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

Document AVC-346
September 1992

INTERNATIONAL ORGANISATION FOR STANDARDIZATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND ASSOCIATED AUDIO

ISO/IEC JTC1/SC2/WG11
MPEG92/
24 September 1992

SOURCE: John Arnold, Australian Defence Force Academy (on behalf of Australian Universal Video Coding Project)

TITLE: Results of Core Experiment I.1 - Interlace-in-Interlace Extraction

PURPOSE: Information

Abstract

This document reports on the results obtained after performing MPEG Core Experiment I.1 - Interlace-in-Interlace Extraction. Comparisons are made with a simple upper 4x4 extraction technique both in terms of PSNR and bit rate in each layer. In addition, the effect of using simple extraction on images coded in field mode and adaptive field/frame mode are presented.

Introduction

This core experiment was proposed as it was recognised that a two-layer scalable video hierarchy made up of interlaced services at both layers will be important in future applications. In these applications, the quality obtained from each layer may be critical.

A scale-4 service is obtained by taking a 4x4 subset of the 8x8 DCT co-efficients produced by the encoder. In its most simple form, this is done by taking the upper 4x4 block of co-efficients as shown in Figure 1.

```
X X X X * * * *
X X X X * * * *
X X X X * * * *
X X X X * * * *
* * * * * * * *
* * * * * * * *
* * * * * * * *
* * * * * * * *
```

Figure 1 Simple Extraction of a 4x4 Block of DCT Co-Efficients

Three other extraction techniques are proposed in this core experiment as shown in Figure 2.

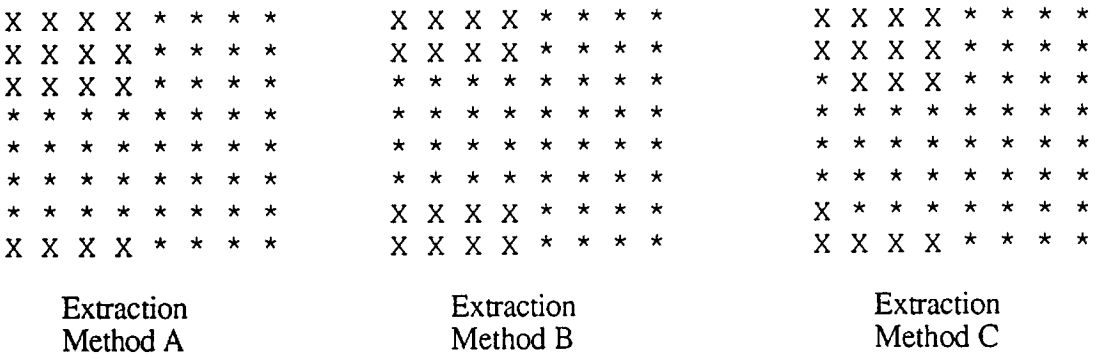

Figure 2 Extraction Methods Proposed in Core Experiment

Experimental Method

A full two layer scalable bit stream was produced as described in TM2. The extracted 4x4 block of DCT co-efficients was coded and transmitted in the slice layer while the 8x8 block of DCT co-efficients (with co-efficients already transmitted set to zero) was transmitted in a scaled slice. Standard zig-zag scanning order as described in TM2 was used.

Coder Parameters

Chroma sub-sampling:	4:2:0
Bit Rate:	4.0 Mbit/s
I-frame Spacing:	15
P-frame Spacing:	3
DCT coding:	Frame Mode
Motion Compensation:	Frame mode
Motion Vectors:	Half Pixel accuracy at scale-8
Run/Amplitude VLC Tables:	As defined in TM2 Appendix D

A separate decoder was used to produce the scale-4 reconstructed image sequence. The 4x4 DCT co-efficients were extracted from the slice layer while the slave slice was ignored. Quarter pixel motion estimation accuracy was obtained using the interpolation technique described in TM2 Appendix D applied to the previous scale-4 reconstructed I and P frames.

Test Sequences:	Bicycle and Cheer Leaders (704 x 480)
Number of Frames:	45 (3 GOPs)

PSNR measurements are made on luminance only with respect to a SIF interlaced (SIF-I) sequence generated in accordance with TM2.

Results of Core Experiment

The results for the basic core experiment at scale-4 are summarised in Tables below.

	Simple Extraction	Method A	Method B	Method C
PSNR (dB)	22.90	23.33	22.80	22.97
Data Fraction (%)	70.4	74.3	72.5	72.3

Table 1 PSNR and Fraction of Data Used for Scale-4 Reconstruction for Each Extraction Technique (Bicycle)

	Simple Extraction	Method A	Method B	Method C
PSNR (dB)	24.00	24.79	24.54	24.64
Data Fraction (%)	71.4	76.8	75.6	75.6

Table 2 PSNR and Fraction of Data Used for Scale-4 Reconstruction for Each Extraction Technique (Cheer Leaders)

These results show that in terms of PSNR, the Method A extraction technique produces the highest quality scale-4 reconstruction for both sequences. For the Bicycle sequence, Method C also produces a slightly superior result compared to simple extraction while for the Cheer Leaders sequence all three of the other methods are superior to simple extraction.

It is also worth noting the large volume of data which is required to reconstruct the scale-4 reconstruction. The SIF size service would require approximately 3 Mbit/s (i.e. around 75% of the total rate) while the full size service requires only an additional 1 Mbit/s. This arises because:

(i) all overhead information is transmitted in the lowest layer.
(ii) the DCT co-efficients transmitted in the lowest layer contain a large proportion of the total information.

The quality obtained at the SIF level given this high data rate is only fair subjectively.

Using a layered approach causes some increase in the total amount of information to be transmitted. Rate control therefore results in a decrease in reconstructed quality at the scale-8 level. This effect is summarised in Table 3 for the bicycle sequence.

