Drift Minimisation In Frequency Scalable Coders Using Block Based Filtering

A. W. Johnson , J. Princen

Applied Research & Development.
Telecom Australia,
770 Blackburn Rd, Clayton Victoria 3168.

Abstract

This paper presents a procedure for the design of quarter pixel precision phase shifting filters for use in an open loop frequency scalable video coding architecture. Simulations have been performed comparing different quarter pixel phase shifting filters. Results show that the performance of five tap block based filters proposed in this paper is significantly better than the bi-linear filters of the MPEG test model, and approach the performance of the more complex minimum drift filters.

1 INTRODUCTION

Frequency scalability is under study in the MPEG 2 video coding algorithm currently undergoing standardisation [1]. Scalable coding introduces some interesting features. For example it allows a scaled version of the signal to be decoded by a reduced complexity decoder, it provides a way to implement Fast-forward and Fast-reverse viewing of encoded bit-streams and it provides high error resilience if prioritised transport is used [1].

A scalable bit stream is produced by partitioning the set of MxM transform coefficients at the encoder into subsets (i.e 2x2, 4x4, NxN) so that a low resolution video sequence can be extracted. A reduced resolution signal is generated by performing NxN point inverse Discrete Cosine Transforms (DCT) using only the first NxN coefficients generated by MxM point DCTs [6]. There are two alternative frequency scalable encoding architectures which have been proposed. These are termed closed loop (embedded) and open loop (non-embedded). The concept of embedded and non-embedded coding originated with DPCM coders where the methods used to achieve hierarchical coding and their advantages were first recognised [2]. These same concepts apply directly to frequency scalable coders. A salient feature of an embedded coding system is that it is capable of reconstructing a high quality reduced resolution signal. The open loop frequency scalable coder is non-embedded because the input signals to the predictor at the encoder and the reduced resolution decoder can diverge, resulting in a loss of signal quality. In the MPEG framework, this behaviour is termed drift [3]. The non-embedded nature of the open loop architecture occurs only when motion compensation is used, and only if the motion vectors are not integer multiples of the transform size. It is apparent that this architecture can be forced to be an embedded system by restricting coding to intra (still image) and/or inter (frame difference), or by forcing the prediction signals at encoder to be equivalent to those extracted at the decoder. The latter is achieved by introducing an additional motion compensation loop at the encoder for each decoder resolution desired making it closed loop. A consequence of this is an increase in encoder complexity and a reduction in coding efficiency. A loss in performance up to 1.3 db has been observed when comparing the performance of two layer frequency scalable embedded to non-embedded coders [4].

A non-embedded frequency scalable system maximises the efficiency of coding and produces a hierarchical system for little additional cost in encoder complexity. This is achieved however, at the expense of some loss in the quality of the decoded reduced resolution signal. As the video coding process is recursive, the non-embedded coder has intra coded frames periodically inserted to correct the accumulation of differences between the decoder and encoder prediction signals.

This paper looks at the problem of maximising the quality of a decoded reduced resolution signal generated from an open loop frequency scalable coder. This is accomplished using phase shifting filters that incorporate both knowledge of the decimation process used to generate the reduced resolution signal, and the interpolation used for sub-pixel motion compensation at the encoder.

2 OPEN LOOP FREQUENCY SCALABLE VIDEO CODING

The two layer, open loop, frequency scalable encoder architecture proposed in MPEG 2 is shown in Figure 1 [1]. Note that motion compensation is performed on the full resolution signal to half pixel precision using a simple bi-linear interpolation filter. Both the low and high resolution decoders receive the same full precision motion vectors.

