ISO/IEC JTC1/SC29/WG11
MPEG93/005
January 1993

CCITT SGXV
Working Party XV/1
Experts Group for ATM Video Coding
(Rapporteur's Group on part of Q.3/XV)

Document AVC-429
21-22 January 1993

Source:	**UK**
Title:	**Demonstration of compatibility with H.261**
Purpose:	**Information**

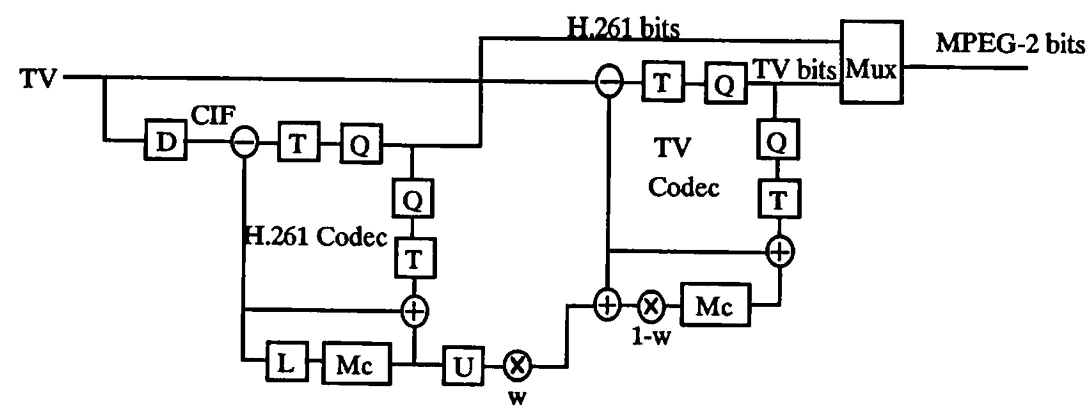
1. Introduction

There is a requirement that H.26x/MPEG-2 shall support compatibility with CCITT Recommendation H.261. The difference between this compatibility requirement and previous compatibility experiments is that different picture rates are used in the layers: H.261 can code progressive pictures at about 30Hz and MPEG-2 can code interlaced pictures at 25Hz. This document demonstrates that the compatibility extensions provided by the MPEG-2 syntax can support compatible coding with H.261 in the lower layer.

2. Compatibility

Compatible coding can be achieved by the use of layered coding. A companion paper, entitled 'Compatible coding', identifies three forms of layered coding: simulcast, loose-coupling and tight-coupling. Simulcast allows the coding of the layers to be done independently but is inefficient in terms of bandwidth utilisation. Tight-coupling, such as used in the 'frequency split' scheme, is more efficient in terms of bandwidth utilisation than simulcast but places constraints on the layers, both in terms of the coding algorithms that can be used and the resolution and the picture rates that can be coded. Loose-coupling, such as used in the 'spatial split' scheme, is as efficient in terms of bandwidth utilisation as the tight-coupling but does not place constraints on the layers, such as the coding algorithms that can be used and the resolution and the picture rates that can be coded.

Figure 1 shows a loosely-coupled two-layered encoder for TV with embedded H.261 encoding of 30Hz CIF. The H.261 codec processes images that have been down-sampled from TV. This down-sampling process is not subject to standardisation and can be simple or complex.



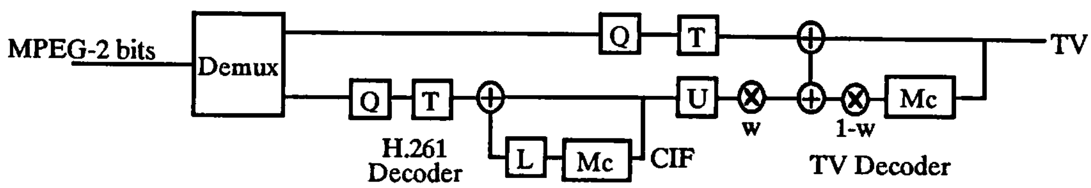
U: Up sampling from 30Hz CIF to TV
D: Down sampling from TV to 30Hz CIF
T: Forward and reverse transform
Q: Forward and reverse quantizer
MC: Motion compensation
L: Loop filter
w: Weighting

Figure 1. MPEG-2 TV/H.261 loosely-coupled compatible encoder.

