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

Document AVC-74
August 1991

SOURCE: Australia

TITLE: The Efficiency of a Layered Pyramid Coder

PURPOSE: Information

Abstract

This document describes the results of an investigation into the efficiency of a layered Laplacian pyramid coding scheme which provides interworking between video services operating at two different spatial resolutions. The efficiency is compared to that of a single layer system which codes the highest resolution signal. The layer signals were coded using techniques similar to those used in H.261 and MPEG1 and the scheme could provide backward/forward compatibility with these standards. The results indicate that the Laplacian Pyramid scheme is potentially as efficient as a single layer coder.

1. Introduction

Flexible layering [AVC-35] has been recognised as a possible approach to video service interworking on the B-ISDN. Concern has been expressed, however, about losses in coding efficiency due to the layering process [AVC-65R]. This document discusses the efficiency of layered coding using a Laplacian pyramid structure which can achieve flexible interworking between video services. The performance of the layered scheme is compared to that of a conventional single layer approach. The results show that the layered coding scheme is not, in principle, less efficient than a single layer coder.

2. The Layered Laplacian Pyramid Interworking Scheme

Figure AVC-74/1 illustrates the construction of a two layer Laplacian pyramid codec capable of interworking with other codecs operating at two different spatial resolutions. The layer 0 signal is formed by spatial lowpass filtering and decimation of the input. In our experiments, the input was a CIF resolution signal and the lower layer sequence had a QCIF resolution. The low resolution signal is coded, decoded and then interpolated. This interpolated signal is subtracted from the original to form the layer 1 enhancement signal, which is subsequently coded. At the decoder, a reconstructed QCIF resolution signal is obtained by decoding the layer 0 signal. This is then interpolated and added to the decoded layer 1 enhancement information to form the CIF resolution sequence.

Thus, the high frequencies of the original image are transmitted in the upper layer while the lowpass information is primarily contained in the lower layer. The decomposition of the original video information into two layers offers tolerance to cell loss. Since the important lowpass information is contained in Layer 0, this information can be transmitted with higher priority (and thus a lower cell loss rate) than Layer 1. Cell loss from Layer 1 information will then have only a limited effect on the quality of the reconstructed CIF sequence.

3. Compatibility and Interworking Capabilities

The pyramid coder described is fully upward/downward compatible with single layer QCIF resolution coders. The scheme is also capable of providing flexible layering [AVC-35]. That is, by switching the lower layer coding off, the scheme functions as a single layer coder at CIF resolution, which could interwork with other schemes such as H.261 and MPEG1 at this resolution.

