Telecommunication Standardization Sector
Study Group 15
Experts Group for ATM Video Coding
(Rapporteur's Group on Part of Q.2/15)

Document AVC-482
March 1993

INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO-IEC/JTC1/SC29/WG11
CODED REPRESENTATION OF PICTURE AND AUDIO INFORMATION

ISO-IEC/JTC1/SC29/WG11/MPEG92/373

Source: M. E. Nilsson, I. Parke, BT Labs
Title: Enhancement of spatial scalability for reduced cost decoders
Purpose: Discussion

This contribution describes the current scheme used to achieve spatial scalability, and a variation of it that allows reduced cost decoders. The two schemes are merged into one hybrid scheme which can be configured to allow the use of reduced cost decoders for applications for which this is necessary. Finally the two approaches are compared in terms of implementation aspects and performance under channel error conditions.

1. Current spatial scalable scheme

The current method of spatial scalability is based on a coder using the coded pictures from another coder as a possible prediction mode. This spatial prediction can be weighted with the motion compensated temporal prediction to form spatio-temporal weighted prediction. The encoder required for this scheme is shown in figure 1, and the decoder is shown in figure 2.

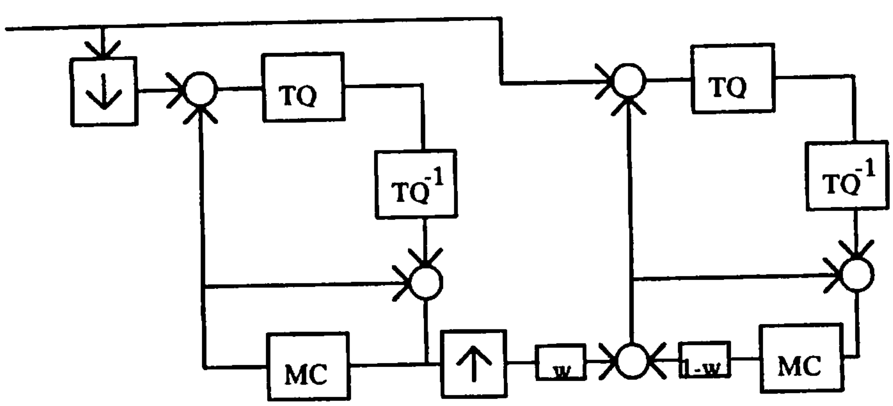

Figure 1. Current spatial scalable encoder.

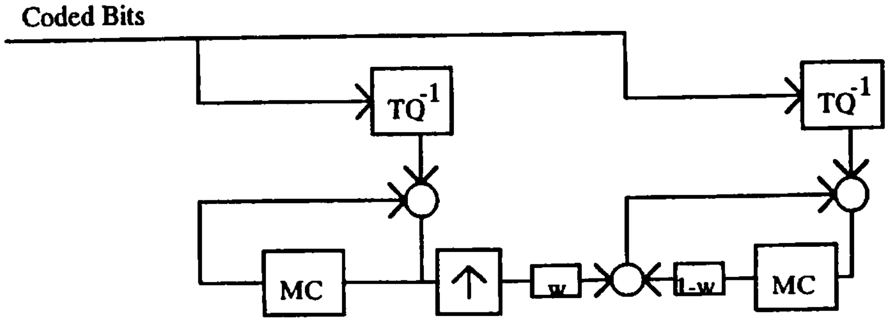

Figure 2. Current spatial scalable decoder.

This type of scheme has previously been referred to as a 'loosely-coupled' scheme. This simply means that the coding and the source formats used in the various layers can be chosen independently. This enables compatibility to be achieved when the picture rates in the layers are different. This scheme has been shown to achieve good picture quality with H.261 being used to code low resolution progressive pictures at 30Hz and MPEG-2 to compatibly code higher resolution interlaced pictures at 50Hz.

2. A spatial scalable scheme for reduced cost decoders

It has been argued in the past that, [illegible] current spatial scalable scheme requires two sets of picture stores in the decoder to reconstruct the hi[illegible]yer pictures, the cost of higher layer decoders is not minimised. In order to reduce the cost of these decoders, a version of the spatial scalable scheme which requires only one set of picture stores in the decoder, has been suggested [1].

