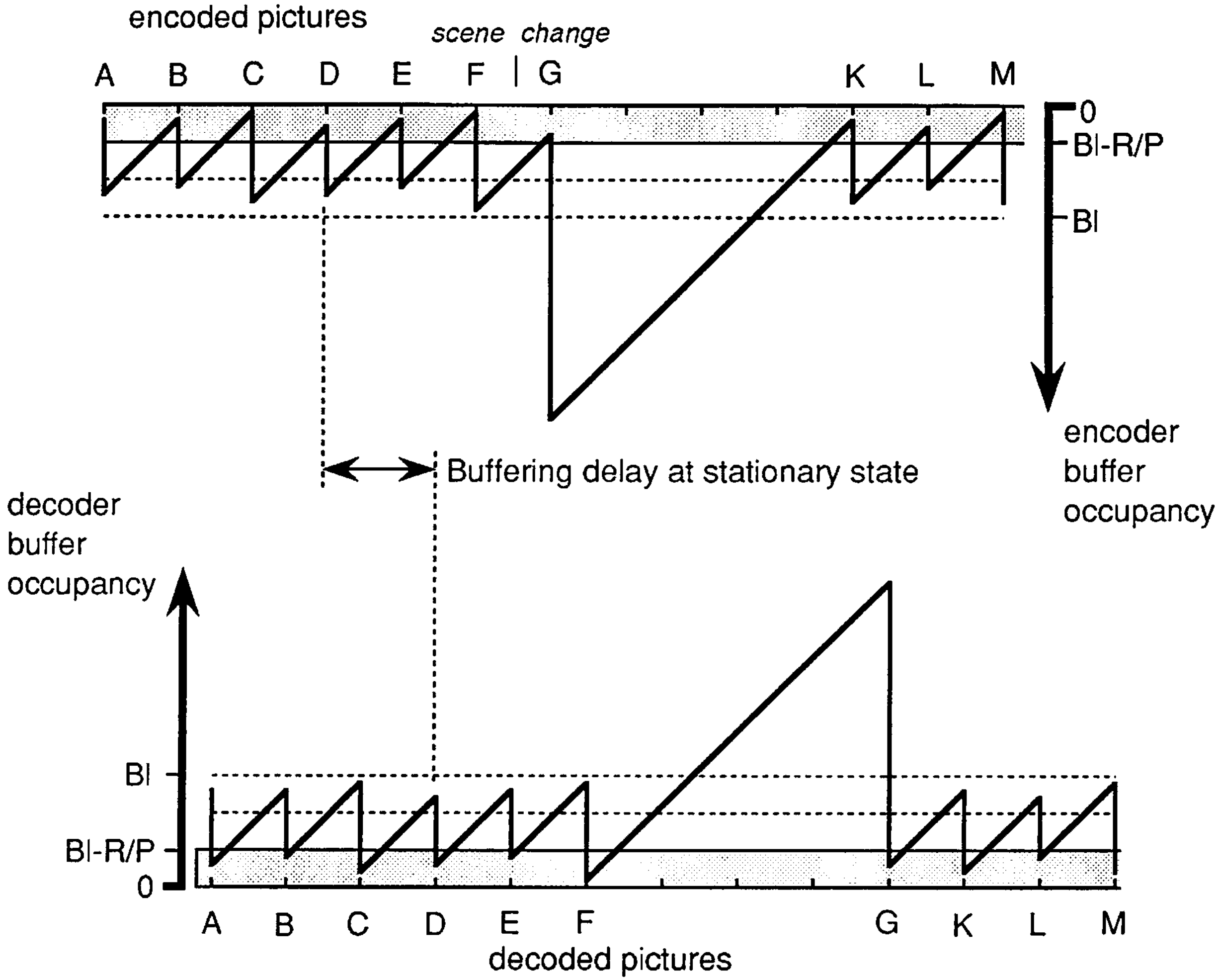
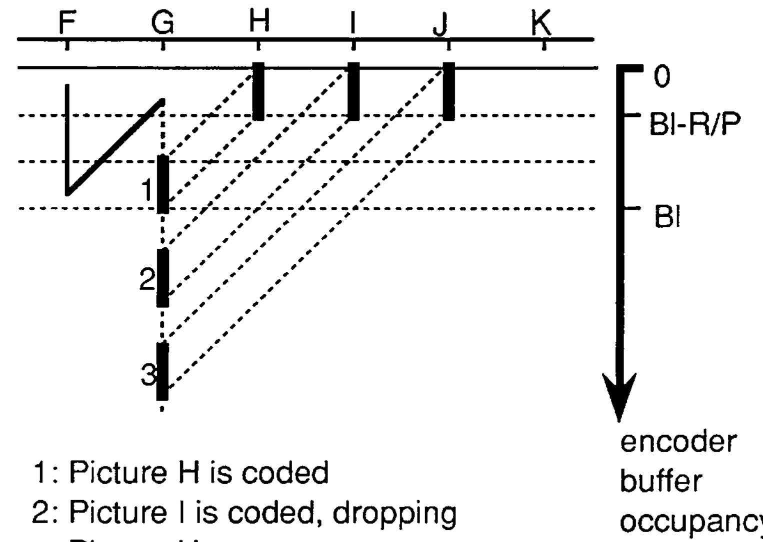


