CCITT SGXV
Working Party XV/1
Specialists Group on Coding for Visual Telephony

Document #104
May 1, 1986

SOURCE: CHAIRMAN OF THE SPECIALISTS GROUP ON CODING FOR VISUAL TELEPHONY

TITLE : REVISED SPECIFICATIONS FOR THE REFERENCE MODEL

At the last Tokyo meeting, a reference model was agreed upon to further investigate the quantization strategy for the n x 384 kbit/s codec (see Annex 2 to Document #103R). It has been found from the preliminary implementations of the model, however, that some clarifications and improvements to the model are required. Consequently, here are proposed specifications for a modified reference model based on the comments provided by those colleagues in Europe and Japan.

Since the reference model is only a tool for the algorithm study, all the members are requested of their efforts to find an efficient but simple and elegant solution by utilizing the reference model. More specifically, not only a total proposal but also the most decisive elements improving the performance should be clarified compared with the reference model, so that all the members can easily understand the best possible combination.

---

1) The reference simulation model has a configuration as shown in Figure 1.

2) The same sequences at the Common Intermediate Format must be used: 'Miss America' and 'Graphics' have already been distributed; 'Checked Jacket' will be provided by Japan; 'Split-Trevor' will be provided by Europe.

3) Number of coded pels per line is reduced to 352 by discarding the first 8 pixels.

4) Classical DCT (a) of Temporary Document No. 3.

5) When we count the information rate, only number of bits necessary for transmitting block attributes and block data are calculated. Those bits necessary for picture attributes and group of block attributes are neglected, because they are only several tens bits and small enough compared with the average number of bits per picture (20 or 30 kbits corresponding to 2:1 or 3:1 temporal subsampling) or per group of block (1111 or 1666 bits).

6) Block types are defined as shown in Figure 2. Attribute #1 (Fixed/Non fixed) addresses to conditional replenishment, which is not used in the reference model.

7) Number of bits per block required to transmit each block type are shown in Table 1, where the block attribute 'fpvt' to identify the block type is represented with 2 bits.

Note: Original specification was 3 bits per block meaning 7128 bits per picture. As this overhead is too heavy, and introduction of VLC to distinguish the five different Block Types can reduce the wordlength to 2 bits per block, an average of 2 bits has been assigned.

8) Coded/Non coded Attribute value is determined after finding out whether the block includes one or more transform coefficients which are outside the range of - 1.5g to + 1.5g.

9) Buffer size and buffer state is written in kbits. 'k' represent 1000 in this case.

10) Block data is logically represented as follows (see Figure 3).

[Data] = q = [index(0,0)][index(0,1)][index(1,0)][index(2,0)] [index(1,1)]...[EOB]

Fig. 3

(0,0)	(0,1)	(0,2)	---	(0,7)
(1,0)	(1,1)	(1,2)	---	(1,7)
---	---	---	---	---
(7,0)	(7,1)	(7,2)	---	(7,7)

11) Diagonal scanning: see Annex 1.

12) Quantizer, VLC and EOB: see Annex 2.

13) For the DC component of intra-frame coding, 11 bit linear quantization is used.

Note: Experience shows that fine quantization of the DC component plays an important role in getting good pictures.

14) Inter/intra modes and how to handle a scene cut: see Annex 3. It should be noted that the same technique is applied for detecting a scene cut as distinguishing inter/intra modes. As for the first picture of the test sequence, we can use an apriori knowledge. This will be allowed because the situation can be seen as an initial condition of equipment power switch on.

15) The following temporal subsampling to CIF is used.

Miss America	2:1
Checked Jacket	2:1
Split-Trevor	3:1
Graphics	3:1

16) Buffer size: a size of 30 kbits is intended. However, for the reference model simulation purpose, the buffer fullness may exceed 30 kbits to avoid special anti-overflow measures (see the next item).

Note: It is expected that the present improvements of the model would not make the buffer fullness go over 30 kbits.

17) Buffer control: see Annex 4. It should be noted that coding is continued with g = 32 even when the buffer fullness reached 30 kbits.