This method of spatial scalability is based on a coder using the coded prediction error from another coder as a possible prediction of the prediction error. This spatial prediction error estimate from the other coder can be considered to be multiplied by a weight before being used to predict the prediction error of the higher layer coder. Until now, only weights of 0 and 1 have been used: these correspond to not using the prediction error estimate at all, or using it all. The encoder required for this scheme is shown in figure 3, and the decoder is shown in figure 4.

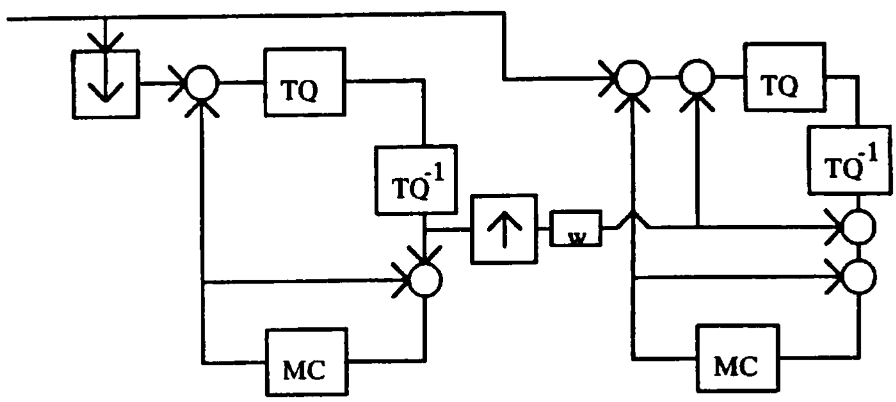

Figure 3. Alternative spatial scalable encoder.

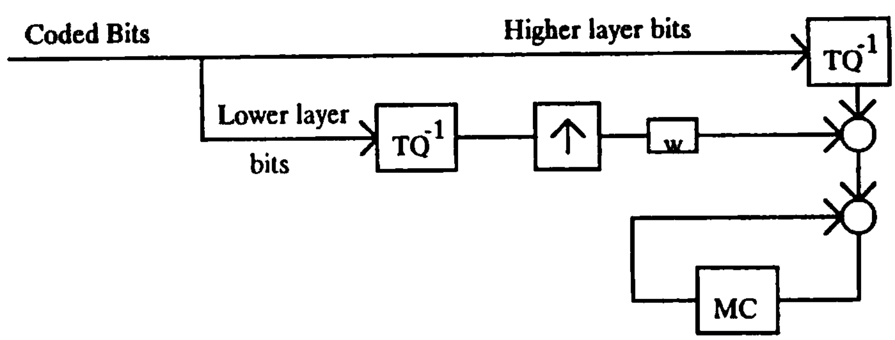

Figure 4. Alternative spatial scalable decoder.

This type of scheme has previously been referred to as a 'tightly-coupled' scheme. This simply means that the coding and the source formats used in the various layers can not be chosen completely independently. The scheme works well when the two coders are operating in similar ways, when both are using the same prediction modes and the same motion vectors, etc. Although in principle the scheme can handle different picture rates in the layers, the performance of the scheme would probably not be very good.