Figure 1 Decoder and encoder buffers for M=1 frame coding



1: Picture H is coded
2: Picture I is coded, dropping Picture H
3: Picture J is coded, dropping Pictures H and I

Figure 2 Information generation of Picture G

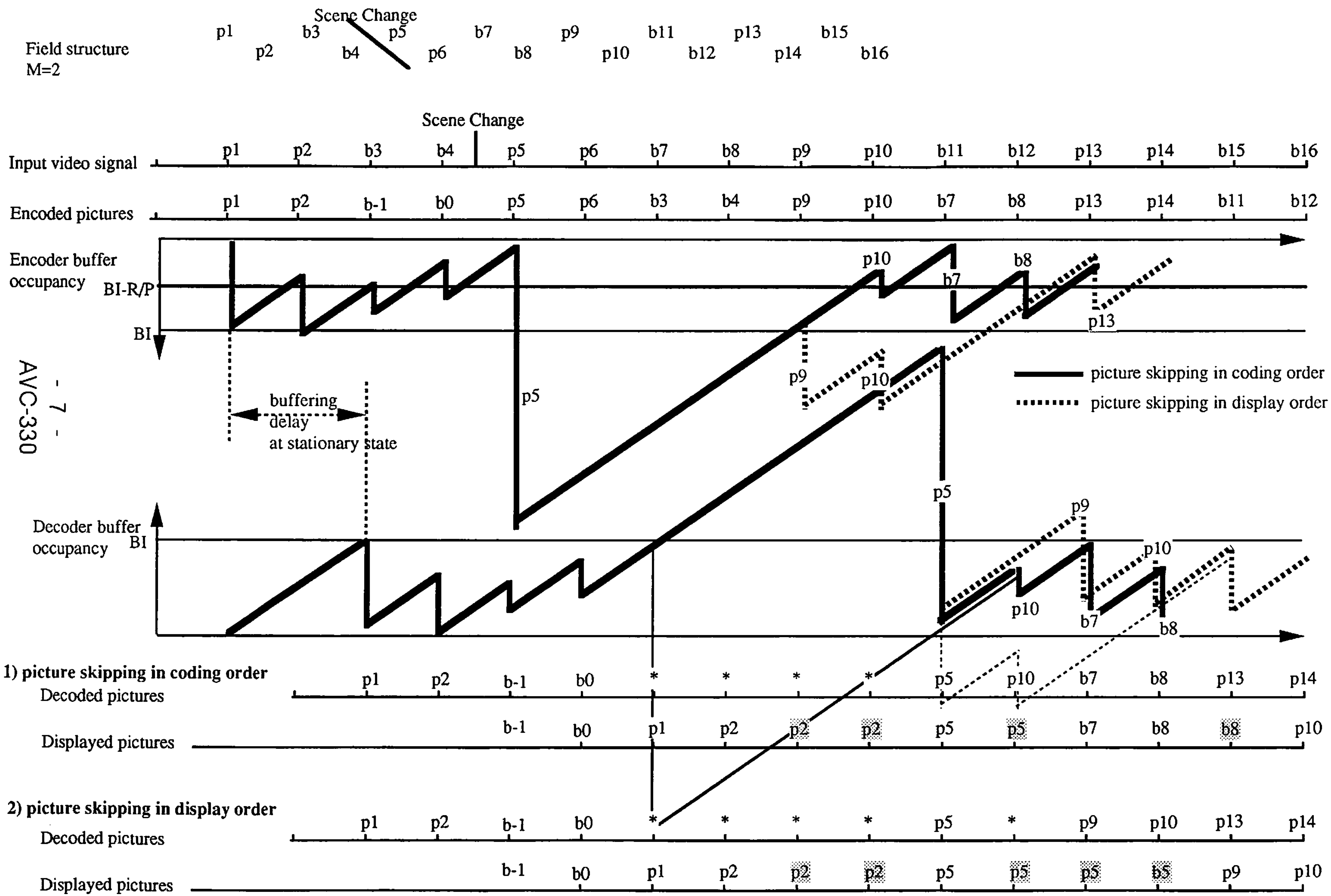


Figure 3 Decoder and encoder buffers for M=2 field coding