ISO/IEC JTC1/SC29/WG11
MPEG 92/ 173
March 23, 1992

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

Document AVC-234
March 18, 1992

SOURCE: Japan
TITLE: Simulation results on compatible coding
PURPOSE: Information

1. Introduction

Two sets of simulations were carried out to study the possibility of compatible coding with prediction mode from base layer. One was carried out in RM8 based coding (fully frame based coding) with uncoded CIF base layer to investigate the upper bound of the improvement of coding efficiency. The other was carried out in TM0 framework (MPEG92/28,80,77) using coded SIF base layer to investigate the coding efficiency in actual coding environments.

2. Summary of the simulation

In the first simulation using uncoded CIF base layer, embedded coding and single layer coding were compared, up/down sampling filter was compared, and the change of macro block type distribution was investigated. (see annex 1)

In the second simulation using coded SIF base layer, comparison between embedded coding and simulcast and up/down sampling filter was included, and comparison between field merging and field skipping as a down sampling method to SIF was also carried out. After that, two methods were tried for further improvement. (see annex 2)

3. Conclusion

(1) It has been verified that the coding efficiency does not deteriorate in embedded coding if upconverted CIF or SIF pictures are carefully used as a prediction mode.
(2) In our trial, the difference between the embedded and simulcast in actual coding framework was 0.55 to 0.79 dB in SNR at 4 Mbps.
(3) In this method, the effect of prediction mode from base layer seems to be limited to such cases that the original picture has less high frequency component or has large motion.
(4) It seems to be necessary to improve prediction mode from base layer especially in P picture and B picture.

) embedded と simulcast table 2.
同じビットレート 1.15 MPEG1

annex 2 interpolate されているところ

base layer

ANNEX 1 Simulation experiment with CIF pictures as a base layer

1. Purpose

The following experiments aim to study upper bound of the coding efficiency improvement when the CIF picture is used as a reference other than previous 4:2:2 coded picture.

2. Simulation method

In this experiment, 4:2:2 picture was transformed to CIF and again transformed to 4:2:2 picture and used as a reference in one of the prediction mode. Here, the purpose is to know the upper bound of the coding efficiency improvement, so coding for CIF was not executed although the reference signal in the actual layered coding contains quantization noise.

Coding algorithm was RM8 based one and only the forward prediction mode was used. Rate control was not executed and rate-distortion characteristics were compared using fixed step sizes. In the CIF prediction mode, motion compensation was not applied. Sum of the square prediction error was used for macro block type decision. Search area of motion vector detection for the previous coded frame was ±14.5 (horizontal), ±10.5 (vertical).

filter 1

vertical decimation : filter as described in Doc. #55 (Sep. 1985) with modified coefficients
vertical interpolation : filter as described in Doc. #55
horizontal decimation : filter described in SM3
horizontal interpolation : repeat pixel value

filter 2

vertical decimation : same as filter 1
vertical interpolation : same as filter 1
horizontal decimation : 21 tap filter which has narrower pass-band and stronger stop-band attenuation than SM3 filter
horizontal interpolation : filter described in SM3

3 Simulation result

Figure 1 shows the obtained rate distortion curve. Figure 2 shows the selected macro block type when the fixed step size was 14. From these results, the following points were clarified.

1) Applying the new reference does not cause degradation of coding efficiency.
2) Prediction from base layer seem to contribute in the following cases.
/ Original picture contains little higher frequency component and CIF is close to the original.
/ Picture contains large motion and motion exceeds the search area from the previous picture. Scene change and uncovered background are contained in this category.
3) If the layered coding is applied,
/ INTRA macro block decreases to nearly none
/ MC-FIX macro block (contains the case of referring to the base layer) increases

in the selected macro block type chart.
4) Comparison between filter 1 and filter 2 as up and down sampling filters shows a big difference. Careful selection of up/down sampling filter seems to be necessary. If filter 1 is used, CIF contains aliasing noise and pixel value repeating will cause larger error to the reference.

5) Motion compensation for base layer (search range was ±2.5) was also tried, but in some cases, coding efficiency decreased. In that case, not only the motion vector information but also DCT coefficient information was increased. This means that a block with less squared prediction error sometimes generates more information after DCT. If the macro block type decision is executed based on the generated amount of information or its estimation, this situation will be improved. Motion compensation on the base layer reference picture may be effective if they are not co-located with the original picture.