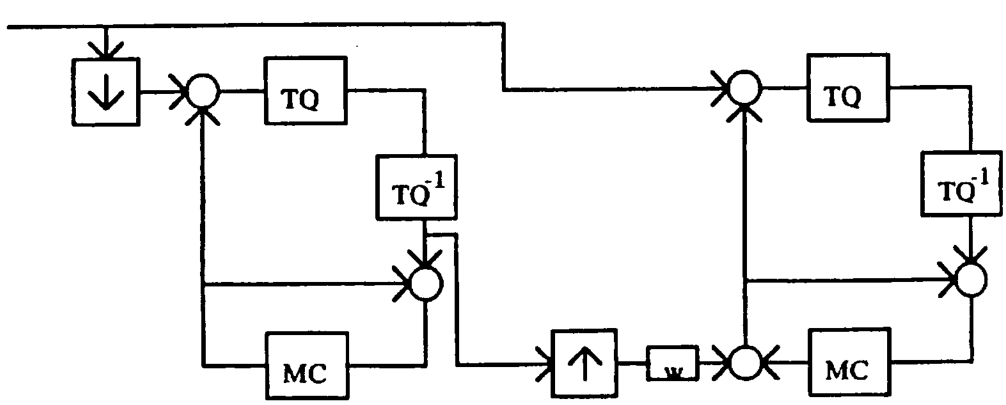

Figure 5. Reorganised alternative spatial scalable encoder.

3. Hybrid spatial scalable scheme

The two methods for achieving spatial scalability outlined above are not totally different. The main difference concerns the point in the lower layer coder from which the prediction for the upper layer is taken. In the current scheme it is taken after the adder and consequently represents a coded picture; this picture can then be weighted with the temporal prediction to form a spatio-temporal prediction. In the alternative scheme, it is taken before the adder and consequently represents a coded prediction error. This coded prediction error can then be weighted and subtracted from prediction error of the higher layer coder. However, as the order of subtraction's is irrelevant, the coded prediction error could be added to the temporal prediction before subtracting from the source, without changing the functionality at all. This is shown in figure 5.

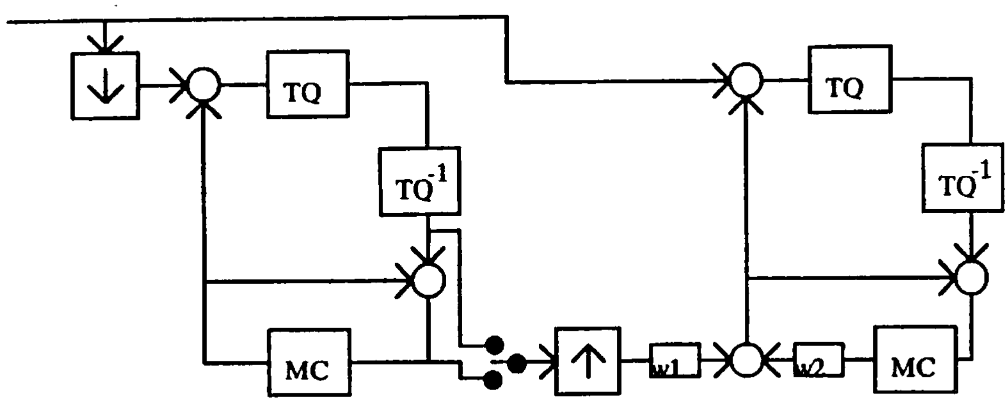

Figure 6. Hybrid spatial scalable encoder.

When drawn in this way, the similarity of this alternative spatial scalable scheme and the current scheme is obvious. Figure 6 shows a hybrid encoder with both schemes incorporated, with a switch being used to select the point from which the lower layer data is taken. For the current scheme, the weights w1 and w2 would be related by w1 + w2 = 1. For the alternative scheme the weight w2 would always be 1, and the weight w1 would be 0 or 1. Figure 7 shows the decoder.

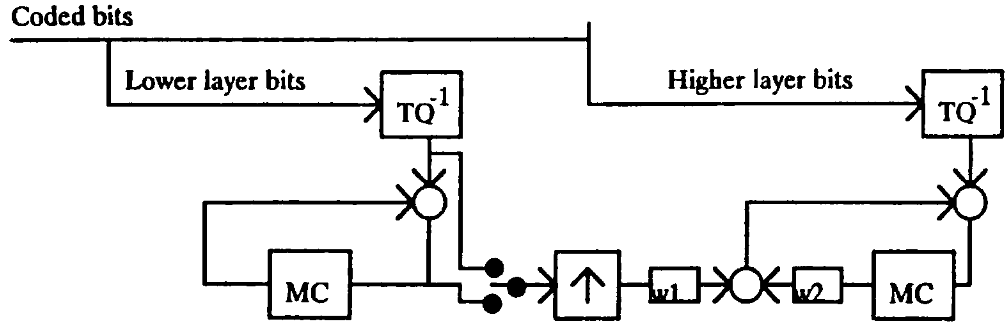

Figure 7. Hybrid spatial scalable decoder.