The TV coder processes the TV images. This part of the coding can utilise a prediction from the locally decoded pictures of the H.261 coder after suitable up-sampling. The up-sampling process is subject to standardisation and is still under study.

When the TV coder is allowed to form predictions using the locally decoded pictures of the H.261 coder after suitable up-sampling, it is operating in a loosely-coupled fashion. When this choice is disabled, it is operating as simulcast.

The corresponding decoder is shown in figure 2. It should be noted that many of the functions for the H.261 coder/decoder and TV coder/decoder are the same. Hence additional hardware is not needed for these functions provided a 30% increase in throughput can be achieved.



U: Up sampling from 30Hz CIF to TV
D: Down sampling from TV to 30Hz CIF
T: Forward and reverse transform
Q: Forward and reverse quantizer
MC: Motion compensation
L: Loop filter
w: Weighting

Figure 2. MPEG-2 TV/H.261 loosely-coupled compatible decoder.

3. Format conversion

Standards conversion between TV and CIF requires horizontal and vertical-temporal sample rate changing. For horizontal sample rate conversion, a horizontal filter such as those described in Test Model 3, section 3.3, can be used. For the vertical sample rate conversion, a vertical-temporal filter is required to give good motion performance and vertical resolution.

The 704x576 25Hz interlaced source was resampled using non-adaptive filtering with a three field aperture as described in MPEG 92/758 to 704x576 50Hz progressive, which was then horizontally and vertically filtered to 352x288 50Hz progressive. These fields were then sub-sampled to 30Hz.

After coding according to CCITT Recommendation H.261, the 30Hz CIF progressive pictures were temporally up-sampled to 50Hz by picture repeating. These fields were then horizontally and vertically interpolated to 704x576 50Hz progressive and sub-sampled to give 704x576 25Hz interlaced.

The temporal sub-sampling and picture repeating described above is illustrated in figure 3. It can be seen that the pattern repeats after every five input frames.

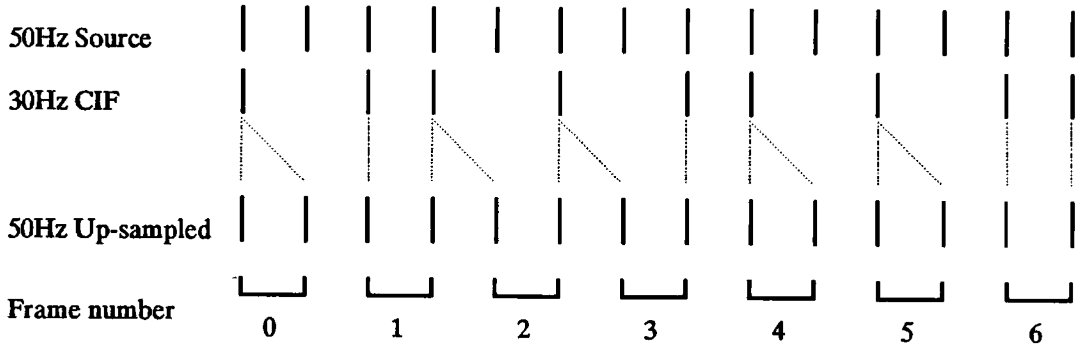


Figure 3. The temporal relationship between source, CIF and up-sampled fields.

4. Encoder Parameters

4.1. Base layer Parameters

The base layer source pictures were progressive at 30Hz. They were coded at 0.5Mbit/s using the H.261 coding algorithm.

4.2. Top layer Parameters

The encoder is Test Model 3, with M=3 and N=Sequence length, and with frame/field adaptive prediction. Compatible prediction from the reconstructed CIF pictures was done using the syntax for compatibility defined in the proposal of MPEG92/651. This syntax indicates, at the sequence layer, the macroblock tables to be used. This greatly simplifies the syntax at the macroblock layer.

The motion estimation was done on the source pictures to half pel accuracy and refined in the coding loop by +/- 1.5 pels. The compatible prediction used the spatio-temporal weighting technique. Table 1 gives the weightings and the fixed length codes used to represent them. The decision to use compatible coding is made on a macroblock basis. The weightings of table 1 apply to predicted macroblocks. In the intra picture only two modes are allowed: pure compatible or pure intra.

The bit rate used in this layer was 1.5Mbit/s.