4. Conclusion

Simulation was carried out to study the possibility of the layered coding which uses base layer picture as a candidate of prediction. This method does not lose coding efficiency comparing simulcast, but its effect seems to be limited to some cases such as the original picture has less high frequency component or has large motion.

Figure 1

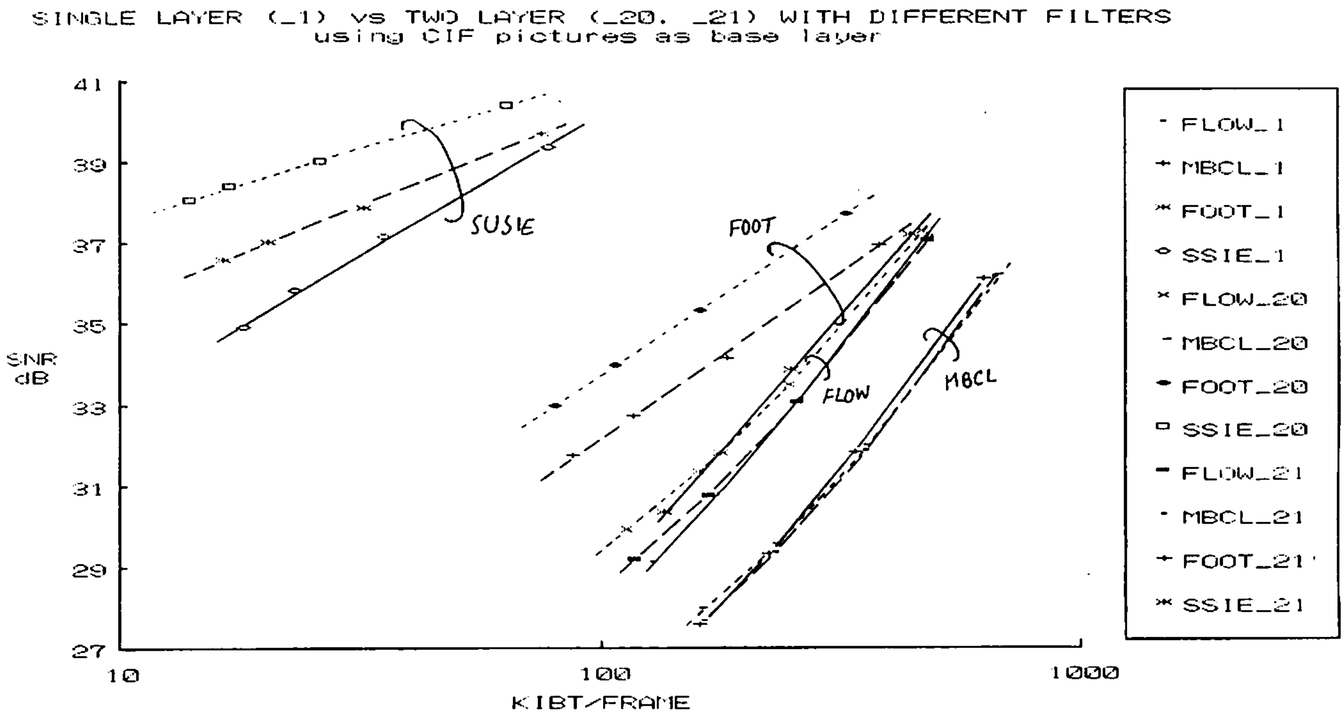
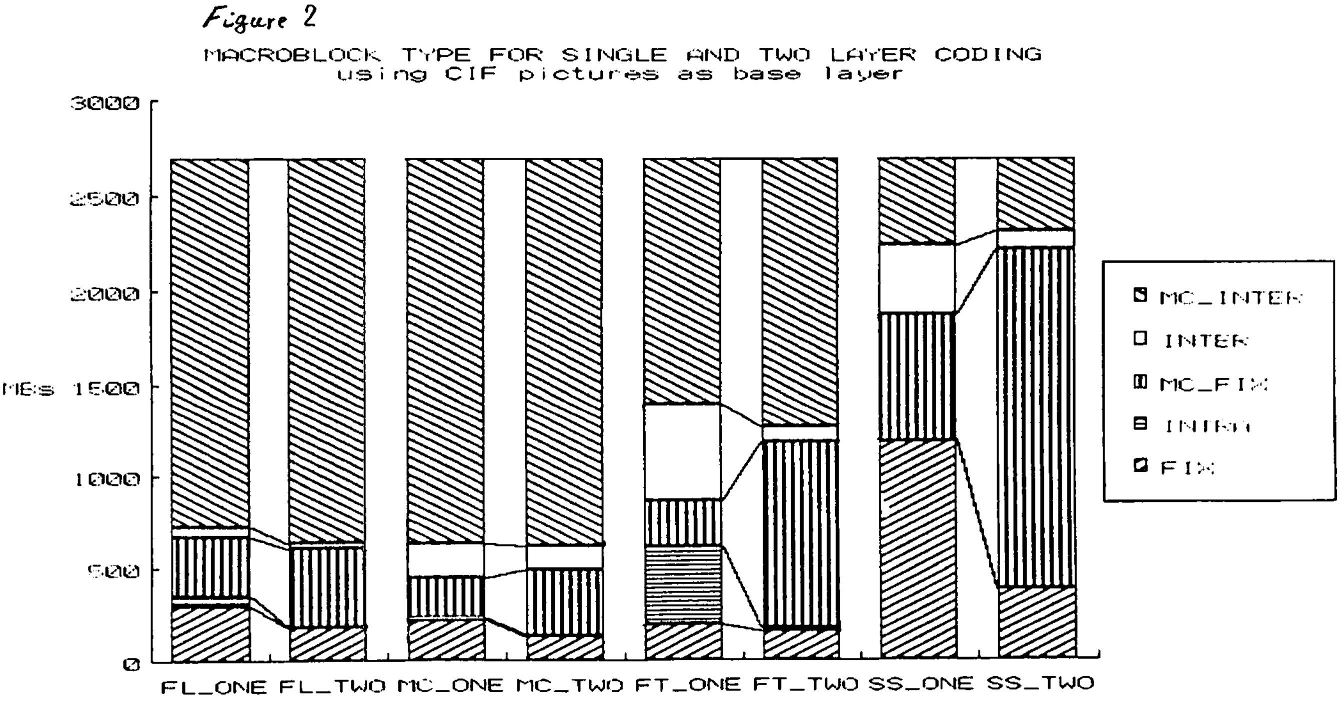