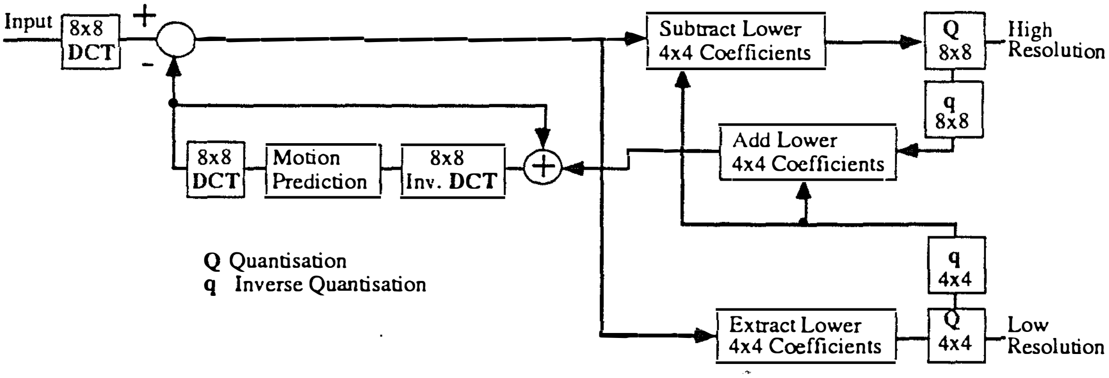

Figure 1 Two layer Single Encoding Loop Frequency Scalable Architecture.

The low resolution decoder has only previously decoded images, from which a prediction signal can be made. To maximise the reconstructed signal quality, it is important that the prediction signal at the low resolution decoder is obtained at quarter pel accuracy (i.e the the full resolution motion vectors are used). A low resolution prediction signal of quarter pixel precision can be obtained by applying a phase shifting filter to a signal extracted at single pixel precision. The phase shifting filter is used to introduce a fractional shift. Figure 2 illustrates a multirate structure capable of generating non-integer shifts. With M equal to four, and using a 7 tap low pass FIR filter whose impulse response is a simple triangle, the equivalent filtering operation for this structure reduces to the quarter pixel precision bi-linear filters specified in section D3 of the MPEG 2 test model [1].

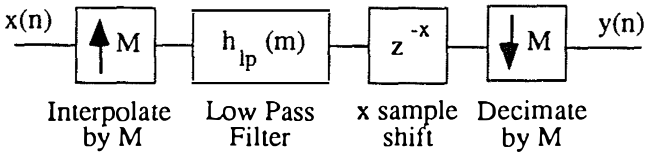

Figure 2 A Multirate structure for realising a fixed shift of x/M samples.

The performance of the open loop scalable coding architecture will be maximised if phase shifting filters at the decoder are optimised for both the filters used to perform half pixel motion compensation, and the equivalent decimation filters used to create the reduced resolution signal. It is readily apparent that the former is employed by the bi-linear filters when implemented as phase shifters, however in the latter case no knowledge is incorporated in their design of the decimation operation using the DCT. Four tap shift invariant filters have been proposed in [7] which do, to a limited degree, take account of the DCT decimation in their design. As a result of their length and the fact they are shift invariant, these filters require additional processing when extracting a prediction signal in the vicinity of a frame boundary. This problem, while easily overcome by extending the frame size by one pixel about the frame boundary, adds to the complexity of the low resolution decoder.

An alternative which does not introduce any frame boundary prediction problems is an approach termed "minimum drift" in the MPEG frame work. The minimum drift approach interpolates the prediction frame using block wise DCT interpolation, extracts the predictions signal at the full frame resolution and then decimates this using the DCT to generate a reduced resolution prediction signal. When using 8x8 transforms at the encoder, it is a relatively simple exercise to show that this combination of operations reduces to the implementation of an 8 tap block based shift varying filter.

3 BLOCK BASED PHASE SHIFTING FILTERS

The "minimum drift" filters add considerable complexity to the low resolution decoder. Filters of reduced complexity can be determined by incorporating the techniques of DCT decimation and interpolation into the multirate phase shifting structure of figure 2.