18) Buffer fullness simulation: Since quantizer step size is controlled on a block-by-block basis, the buffer fullness B(n) is calculated at every block as follows.

$$B(n) = B(n-1) + I(n) - A/2376$$

where
- B(n) : Buffer fullness at the end of the n-th block
- B(n-1): Buffer fullness at the end of the (n-1)-th block
- I(n) : Number of bits spent to transmit the n-th block (cf Table 1)
- A : Average number of bits transmitted per picture
  - = 40000 for 2:1 temporal subsampling, scene cut
  - = 20000 for 2:1 temporal subsampling, other pictures
  - = 60000 for 3:1 temporal subsampling, scene cut
  - = 30000 for 3:1 temporal subsampling, other pictures

It should be noted that B(n) be calculated using 'real' type variable.

Note: If we would simulate the hardware codec more exactly, above value of A is not correct. For example, in case of '2:1 temporal subsampling, other pictures', it should be 10000 and we have to subtract 10000 at the end of each coded picture. However, for our reference model purpose, this strictness will not be necessary.

19) Motion estimation is achieved by block matching technique with a window search of $\pm$ 7.

- Integer displacement only.
- Coarse-fine 3 steps algorithm (see Annex 5) is used.

  Stage 1: Search at 0 and $\pm$ 4 pels.
  Stage 2: New center at minimum error position of stage 1. Search at 0 and $\pm$ 2 pels (referring to the new center).
  Stage 3: New center at minimum error position of stage 2. Search at 0 and $\pm$ 1 pels (referring to the new center).

  Search area is restricted to the inside of the 352 pels x 288 lines even when the block is located at the edge of the picture.

- Non-zero motion vectors are transmitted by using an 8 bit FLC.

20) No motion vector for chrominance.

21) Motion vectors are forced to zero if

$$\sum_{0}^{7}\sum_{0}^{7} \left| \text{frame difference} \right| < 64 \times 3.$$

Note: It was observed in the preliminary implementation that a lot of motion vectors are generated in the still backgrounds due

to noise contained in the test sequence, needing 7 - 10 kbits per frame. Examples of the scheme to suppress unnecessary motion vectors are found in Documents #90 and 91. One of the simplest schemes is employed for the reference model.

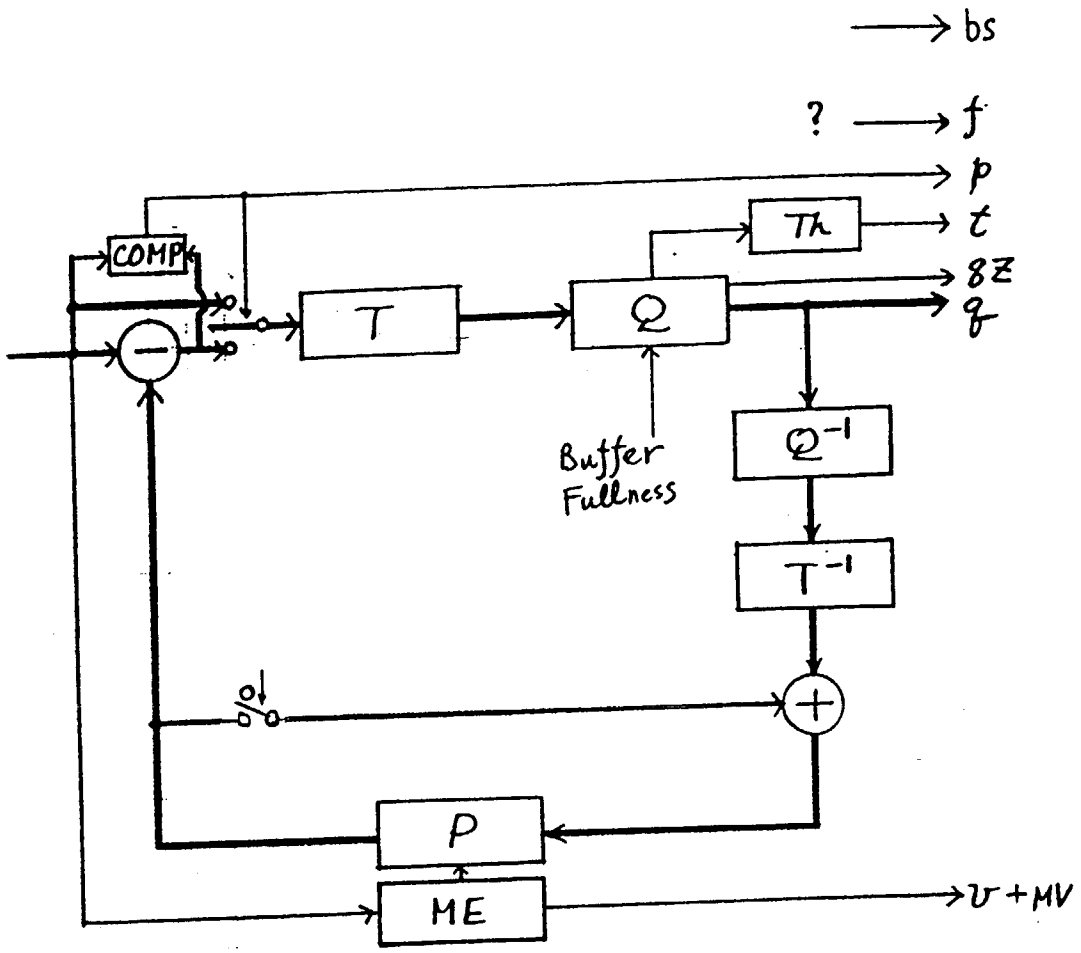
22) Evaluation function for motion estimation is a sum of absolute difference concerning to all of the 64 pels in a block.

