ISO-IEC/JTC1/SC2/WG11
CODING OF MOVING PICTURES AND ASSOCIATED AUDIO

ISO-IEC/JTC1/SC2/WG11
MPEG92-

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

Document AVC-245
March 13, 1992

SOURCE: G. Bjøntegaard, Norwegian Telecom Research.
On behalf of the Eureka VADIS project.
TITLE : TM0 simulations and improvements.
PURPOSE: Discussion

1. Coding within TM0.

The full TM0 has not been implemented yet. A subset of TM0 is therefore used as a reference in this document. Furthermore, the "bitstream" decoder has not been completed yet. Still the simulation program has been based on "NTA Kurihama proposal" which performed bitstream decoding. I am therefore quite confident that there are no big errors in the simulation results presented here.

1.1. The subset of TM0.

- Only M=1 is implemented. The main purpose of the document is to identify methods for improving low delay coding (M = 1).
- The DCT is performed only frame based. One of the reasons for this is that field coding of chroma is still not clearly defined in TM0.
- The TM0 buffer control is not implemented. Instead there is the "NTA Kurihama" buffer control. However, the sequences used are the 50 first frames of MOBCAL, FLOWERGARDEN and TENNIS where buffer regulation is not critical.
- The mode selection is only between INTRA and INTER. Motion compensation is performed field based with two vectors pr. macro block (as defined in TM0). This mode seems to give better results than using frame based prediction only. Selection between the two would give a slight improvement.

field model

2. Changes relative to TM0

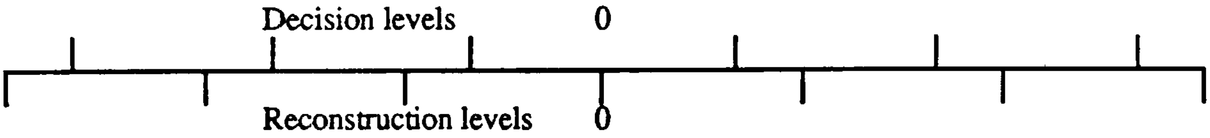
2.1. Quantizing transform coefficients (PLUS1).

In some of the European proposals for the Kurihama test (6 and 7) a quantizer different from the one defined in TM0 was used:

Haskell o precale
o temporal filtering

LEVEL = sign(COEFFICIENT)*(lCOEFFICIENTl + OFFSET)/(2*MQUANT)
OFFSET = (2*MQUANT)//3
RECONSTUCTED_COEFFICIENT = (2*MQUANT)*LEVEL

This last quantizer is better fitted to the statistics of the transform coefficients than the quantizer defined in TM0. It means that the reconstruction levels are equally spaced, whereas the decision levels in the quantizer are shifted away from zero as indicated in the figure.



In the present simulations, an adaptive version of the above quantizer was used. OFFSET was set to one of three values:

OFFSET_1 = 0.4*MQUANT,
OFFSET_3 = 0.6*MQUANT,
OFFSET_5 = 0.8*MQUANT,

The decision of which "OFFSET" to use is the same as used for chosing the VLC with 1, 3 or 5 bits for EOB in the adaptive VLC method (see section 2.4 and doc.MPEG91/206).
This same quantizer was used for INTRA and INTER blocks. It should be noted that the use of OFFSET in the quantizer affects only the encoder. The decoder will only know the LEVEL value and do the reconstruction accordingly.

2.2. Using decoded picture for half pixel motion search (PLUS2).

In TM0 all the motion vector search is performed relative to the original images. As a change to this, I have performed the half pixel motion estimation relative to the decoded picture. The main reason for doing this is the following:
- The half pixel accuracy motion vector is more important concerning filtering of the prediction than concerning the accuracy of displacement. To optimise the filtering, it is necessary to do the comparison to the same picture that will be used for prediction (locally decoded). This feature is named PLUS2.

2.3. Extension of the field prediction (PLUS3).

The bidirectional interpolation used for B-frame prediction implies using an average from two fields/frames for prediction. This predictor gives good results. Some of the same effect may be obtained for P-frame prediction. In this document I describe how this may be done by using the average of two fields in the previous decoded picture. The procedure is illustrated in the figure below.