Figure 2



ANNEX 2 Simulation experiment with SIF coded pictures as a base layer

1. Introduction

Some simulations were carried out concerning compatible coding using the TM0 framework (MPEG92/28,80,77). Compatible coding was realized by appending the prediction mode from the up-sampled picture from the MPEG1 local decoding loop (MPEG92/76).

Simulations were carried out for the 4 Mbps total bit rate case (not for the 9 Mbps total bit rate case). The following 5 items are included in the simulation results.

(1) Comparison between embedded and simulcast
(2) Comparison among 3 down-sampling filters
(3) Comparison among 3 up-sampling filters
(4) Comparison between field merging and field skipping as a down-sampling method to SIF
(5) Some trials for improvement

All of the simulation results are given as mean SNR values calculated for 2 second sequences, and the SNRs are shown in the form of;
mean SNR for a total scene (mean SNR for I pictures : mean SNR for P pictures : mean SNR for B pictures).

The test sequences used for the simulation were " Mobile & Calendar ", " Flower Garden " and " Football ".

2. Embedded vs simulcast

Table 1 shows the simulation results for TM0 (4 Mbps), TM0 (2.85 Mbps = simulcast case), and TM0 with prediction from MPEG1 (= embedded case). In the embedded case, up- and down-samplings were carried out as described in SM3.

The results show that embedded coding gives 0.38 to 0.50 dB higher SNR values than simulcast, but it still gives 1.28 to 1.50 dB lower values than the non-compatible mode (TM0 4 Mbps).

In Table 1, the percentage of MPEG1 prediction mode selection is also included for each picture type. We can see from this result that prediction from MPEG1 is effective and that this prediction type is selected for almost all macro blocks in I pictures, but this selection type is not selected so much in P pictures and B pictures.

To see this situation more clearly, another simulation was carried out to compare embedded and simulcast. In this simulation, the GOP structure was changed to " I P P P ....." like RM8. Table 2 shows the result. The result shows that embedded coding and simulcast give nearly the same performance in this GOP structure.

Considering the fact that this GOP structure might be chosen for communication field application, some effort seems to be necessary to make the prediction mode from MPEG1 more effective in P and B pictures.

3. Down-sampling filter

When the bit rate for MPEG1 is limited to 1.15 Mbps, the quantization noise included in the MPEG1 locally decoded picture depends on the power contained in SIF.