23) No pre-processing, no post-processing is allowed.

24) No filtering in the loop is allowed.

25) Presentation of new algorithms or parameters: use the window method proposed in Annex 6. It is suggested, in order to minimize the amount of sequences to be processed, to use the most critical sequence for the comparison between the Reference and the improvement on the Reference. The most critical sequence may depend on the type of processing. The ultimate results should be shown, however, by presenting again all the sequences.

End



bs : Buffer State
f : Fixed / Non Fixed (Attribute ≠ 1 / Doc. #69)
v : No MC / MC (Attribute ≠ 2 / Doc. #69)
t : Non coded / Coded (Attribute ≠ 3 / Doc. #69)
p : Intra / Inter (Attribute ≠ 4 / Doc. #69)
qz : Quantizer indication
q : Quantizing index + EOB
MV: Motion Vector

Fig. 1 REFERENCE MODEL

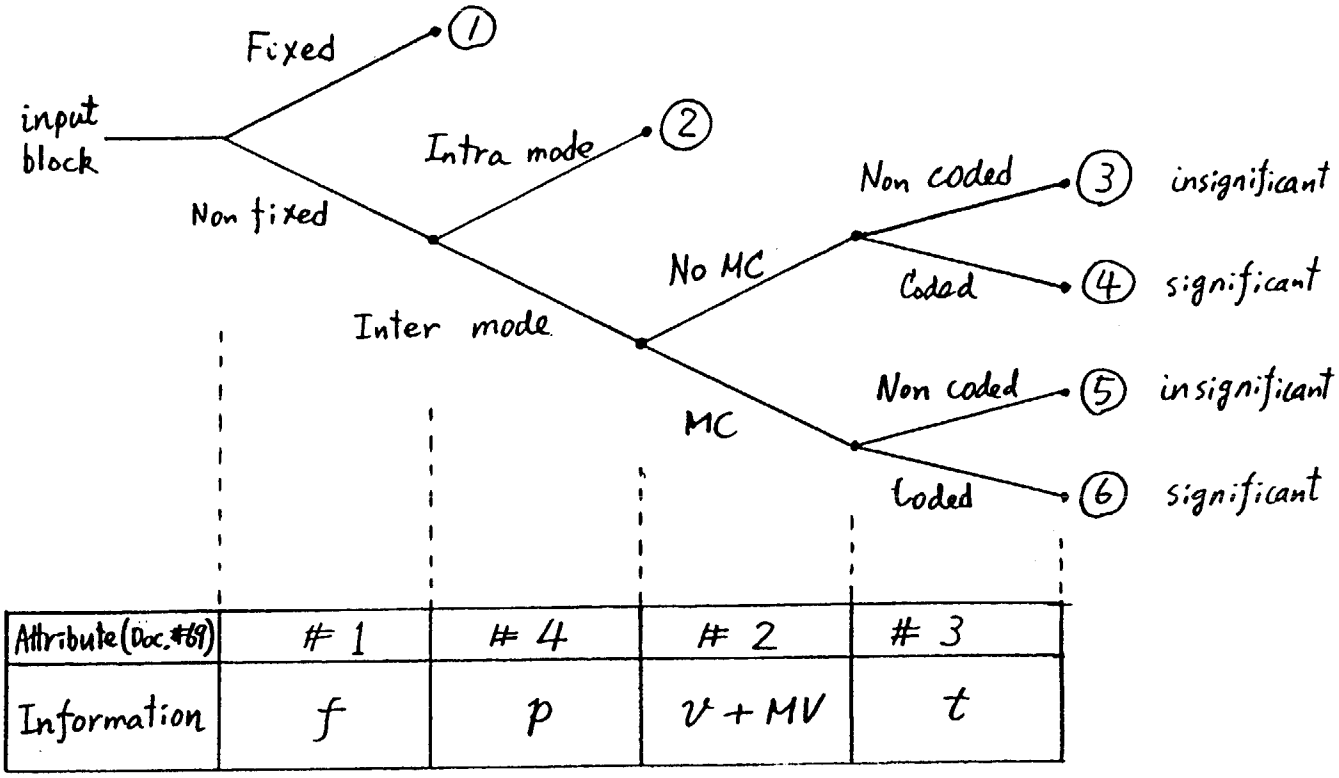


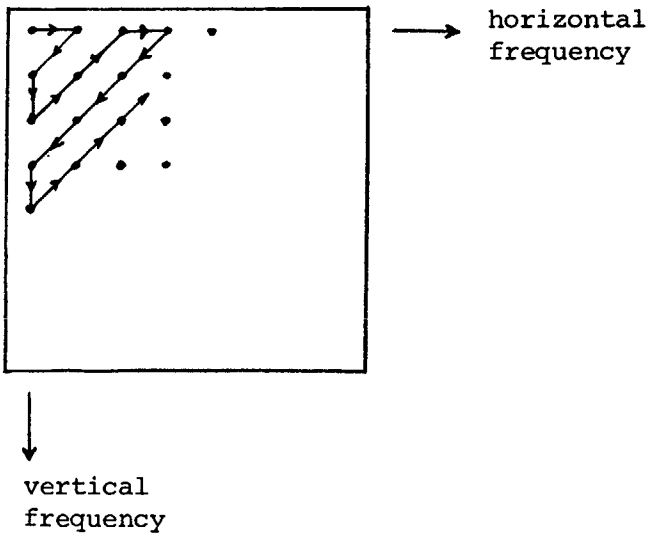
Fig. 2 Block type

Table 1 Number of bits per block

Block Type	Block Attribute f, p, v, t	Quantizer	MC vector	Quantization Index	Total	Note
①	2	0	0	0	2	Not used in the ref. model
②	2	3	0	[Data]	5 + [Data]	
③	2	0	0	0	2	
④	2	3	0	[Data]	5 + [Data]	
⑤	2	0	8	0	10	
⑥	2	3	8	[Data]	13 + [Data]	

Annex 1: Diagonal Scanning

The transform coefficients are transmitted in the following way:



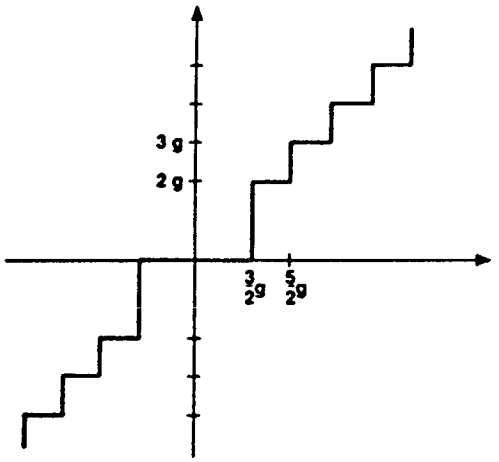
That means:

DC, first horizontal frequency, first vertical frequency, ...

Horizontal frequency corresponds to vertical lines in the picture.

Annex 2: Quantizer, VLC and EOB

Uniform quantizer is defined by the step size g. The quantizer has a threshold with value 1.5g.