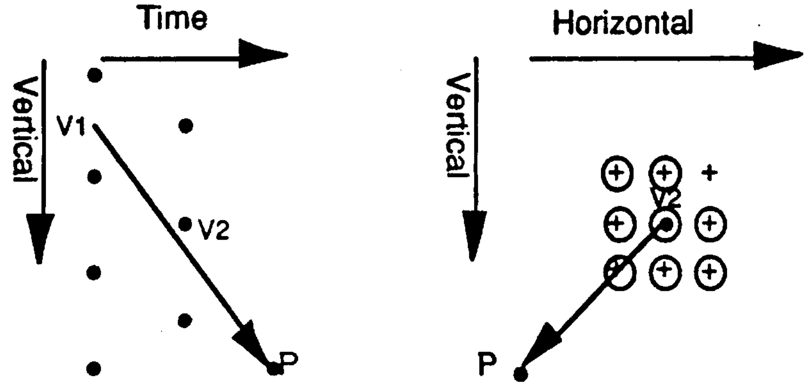
P is a point in the field to be predicted. After the 1/2 pixel search the vector V1 is found to be best. (V1 is field vector with half pixel resolution. The information on which field V1 refers to is available.) The following steps are taken to obtain a predictor based on two fields:

- Calculate the vector V2. This is the "corresponding vector" from the other field. It is obtained by linear scaling of the vector V1. Its value is rounded to the nearest 1/2 pixel value. The expression for V2 is:

V2 = NINT(V1•F)   where F depends on which fields P and V1 belong to. The values are given in the table below.

P	V1	F
1. field	1. field	0.5
1. field	2. field	2.0
2. field	1. field	0.67
2. field	2. field	1.5

- Around V2, 8 points are located as indicated by circles in the Horizontal/Vertical plot.
- The prediction can either be:
  - The block defined by V1 alone or
  - An average between the block defined by V1 and the block defined by one of the 8 vectors around V2.



+ Half pixel point

● Integer pixel point

○ Search points for extended prediction

The total signalling for a field vector is therefore:
- One half pixel vector as defined in TM0.
- 1 bit (B1) to tell if this vector is used alone (B1=0) or if a combination of two vectors are used (B1=1)
- If B1=1, additionally 3 bits is is sent to tell which of the 8 positions around V2 to use.

The extension implies an additional search of 8 positions. However the data would already be in the "calculation unit" so no additional data access to memory is needed.

The extension is named PLUS3. It implies using more bits on prediction. The benefit is a better predictor.

2.4. Use of adaptive VLCs for coefficient coding (PLUS4)

In the NTA Kurihama proposal, adaptive VLCs for coefficient coding was used. The same VLCs have been tried in the present simulations. To summarise the essence of this scheme:

- Three VLCs with 1, 3 or 5 bits for EOB was defined.
- For each coefficient to be coded, one of these VLCs was chosen. The choice was fully determined by the following 3 parameters:

- LEVEL of the last transmitted coefficient.
- RUN before the last transmitted coefficient.
- Number of coded coefficients in the last MB. For the first coefficient in a block this parameter defines the VLC alone.

The VLCs were developed for the conditions:
- 4.2.2 signal which means 4 luminance and 4 chrominance blocks in a MB.
- Coded Block Pattern was not used. It was therefore relatively many blocks without coefficients

In the present model 4.2.0 format and CBP are used. This means that the statistics of the coefficients to be coded becomes different. This should be taken into account and a new set of VLCs be developed. This has not been done so far.

2.5. Use of 3-D VLC for block scanning in coefficients (PLUS5).

In the paragraph above, the VLCs were adaptively chosen - mainly to get a VLC with the right number of bits used to code EOB.

Here is another approach to obtain the same "adaptivity". There are 3 rather than 2 entry value to the VLC:

- LEVEL
- RUN
- Whether the coefficient to be coded is the last in a block or not.

This implies:
- End Of Block (EOB) does not have to be coded since the VLC itself signals when there are no more coefficients to be coded.
- The code is "self adaptive" concerning use of bits to signal "last coefficient" - because "last coefficient" is an entry to the VLC.
- The use of this VLC requires that CBP (or a similar function) is used.
- There is no distinction between coding the "first" or "not first" coefficient in a block as it is in TM0.

The construction of the CODES are similar to the ones described in the NTA proposal of Kurihama. The number of bits for the different code words for the new VLC are as follows:

Bits for "NOT LAST COEFFICIENT".

```
            LEVEL
  3  5  6  7  8  8  8  9  9  10 10 11 11 12 12 .....
R 4  7  8  9 10 11 ....
U 5  8  9 10 11 12..
N 5  9 10 11 12..
  6 10 11 12..
  6 11 12..
  7 12..
  7 13..
  8 ..
  8 .
  .
```

Bits for "LAST COEFFICIENT".

```
                LEVEL
         5 10 11 12 13 14 ..
       R 6 11 12 13 14 ..
       U 7 12 13 14 ..
       N 7 13 14 15.
         8 14 ..
         8 15 ..
         9 .
         9 .
        10 .
        10 .
```