Considering this point, the coding performance may be different according to the bandwidth of the down-sampling filter. In this simulation, three down-sampling filters were compared with each other. The first was the filter described in SM3. The second was the first one plus the same filter for the vertical direction. The third one was the same 7 tap filter for the vertical direction but a narrower 15 tap filter for the horizontal direction. The up-sampling filter was the 7 tap filter described in SM3 for all cases.

Table 3 shows the result. The narrowest filter gave the best performance for all the three sequences, but the difference between the best one and the worst one was about 0.1 dB and the difference according to the down-sampling filter was small in this trial.

4. Up-sampling filter

When the prediction mode from the locally decoded MPEG1 picture is selected, it is expected that the quantization noise can be controlled by an up-sample filter, because an up-sample filter works as a loop filter in H.261. In this simulation, three filters were compared. All of the three filters were a 7 tap filter and the passing bandwidth was different with each other. The widest passing-band filter was the one described in SM3. The down-sampling filter was the 7 tap filter described in SM3 for all cases.

Table 4 shows the result. The second narrowest filter gave the best performance for all the three sequences, but the difference between the best one and the worst one was about 0.1 dB and the difference according to the up-sampling filter was also small in this trial.

5. Field merging vs field skipping

In all the simulations above, SIF was made by selecting a field of one parity (skipping the field of the other parity), filtered, and down-sampled. So, the prediction from MPEG1 becomes difficult to match the skipped field. This seems to be one reason for the MPEG1 prediction mode not being selected so much in P and B pictures. Here, SIF was made by merging both parities of the field, filtered by the 7 tap filter described in SM3, and was compared with field skipping.

Table 5 shows the comparison between field merging and field skipping as a down-sampling method to SIF. This result shows that field merging gives 0.05 to 0.26 dB better SNR values. In this comparison, the quality for MPEG1 decoded picture should also be considered. If field merging is selected, there will be an annoying artifact at the edge of pictures with a large motion.

6. Other trials for improvement

Two methods were tried to make the prediction mode from MPEG1 more efficient or more likely to be selected for further improvement.

(1) Mode decision rule

In TM0, mode selection is done only by comparing the power of the predicted errors. But when comparison between the TM0 prediction mode and MPEG1 prediction mode is done, the motion vector information amount should also be considered because it is not necessary to send motion vectors in the MPEG1 prediction mode. A new mode decision rule was compared to the old one considering this point.

Table 6 shows the result. For the sequence " Football ", the new decision rule gave a 0.16 dB higher SNR, but the new decision rule gave 0.02 dB lower SNR for the other two sequences. Prediction from the MPEG1 mode was more likely to be selected in all cases.

(2) MPEG1 prediction mode for only odd fields

When SIF is made by field skipping, the skipped field will not be well predicted by the MPEG1 decoded picture especially for pictures with a large motion. This problem is considered to be solvable, for example, by applying the MPEG1 prediction mode to only odd fields.

Table 7 shows the result. A 0.41 dB gain was obtained for the sequence "Football", and a 0.24 dB gain was obtained for the sequence " Flower garden ". But a 0.07 dB lower SNR was obtained for the sequence " Mobile & Calendar ". But " Mobile & Calendar " is a very special picture with slow panning and a very high resolution. So, this technique seems effective for usual pictures .

7. Conclusion

The following were clarified as a conclusion..

(1) Comparison between embedded coding with SM3 and simulcast showed that embedded coding gave 0.38 to 0.50 dB better SNR than simulcast, but the prediction mode from MPEG1 was not so effective in P and B pictures. Considering that a low delay may be chosen in the communication field application, some effort to make the prediction mode from MPEG1 more effective seems necessary.
(2) Variations in the down-sampling filter and the up-sampling filter gave little effect to coding performance in the trials.
(3) The mode selection rule has some room to be improved. A mode selection rule considering the motion vector information amount seems to be effective especially for pictures with a large motion.
(4) When the SIF is made by field skipping, some special efforts seems necessary to predict the skipped field. Comparing field skipping and field merging as a down-sampling method to SIF, field skipping gives better performance if some special prediction is applied to the skipped field.
(5) According to the two improvements above, the prediction mode from MPEG1 became more likely to be selected.
(6) Comparing the best scores for embedded coding and simulcast, the differences were from 0.55 to 0.79 dB.

The trials in this document are not necessarily the best ones for each item, and further improvement can be expected.

END

The following are the simulation results. In the table, basic embedded coding means the TM0 based coding embedded with the SIF made by field skipping and down- and up-sampled by filters described in SM3.