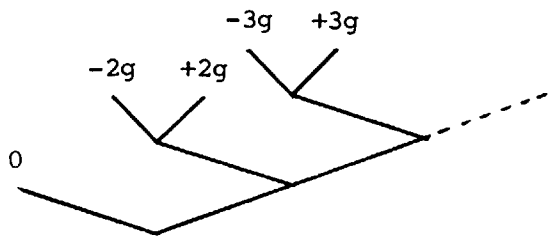
Example: A transform coefficient F

$$1.5g \leq F < 2.5g$$

is quantized to the value $\hat{F} = 2g$.

The dynamic range is + 2048. Step size is related to uniform transform, alternative (a)/Temporary Document No. 3.

The same quantizer is used in both inter- and intra-frame coding and also in both luminance and chrominance coding. Quantized transform components are coded with the following VLC. An exception is in DC components of intra-frame mode, which are 11 bit linearly quantized and coded with an 11 bit FLC.



0 → 0
-2g → 100
+2g → 101
-3g → 1100
+3g → 1101
.
.
.

EOB (End Of Block) is transmitted by indicating the number of non-zero components in each block with the following VLC.

1 non-zero	1 bit
2 non-zeroes	2 bits
3 non-zeroes	3 bits
.	
.	
.	

Annex 3: Inter/intra Mode Switch and Scene Cut

1. Inter/intra-frame switch

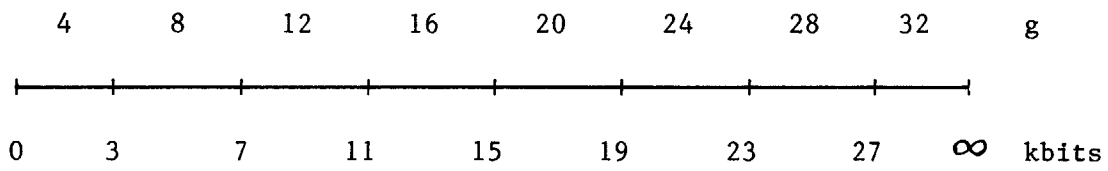
   a) Block based.
   b) The block values are first subtracted from a median grey level (level 128) and then squared. If the resulting value is greater than the prediction error squared of the corresponding block, then interframe mode is used.

2. Scene cut

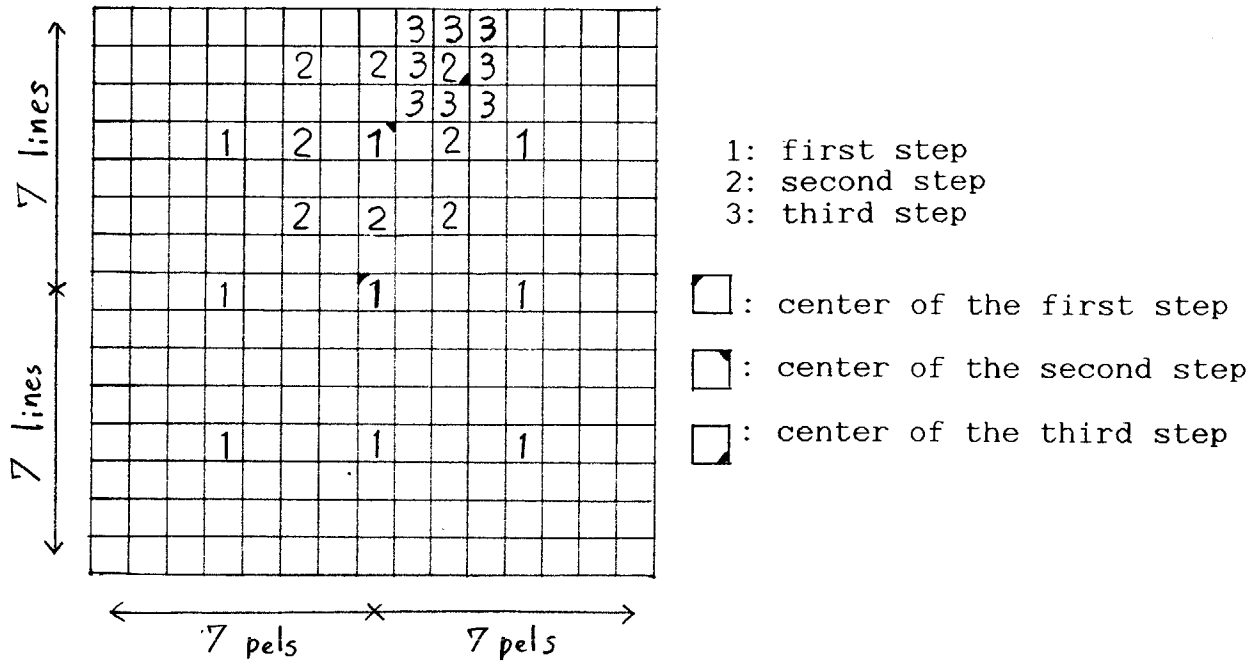
   A scene cut is detected by the same technique as above for the intra/inter mode decision, but the comparison should be carried out over the entire picture. In the case of scene cut, intraframe mode is used. To overcome the problem of a buffer overflow, the previous picture should be repeated, that means, for the first picture after the scene cut, the amount of bits available is twice compared with the average amount of bits per picture.