	Field 1		Field 2	
Code	CIF	TV	CIF	TV
0 0	1.0	0.0	0.0	1.0
0 1	0.0	1.0	1.0	0.0
1 0	0.5	0.5	0.5	0.5
1 1	1.0	0.0	1.0	0.0

Table 1. Spatio-temporal weightings.

5. Simulation

The source interlaced TV sequences were down-converted from CCIR recommendation 601 resolution to CIF (352x288 progressive fields at 30Hz). These CIF pictures were then coded with 0.5Mbit/s using the H.261 coding algorithm. The resulting coded pictures were up-sampled back to TV resolution and used as the compatible prediction when coding the source TV pictures with an additional 1.5Mbit/s using Test Model 3 with frame/field adaptive prediction and compatible prediction.

The Susie source sequence was used. The two-layered coded TV pictures were compared when coupling between the layers was enabled (spatial split) and when coupling was disabled (simulcast).

6. Results

The results are summarised in tables 2 and 3. A tape will be shown.

It can be seen that the loosely-coupled (spatial split) scheme achieves significantly better subjective picture quality than, as well as a luminance SNR gain of 0.74dB over, the zero-coupled simulcast scheme.

Figure 4 shows a plot of luminance SNR against frame number for simulcast and compatible coding. It can be seen that there is little difference in SNR at the beginning of the sequence and at the end, when there is little motion and the SNR is high. During the middle of the sequence when the motion is rapid and not just a translation in two dimensions, the motion compensation does not perform as well as before, and the SNR falls. However, in the loosely-coupled case, the SNR does not fall as much as it can take advantage of the lower layer reconstructed picture for use in prediction. The maximum gain over simulcast is 3.3dB. This has a significant effect on subjective picture quality.

Item	All	Intra	Predicted	Interpolated
Number of pictures	121	1	40	80
SNR for Y	39.37	41.19	39.57	39.24
SNR for U	44.86	46.43	44.89	44.82
SNR for V	45.01	46.43	45.02	44.98
Mean value of QP	9.23	4.23	5.40	11.21

Table 2. TM3 statistics for zero-coupled simulcast coding of Susie. Bit rate: 0.5+1.5Mbit/s.

Item	All	Intra	Predicted	Interpolated
Number of pictures	121	1	40	80
SNR for Y	40.11	41.64	40.35	39.97
SNR for U	45.10	46.20	45.28	45.00
SNR for V	45.25	46.30	45.40	45.16
Mean value of QP	6.69	2.51	3.82	8.18

Table 3. TM3 statistics for loosely-coupled compatible coding of Susie. Bit rate: 0.5+1.5Mbit/s.

Figure 5 shows the number of compatible macroblocks per frame, but only for predicted pictures. It shows how many were predicted from field 0 only, field 1 only and both fields; the number of averaged macroblocks is not shown. It can be seen that the number of these 'pure' compatibly predicted macroblocks varies across the sequence, peaking when the motion is rapid in the middle of the sequence. This is in agreement with the discussion above on the variation through the sequence of luminance SNR.

Figure 3 illustrated the temporal sub-sampling and picture repeating process used in the up and down-sampling to/from CIF, which can be seen to repeat every five input frames. Frames 1, 6, 11, ... of the up-sampled CIF are temporally correct, while frames 0, 4, 5, 9, ... have field 0 temporally correct and field 1 incorrect and frames 2, 3, 7, 8, ... have field 1 temporally correct and field 0 incorrect. This correlates exactly with the compatible modes chosen as illustrated in figure 5: frames 1, 6, 11, ... have many macroblocks with both fields coded compatibly, frames 0, 4, 5, 9, ... have many macroblocks with only field 0 coded compatibly and frames 2, 3, 7, 8, ... have many macroblocks with only field 1 coded compatibly.

The temporal imperfections of the up and down-sampling process therefore have an effect on the performance of the compatible prediction. Presumably if all fields were temporally correct, more macroblocks with both fields compatibly predicted would be coded, and the gain over simulcast would be even greater. Temporal interpolation could be used to produce temporally correct fields, but would cause spatial distortion, blurring, which would not aid prediction. Alternatively, motion compensated interpolation could be used. This could be implemented by associating field motion vectors with compatibly coded macroblocks as appropriate. This requires further consideration.

7. Conclusion

It has been shown that the 'spatial split' scheme from Test Model 3 provides the possibility of compatible layered coding with 30Hz pictures coded according to H.261 in the lower layer and 25Hz interlaced pictures coded in the upper layer. However, it does not allow the up-sampling process to be specified: the requirements of the general up-sampling process needs further study and a subsequent syntax definition is necessary.