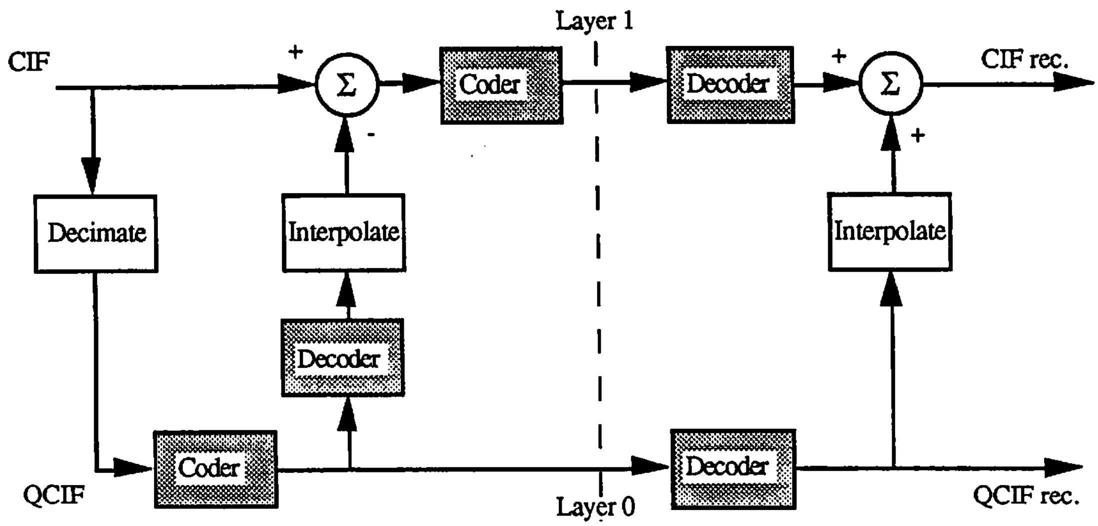

Figure AVC-74/1. Two-Layer Laplacian Pyramid Interworking Scheme

Layering schemes based on this pyramid architecture may be easily extended using further layers to provide interworking over a larger range of resolutions. The scheme allows service enhancements to be introduced in a compatible manner by adding layers to the top of the hierarchy. Lower resolution layers can also be dropped as they fall into disuse, without affecting high resolution services.

Since the decomposition, or layering, is separated from the coding it is possible to use a range of different coding schemes for the layer signals. The structure can be made compatible with existing standards, such as H.261 or MPEG1. It is also possible to introduce coding schemes which are specifically designed to code the layer signals and coding schemes which may be more suitable for the B-ISDN [AVC-73].

4. Coding Performance

Interworking is a key requirement of video coders which will operate over the B-ISDN. However, it is important that the coding efficiency of multi-layer interworking codecs is at least comparable to that achieved by single layer schemes. To estimate the bit-rate efficiency of the two-layer pyramid scheme, a number of computer simulation studies have been performed using CIF resolution source material.

The purpose of the study was to investigate the coding efficiency of the Laplacian pyramid interworking scheme when combined with a coding method that fits in well with existing coding standards. For this reason, Hybrid DPCM/Transform coders, similar to H.261, were adopted to code both the layer sequences (see Appendix A). Each coder was designed to operate in variable bit rate mode, by fixing the quantizer step-size.

The coders on each layer operate independently. In this way, it is possible to achieve flexible interworking with existing standard coders at either CIF or QCIF resolution. The efficiency of the pyramid coder was investigated separately for intraframe, interframe and interframe+motion compensation (MC) modes. The Entropy of the non-zero valued DCT-coefficients was calculated to estimate the efficiency of both single layer and pyramid interworking approaches (see Appendix A).

Figure AVC-74/2 shows the rate-distortion functions for the pyramid codec in different modes with the CIF sequence "TREVOR". Points on the rate-distortion function were obtained using various quantizer stepsizes in the layer coders. The bit rate calculated for the two layer scheme is the sum of the bit rates of the two layers. Overheads associated with motion vectors, etc. have been ignored in this study. The peak signal to noise ratio (PSNR) was calculated as an average for the CIF resolution sequence. The Laplacian pyramid method performed very efficiently in intraframe mode and achieved a coding gain compared to the single layer approach. It becomes less efficient than the single layer coder when interframe prediction and motion compensation is invoked. Nevertheless, the bit rate overhead found for interframe+MC is small compared to the total service bit rate.