Table 1 Comparison between embedded and simulcast

	MOB&CAL	Flower Garden	FOOTBALL
TM0 4Mbps (= non compatible)	28.28 dB (28.88:28.50:28.13)	29.86 dB (31.00:29.72:29.79)	35.71 dB (38.38:35.94:35.39)
Simulcast (2.85Mbps)	26.55 dB (26.85:26.51:26.53)	27.86 dB (26.61:27.39:27.95)	33.85 dB (35.94:33.80:33.67)
Basic embedded	27.00 dB (27.33:27.02:26.95)	28.36 dB (29.16:27.94:28.43)	34.23 dB (36.08:34.41:33.98)
MPEG1 pred. mode selection (%)	(92.94: 9.97: 1.42)	(89.51: 9.29: 1.67)	(70.97:27.47: 3.97)

Table 2 Comparison between embedded and simulcast in IPPPP... mode

	MOB&CAL	Flower Garden	FOOTBALL
Simulcast IPP.. mode	25.06 dB (25.38:25.06)	26.80 dB (29.10:26.77)	34.50 dB (41.35:34.44)
Embedded IPP.. mode	25.04 dB (26.56:25.01)	26.84 dB (30.45:26.80)	34.56 dB (41.36:34.53)

Table 3 Comparison among down-sampling filters

	MOB&CAL	Flower Garden	FOOTBALL
Basic embedded	27.00 dB (27.33:27.02:26.95)	28.36 dB (29.16:27.94:28.43)	34.23 dB (36.08:34.41:33.98)
up:SM3 down:SM3+v7tap	27.01 dB (27.37:27.03:26.96)	28.34 dB (29.17:27.94:28:40)	34.21 dB (36.08:34.39:33.96)
up:SM3 down:h15tap +v7tap	27.07 dB (27.49:27.10:27.00)	28.42 dB (29.36:28.03:28.47)	34.27 dB (36.14:34.47:34.01)

Table 4 Comparison among up-sampling filters

	MOB&CAL	Flower Garden	FOOTBALL
Basic embedded	27.00 dB (27.33:27.02:26.95)	28.36 dB (29.16:27.94:28.43)	34.23 dB (36.08:34.41:33.98)
down:SM3 up:narrower	27.10 dB (27.57:27.13:27.03)	28.46 dB (29.46:28.09:28.50)	34.35 dB (36.27:34.61:34.07)
down:SM3 up:narrowest	27.02 dB (27.45:27.05:26.96)	28.37 dB (29.30:27.98:28.41)	34.33 dB (36.24:34.59:34.05)

Table 5 Comparison between field skipping and field merging

	MOB&CAL	Flower Garden	FOOTBALL
Basic embedded (= field skip)	27.00 dB (27.33:27.02:26.95)	28.36 dB (29.16:27.94:28.43)	34.23 dB (36.08:34.41:33.98)
Field merge	27.05 dB (27.40:27.07:27.00)	28.54 dB (29.41:28.18:28:59)	34.49 dB (36.12:34.96:34.18)

Table 6 Comparison of mode decision rule

	MOB&CAL	Flower Garden	FOOTBALL
Basic embedded	27.00 dB (27.33:27.02:26.95)	28.36 dB (29.16:27.94:28.43)	34.23 dB (36.08:34.41:33.98)
MPEG1 pred. mode selection (%)	(92.94: 9.97: 1.42)	(89.51: 9.29: 1.67)	(70.97:27.47: 3.97)
New decision rule	26.98 dB (27.29:26.99:26.94)	28.34 dB (29.14:27.93:28.42)	34.39 dB (36.01:34.74:34.10)
MPEG1 pred. mode selection (%)	(92.94:10.71: 2.35)	(89.51:11.10: 3.20)	(70.97:44.64:16.31)

Table 7 MPEG1 prediction mode for odd field only

	MOB&CAL	Flower Garden	FOOTBALL
Basic embedded	27.00 dB (27.33:27.02:26.95)	28.36 dB (29.16:27.94:28.43)	34.23 dB (36.08:34.41:33.98)
MPEG1 pred. mode selection (%)	(92.94: 9.97: 1.42)	(89.51: 9.29: 1.67)	(70.97:27.47: 3.97)
Odd field only	26.93 dB (27.15:26.95:26.90)	28.60 dB (28.48:28.52:28:65)	34.64 dB (35.28:35.29:34.35)
MPEG1 pred. mode selection (%)	(92.94: 17.34: 2.46)	(89.51: 51.65: 5.23)	(70.97:79.65:22.34)