	Single Layer	2 Layer Simple Extraction	2 Layer Method A Extraction	2 Layer Method B Extraction	2 Layer Method C Extraction
PSNR (dB)	27.98	26.89	26.94	26.92	27.15

Table 3 PSNR of Scale-8 service for Various Extraction Techniques (Bicycle)

The highest service quality has been reduced by more than 1 dB in all cases.

Comparison with Field Based Coding

A further set of experiments was performed to compare the results obtained with the various extraction techniques described above with those obtained when DCT coding was performed in field mode and in adaptive field/frame mode as described in Section 4.5 of TM2. The simple extraction technique of Figure 1 was employed to produce the scale-4 reconstruction. The results are summarised in Tables 4 and 5.

	Field Mode DCT Coding	Adaptive Field Frame DCT Coding
PSNR (dB)	23.14	23.14
Data Fraction (%)	76.5	77.5

Table 4 PSNR and Fraction of Data Used for Scale-4 Reconstruction for Simple Extraction Technique Applied to Field Mode Coded and Adaptive Field/Frame Mode Coded Sequences (Bicycle)

	Field Mode DCT Coding	Adaptive Field Frame DCT Coding
PSNR (dB)	24.64	24.24
Data Fraction (%)	79.9	78.4

Table 5 PSNR and Fraction of Data Used for Scale-4 Reconstruction for Simple Extraction Technique Applied to Field Mode Coded and Adaptive Field/Frame Mode Coded Sequences (Cheer Leaders)

In both cases, field mode DCT coding with simple extraction performs as well as all of the frame based extraction techniques except for Method A which is slightly superior (approximately 0.2 dB). The adaptive field/frame approach performs equally as well for the

bicycles sequence while performing slightly less well on the cheer leaders sequence. In all of cases which employ field mode, the fraction of the data in the bottom layer is higher than for the frame based case.

Again, some feel can be obtained for the additional overhead caused by the scaling by comparing the PSNR of the scalable and non-scalable coders. This is done in Table 6 for the bicycle sequence.

	Single Layer Field Mode	Single Layer Adaptive Field/Frame	Field Mode Coding Simple Extraction	Adaptive Field/Frame Coding Simple Extraction
PSNR (dB)	27.83	27.98	26.90	27.15

Table 6 PSNR of Scale-8 service for Various Extraction Techniques (Bicycle)

As in the case of the frame mode experiments, the quality of the high level service has been reduced by close to 1 dB in both cases. Note that Method C extraction and Adaptive Field/Frame coding with simple extraction produce the highest PSNR values for the scale-8 decoded service.

Drift In Scale-4 Reconstruction

It is well known that achieving scalability using the method described in this work will lead to some drift between the coder and a scalable decoder. This effect is being investigated in other core experiments. In Figure 6, a plot of PSNR of the scale-4 reconstruction versus frame number is plotted for the bicycle sequence coded in frame mode with Method A used to perform the extraction.

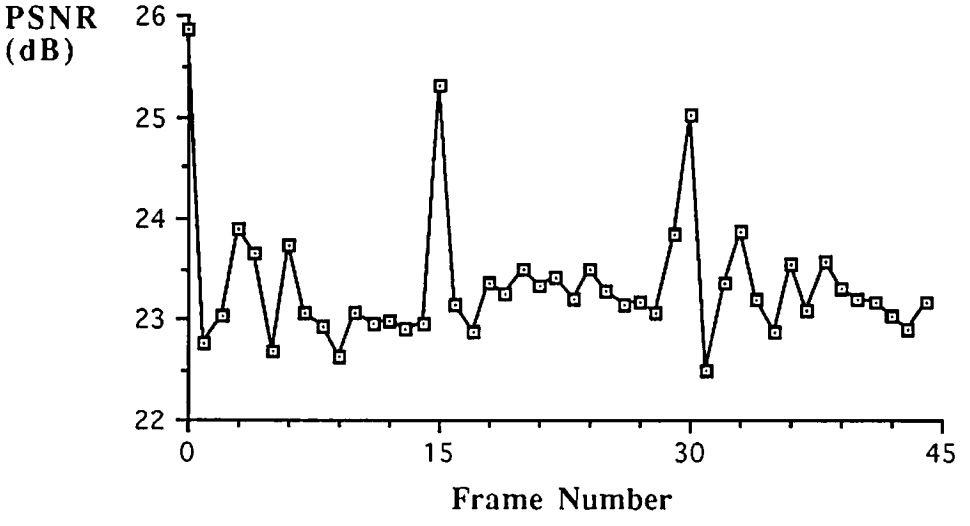

Figure 6 PSNR versus Frame Number for Scale-4 reconstruction of Bicycle Sequence when Method A is Used to Perform the Extraction (Bicycle)

Results are similar for the other extraction techniques. With an I-frame spacing of 15, no significant drift is noticeable. Nevertheless, drift may still be a problem for higher quality services. Further experimentation in this area is needed.

Conclusions

From the experiments performed, the Method A extraction technique provides the highest quality reconstructed scale-4 image sequence. Subjective evaluation of the coded sequences tends to support this conclusion. Using any of these approaches to scaling, however, results in a very high data rate for the scale-4 service (approximately 3 Mbit/s or 75% of the data required for the scale-8 service). In addition, the quality of the scale-4 reconstruction is only fair.

The data volume required for the lower layer could be reduced by confining more of the overhead information to the higher layer or by coding the co-efficients in the lower layer more coarsely with enhancement information transmitted in the upper layer (both of these topics are considered in core experiment I.4). It may well be that the only way to achieve adequate quality in both layers with a more appropriate distribution of data between the layers is to use a coder and decoder with more than one coding loop. Further experimentation in this area is required.