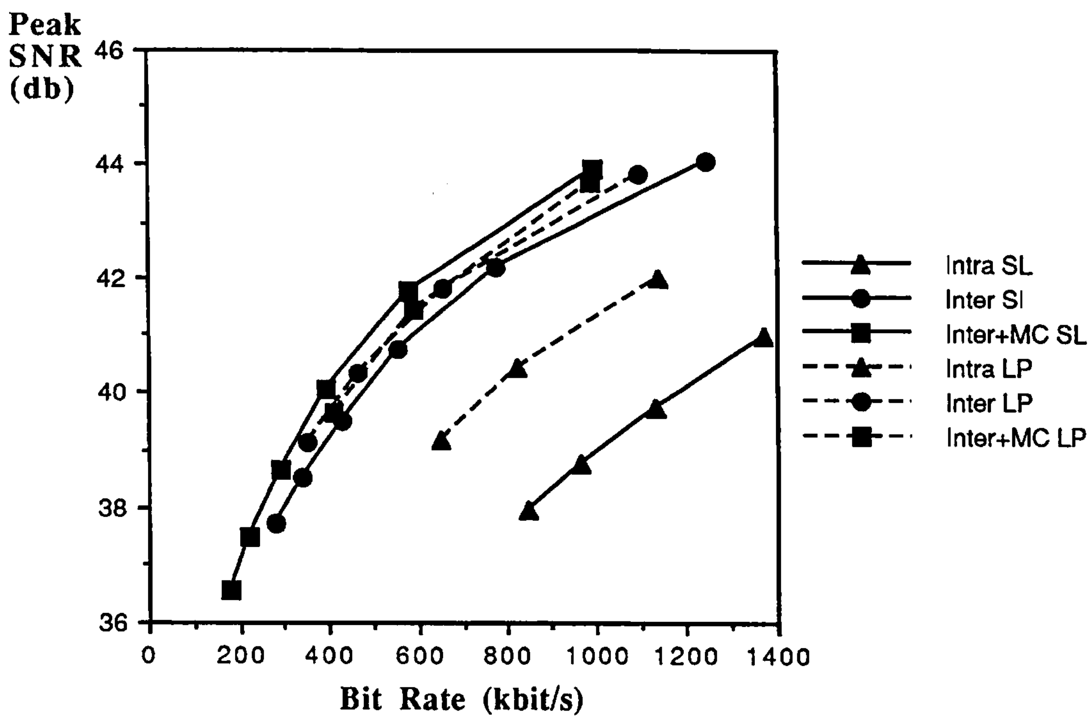

Figure AVC-74/2: Comparison in intraframe, interframe and interframe+MC mode.
SL denotes the results for the single layer approach and LP for the Laplacian Pyramid interworking scheme.
Results were obtained using a 12 tap filter for decimation and interpolation.

Figure AVC-74/3 demonstrates the coding gain that can be achieved in interframe+MC mode by redesigning the filters used in the pyramid decomposition. The filter coefficients are given in Appendix A. Using a 3-tap filter with a lower cut-off frequency for decimation and interpolation increased the efficiency of the two layer coder in interframe+MC mode. Better efficiency was observed when different filters were used for decimation and interpolation. Using a 3 tap filter for decimation and a 5 tap filter with greater stopband attenuation for interpolation, improves performance. The 5 tap filter interpolation filter smooths the quantization noise introduced in Layer 0 more effectively than a 3 tap filter. As a result, the motion compensation on Layer 1 performed more effectively and a coding efficiency comparable to the single layer approach was achieved.

5. Conclusion

The Laplacian Pyramid scheme discussed in this document is capable of implementing flexible layering and is a very effective means of providing interworking between codecs operating at different spatial resolutions. It has an inherent robustness against cell loss and is also capable of providing backward/forward compatibility with existing standards like CCITT Rec H.261 and MPEG1 at any of the layer resolutions.

The simulation studies show that the Laplacian pyramid also serves as an efficient coding scheme, with the potential to encode the information in all layers at a bit rate and quality comparable to single layer coding of the input sequence. Hence layered signal decomposition does not necessarily result in coding inefficiency.