2.6. Use of 3-D VLC for frequency scanning in coefficients (PLUS6).

One of the advantages with the VLC defined above, is that the same VLC may be used for block scanning and frequency scanning. For the frequency scan, the procedure would be like this:

- "Last coefficient" still means last in a block.
- All the quantized values for a SLICE is saved for later coding. The coefficients are then coded by first coding all coefficients in the SLICE with a certain coefficient number
- The blocks with no coefficients (signalled with CBP) are omitted in scanning the first coefficient number.
- When a "last" coefficient is coded, that block is removed from the scanning to be used for the next coefficient number.
- When all the blocks has been "removed" as indicated above, there are no more coefficients to be transmitted.

When the VLC is used this way, it means that exactly the same number of coefficients and zeros are coded in the block scanning and frequency scanning. It would therefore also be expected that the performance of the two would be quite similar.

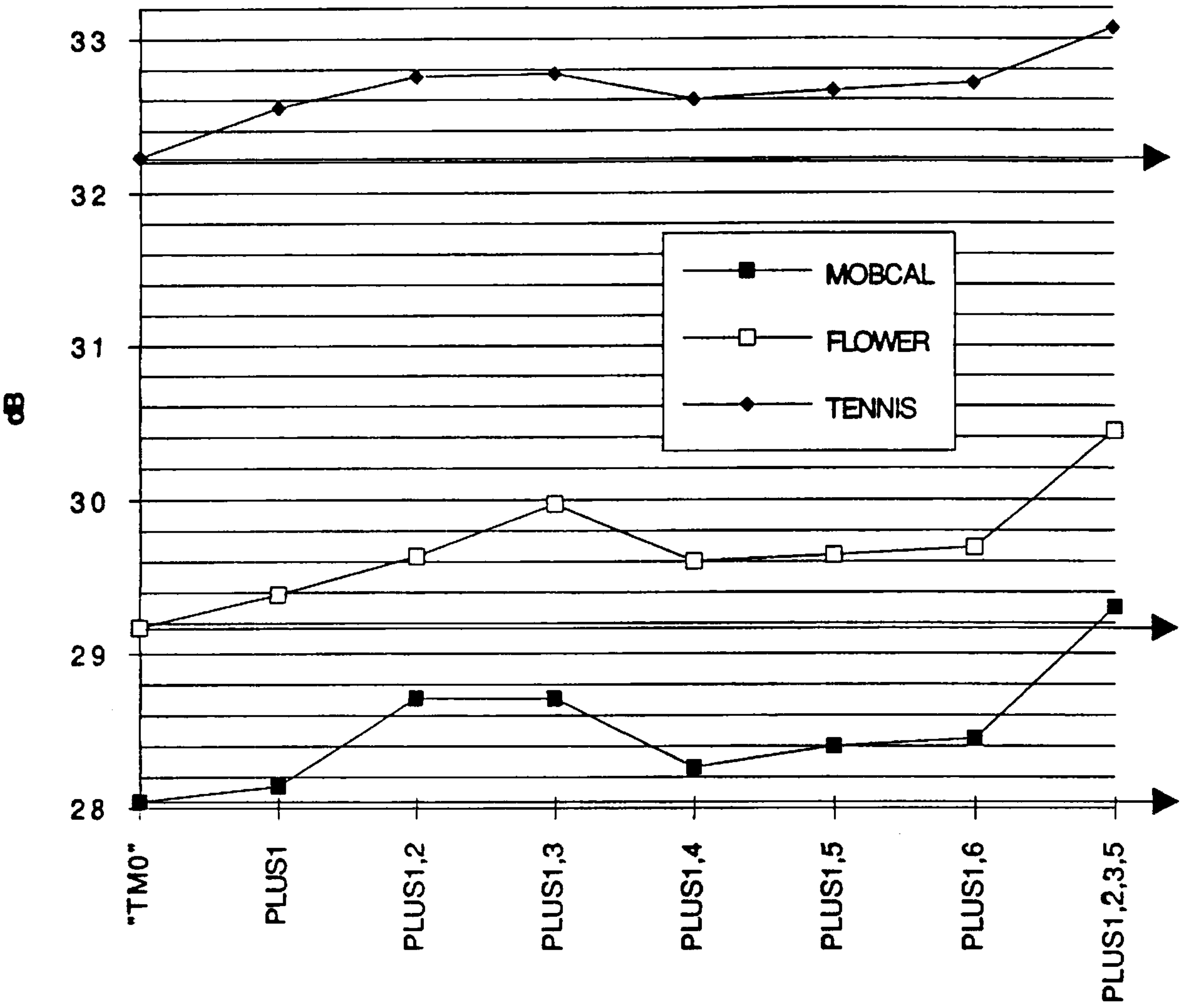
This kind of VLC and frequency scanning is well suited for layering of the bitstream. One would then add a code to the VLC table - End Of Layer (EOL)This would signal that there is no more coefficients connected to a particular layer. Since there will be many coefficients connected to each layer, the code word for EOL would be relatively long (5-7 bits?). The same VLC could therefore be used also in block scanning with negligible loss even if EOL would not be used in that case.