Annex 4: Buffer Control

Quantizer step size g is chosen for each block, and indicated with 3 bits. Step size is chosen as a function of buffer content:



Annex 5: Motion Estimation



Coarse-fine 3 steps algorithm

Annex 6: Presentation of Improvements

Take a 180x288 window from the "Reference Model"-result and take the same window from the improved result; put them next next to each other (so the results can be displayed together).
For the Split/Trevor-sequence and the Checked Jacket-sequence we propose to position the window at (40,0):

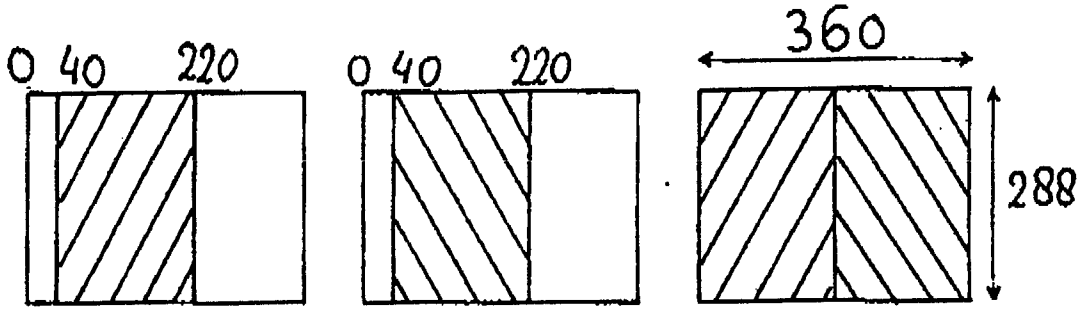


fig. 1 reference model result | improvement | refer. improv. DISPLAY

For the Miss America-sequence we propose to position the window at (90,0):

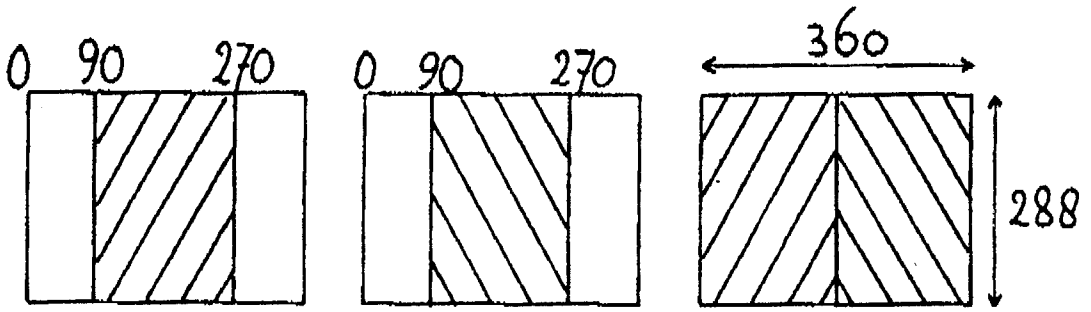


fig. 2 reference model result | improvement | refer. improv. DISPLAY

Note: In figure 1 and 2, the column 0 corresponds to the first column of the modified (352 pels x 288 lines) CIF.