It has been shown that the simplified compatible syntax, proposed to MPEG in November in MPEG92/651, is valid.

It was shown that the two-layered loosely-coupled coding scheme (spatial split) can achieve both better subjective and objective picture quality than the simulcast scheme. In particular, the gain is achieved when the motion compensation in the upper layer is not performing well, as in the case of rapid, rotational or three dimensional motion, as opposed to two dimensional translational motion.

It was suggested that the performance of the loosely-coupled (spatial split) scheme could possibly be enhanced when the picture rates in the layers are different, by the application of motion compensation to the reconstructed lower layer picture before using it as a prediction for the upper layer.

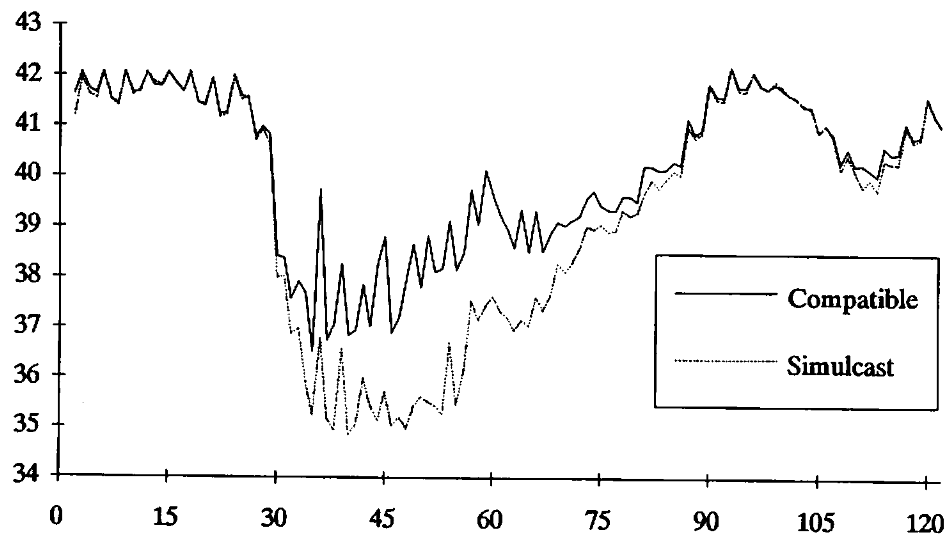


Figure 4. Luminance SNR against frame number for simulcast and compatible coding.

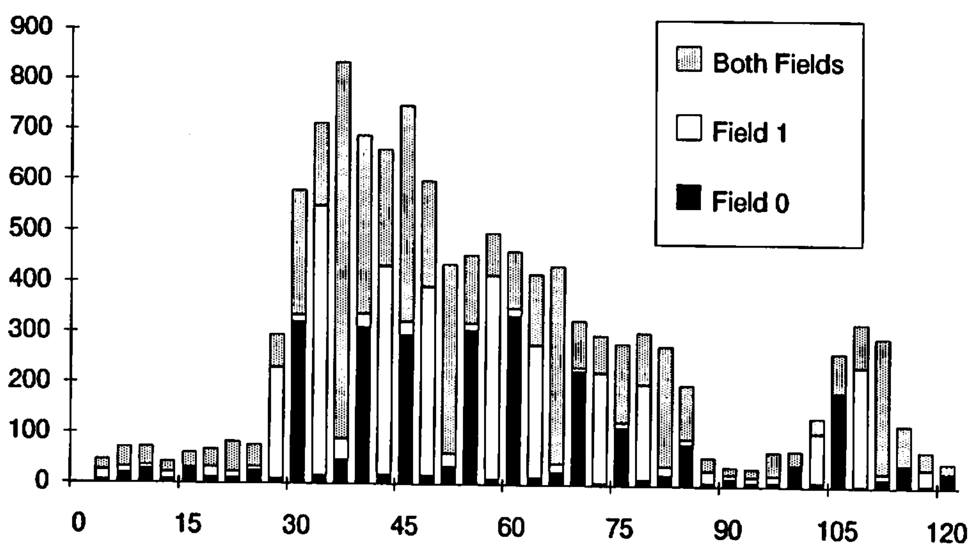


Figure 5. The number of compatible macroblocks/frame for predicted pictures only.