3. Simulation results.

Simulation have been performed on the 50 first pictures of MOBCAL, FLOWERGARDEN and TABLE TENNIS. The bitrate is 4 Mb/s. The SNR values for the luminance signal is given in the table below. The same information is displayed graphically in the figure below.

	"TMO"	+PLUS1	+PLUS1,2	+PLUS1,3	+PLUS1,4	+PLUS1,5	+PLUS1,6	+PLUS1,2,3,5
MOBCAL	28.04	28.14	28.71	28.71	28.26	28.40	28.45	29.30 (+1.26)
FLOWER	29.17	29.38	29.63	29.97	29.50	29.64	29.69	30.44 (+1.27)
TENNIS	32.23	32.55	32.75	32.77	32.60	32.66	32.71	33.07 (+0.84)

SNR gain for different improvements



4. Discussion of the results.

Some comments to the improvements relative to "TM0". When estimating the effect of small changes in SNR, it may be useful to bear in mind that: 0.1 dB ≈ 2% in bitrate. Furthermore, some of the improvements like PLUS 2 and PLUS3 relates only to INTER pictures. The SNR gain is referred to a model where 1/3 of the bits are used for INTRA coding. Referred to an application where less bits are used for INTRA coding, the coding gain with PLUS2 and PLUS 3 could therefore be up to 50% higher.

PLUS1:
The new quantizer gives a gain of (0.1-0.3) dB in SNR. When testing the other features, this new quantizer is included.

PLUS1,2:
Half pixel motion search relative to decoded picture show consistant improvement compared with motion search relative to original picture. When doing the 1/2 pixel search, remember that all 9 positions must be re-estimated - not 8 as in SM3. We must also recalculate the integer pixel position relative to decoded picture for the comparison to be correct.

PLUS1,3:
The defined extension of the prediction makes significant improvement.

PLUS1,4:
Use of the adaptive VLC from the NTA Kurihama proposal gives relatively little gain. The reason for this reduced effect is that TM0 is different from the coding method where the adaptive VLC was used. A new set of VLCs should be developed to see if the gain can be improved.

PLUS1,5:
The 3D VLC for coefficient coding shows a gain of typically 1/4 dB (5% in bitrate).

PLUS1,6:
The same VLC used on frequency scanning shows slightly higher gain than used on block scanning.

PLUS1,2,3,5:
Here all the additional features are included. It is seen that the gain in SNR over TM0 is (1.36 - 1.97) dB. This is very much. For comparison the differences in SNR between 4 MB and 9 MB in the NTA Kurihama proposal for the same sequences were: (3.17 - 4.17) dB.

5. Conclusions

- A quantizer like the one defined in PLUS1 gives some coding gain and may be used for INTRA as well as for INTER blocks. The encoder will decide to what extent "deadzones" or "adaptivity" will be used.
- The half pixel motion search relative to decoded picture gives coding gain. The way this is done is not going to be standardised. However, the test model should reflect the best possible picture quality.
- The prediction using combination of two fields gives significant improvement. This will give special benefit for coding without B-frames, but coding with B-frames would also benefit from it. It should therefore be made room in the test model to do more testing on this feature.
- The 3D VLC shows slight coding gain and is interesting because it may be used for block scanning as well as frequency scanning. Room for further test on this should be made in the test model.

6. Video demonstrations.

All three sequences coded with "TM0" are shown. Likewise, all three sequences coded with "TM0" + PLUS1 + PLUS2 + PLUS3 + PLUS5 are shown.

For comparison I have also included results from an extension of the NTA Kurihama proposal. The differences from what was presented in Kurihama are:

- One vector pr. field is used for prediction. The coding is as described in TM0.
- The prediction extension *PLUS3* is introduced.

The sequences coded are: MOBCAL (125 pictures), FLOWERGARDEN (125 pictures), and TENNIS (125 pictures). The bitrate is 4 Mb/s and all three sequences are decoded from bitstream. The improvements in SNR from the Kurihama results:

MOBCAL	0.93 dB
FLOWERGARDEN	1.40 dB
TABLE TENNIS	0.57 dB