The syntax for the hybrid scheme could be the same as the current spatial scalable syntax. The only addition required is a mechanism for indicating the position of the switch. It may be appropriate to signal this at the sequence layer, and leave it unchanged throughout a sequence. If it were permanently in the 'up' position for a particular application, the lower layer set of picture stores would not be required in the decoder. This provides the capability of allowing reduced cost decoders for some applications.

If the position of the switch can be altered more often than once per sequence, without incurring too much cost in terms of overhead bits, there is a possibility to improve the coding performance. For example, when the same prediction modes are used in each layer, the upper position of the switch may prove to be more effective, while for other macroblocks the lower position may be more effective.

This hybrid approach to spatial scalability makes use of 8x8 DCT's only. This allows main profile decoders to decode the lowest layer of a spatially scalable bitstream. An example of this would be the capability to extract a TV picture from a HDTV spatially scalable bitstream.

4. Picture quality and decoder cost trade-off

This section compares the overall performance and requirements of the two components of the hybrid spatial scalable approach.

4.1. Memory and processing requirements

The decoder for the alternative spatial scalable scheme, shown in figure 4, requires only one set of picture stores to decode the higher layer. The decoder for the current spatial scalable scheme, shown in figure 2, requires two sets of picture stores to decode the higher layer. However, when the lower layer source pictures contain a quarter of the number of pels of the higher layer, as in the case of SIF and CCIR 601 resolution images in the layers, the amount of memory needed by the current scheme is only 25% more than that of the reduced cost decoder.

The same reasoning applies to processing requirements. Although the decoder is usually drawn as consisting of two loops, this is simply to demonstrate the functionality clearly; in a practical implementation, the same processing hardware would be used for both lower and higher layers. The additional processing requirements for the current scheme compared to the alternative one are simply that memory accessing must occur faster, only 25% faster for the example of the previous paragraph. This shared processing hardware architecture is shown in figure 8.

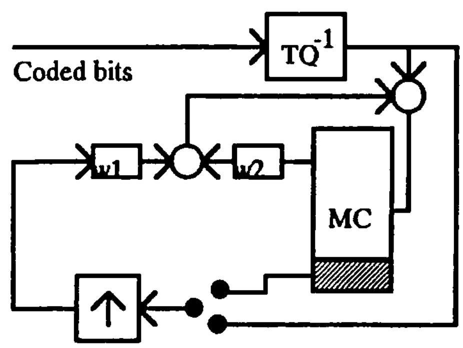

Figure 8. Shared processing architecture of hybrid spatial scalable decoder.

4.2. Error resilience

The presence of the reconstructed lower layer picture in the decoder helps performance in the case of bitstream errors in the higher layer bitstream, as the decoder can in this event 'fall-back' to displaying the lower layer in the areas of the picture affected by errors. The reduced cost decoder does not have this option to substitute the lower layer in the areas of the picture affected by errors.

5. Conclusion

In this contribution the current and an alternative, reduced decoder cost, scheme to achieve spatial scalability is described. This second scheme was shown to be architecturally similar to the current scheme, being described with the existing spatial scalable syntax. The two schemes were merged into one hybrid scheme, with all the functionality of the two individual schemes.

This hybrid spatial scalable scheme can be configured according to the particular application. It can allow the use of reduced cost decoders for applications for which this is necessary, it can allow the flexibility to compatibly code pictures at different picture rates well, for applications where this is necessary, as well as possibly offering the opportunity to improve the coding performance of the spatial scalable approach.

Finally it was shown that in the typical application with four times as many pels in the upper layer as the lower layer, the current scheme requires only 25% more memory and 25% faster memory access than the reduced cost decoder.

6. Reference

[1] MPEG92/165, Compatible coding structure, PTT Research.