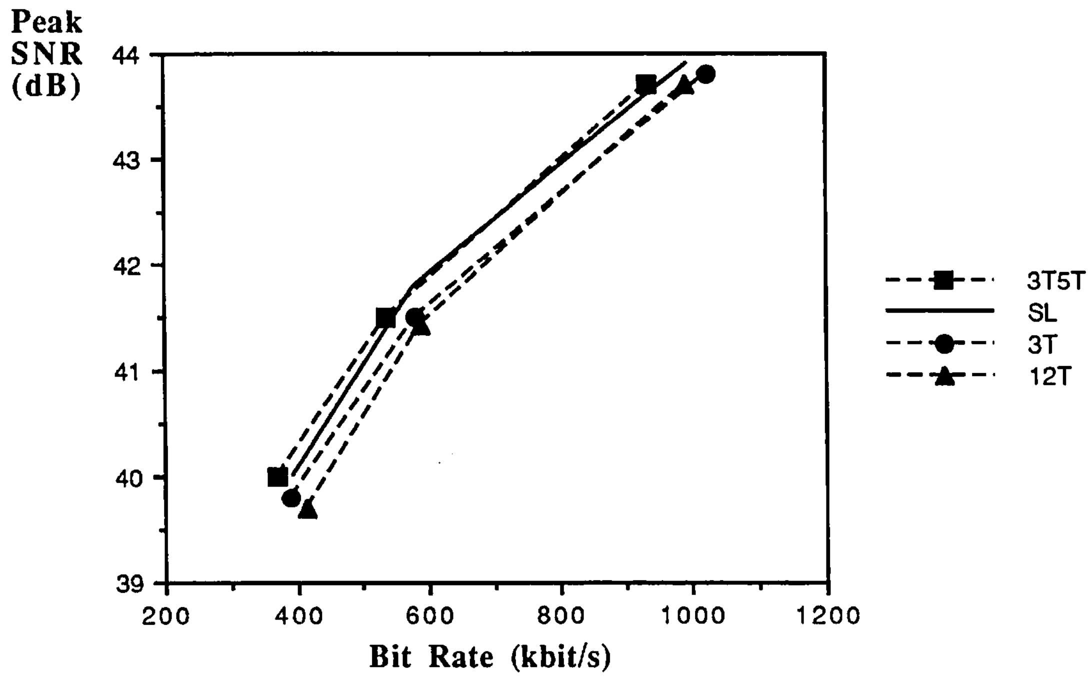

Figure AVC-74/3: Rate-distortion curves in interframe+MC mode using various filters for decimation and interpolation. 12T denotes the 12 tap filter used in the previous experiments. 3T denotes a 3 tap filter for decimation and interpolation and 3T5T the result for a 3 tap filter for decimation and a 5 tap filter for interpolation.

Appendix A. Experimental Set-Up

The encoders for each layer sequence consisted of the basic features of H.261 and MPEG1 standard coders adapted for variable bit rate applications. Coding was performed on luminance only. Each image was divided into blocks consisting of 8x8 luminance pixels. The Discrete Cosine Transform (DCT) was applied either on the original image blocks (intraframe mode) or on blocks of the difference between two consecutive frames (interframe mode). A 3-step block matching motion compensation scheme, applied on blocks of size 16x16 pixels, was used to further exploit temporal redundancies between two frames (interframe+MC mode). The search area was a 32x32 pixel area centred on the 16x16 pixel block. All DCT-coefficients were linearly quantized using the identical values for quantization stepsize and threshold.

The efficiency of the interworking scheme was measured in terms of the entropy of the non-zero, quantized DCT-coefficients. If the levels of the quantized non-zero DCT-coefficients are denoted by $f_1, f_2, ..., f_k$; and if the probability of the level f_i is given by p_i then the entropy is defined as

$$H=-\sum_{i=1}^{K} p_i \log_2(p_i) \qquad \text{(A.1).}$$

The entropy was calculated on a per frame basis.The entropy defined by (A.1) provides a realistic estimate of the bit rate. In this study, 40 frames (frames 35-74) of the videotelephony CIF test-sequence "TREVOR" (a typical head and shoulder sequence with fast motion) were used. An identical single layer coder was applied for comparison purposes to the original CIF resolution sequence. A variety of quantizer step-sizes were used on each layer. The rate-distortion plot only shows the results which yield the minimum bit rate for a given CIF resolution SNR.

The following filters were used in the study:

Co-efficients 3 tap filter: [0.25 0.5 0.25]

Co-efficients 5 tap filter: [0.1 0.25 0.3 0.25 0.1]

Co-efficients 12 tap filter: [-0.0038 0.019 -0.0027 -0.085 0.088 0.48
0.48 0.088 -0.085 -0.0027 0.019 0.0038].