It was shown in [5] that applying an N point synthesis DCT using the first N coefficients generated by an M point analysis DCT (N < M) can be interpreted as decimating a filtered version of a periodic, symmetrically extended sequence of the original data x(n). The filter has a length of 2M taps, is the result of a cosine summation, and for integer ratios of M to N is shift invariant and even symmetric. Equation 1 defines a symmetrically extended periodic sequence generated from a block of B data samples.

$$x_{se}(p) = \begin{cases} x(p \bmod 2B) & 0 \le p \bmod 2B \le B-1 \\ x((2N-1-p) \bmod 2B) & B \le p \bmod 2B \le 2B-1 \end{cases} \qquad 1.$$

Interpolation can also be implemented using the DCT. This is achieved by applying an M point inverse DCT using the first N coefficients generated by an N point analysis DCT (N < M), with the remaining M-N coefficients forced to zero. When the ratio of M to N is an integer, this operation can be shown to be equivalent to taking the original block of N data samples, forming from it a symmetrically extended periodic sequence, inserting M/N zeros between each sample and then filtering the resultant sequence with an 2M tap filter. The filter's impulse response is, for the same values of M and N, identical to the filter used for decimation using the DCT. It takes the form given by equation 2. The term $\varepsilon(k)$ is used to include the effect of any multiplicative modifications in the DCT domain. In this paper, $\varepsilon(k)$ is equal to 1 for each value of k used in the summation. The term $[\alpha(k)]^2$ is equal to 0.5 for k = 0, and 1 otherwise.

$$h_{LP1}(p) = \sqrt{\frac{2}{M}}\sqrt{\frac{2}{N}}\,\frac{1}{2}\sum_{k=0}^{N-1}[\alpha(k)]^2\varepsilon(k)\cos\left[\frac{2\pi k}{2M}\left[p+\frac{1}{2}-\frac{M}{2N}\right]\right] \qquad \text{for } p = 0..2M-1 \qquad 2.$$

Figure 3 illustrates the multirate structure used to implement block based phase shifting filters. This block diagram includes both the filtering associated with the MPEG implementation of half pixel motion compensation, and filtering which incorporates the frequency scalable approach for generating a reduced resolution signal. The purpose of this multirate structure is to manipulate the low resolution decoder's prediction signal in such a manner that the difference between these predictions coefficients transformed using a 4x4 DCT, and the lower 4x4 transform coefficients of the encoder's prediction signal, are minimised.

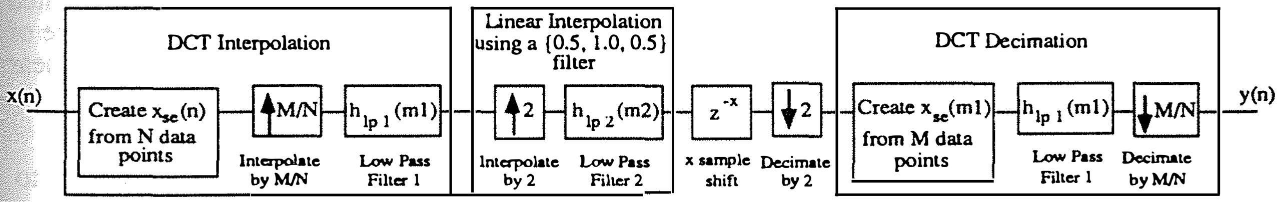

Figure 3 A Block Based Multirate Structure For Realising a Fixed Delay of xN/2M Samples.

The DCT does not perform well as an interpolator. Signal leakage resulting from filtering across the block boundaries of the symmetrically extended periodic sequence, generate the well known edge effects associated with the DCT. It is well known that these artefacts can be reduced if the system is over sampled. This is achieved by incorporating overlap between adjacent transforms. This technique can be incorporated in the block based filters proposed in this paper by giving M and N of figure 3, the values of ten and five respectively. If the value of N used in the design of the block based filters is restricted to one greater than the size of the transform of the lower layer, the problem of extracting a prediction signal at a frame boundary does not occur. The multirate structure of figure 3 then simplifies to a four by five filter coefficient matrix for each value of sub-pixel precision used by the low resolution decoder. The filters are listed in table 1[4].

1.00000	0.00000	0.00000	0.00000	0.00000	zero pel
0.00000	1.00000	0.00000	0.00000	0.00000	
0.00000	0.00000	1.00000	0.00000	0.00000	
0.00000	0.00000	0.00000	1.00000	0.00000	
0.74662	0.32068	-0.09955	0.04745	-0.01521	quarter pel
-0.06633	0.81837	0.31055	-0.09000	0.02740	
0.03404	-0.10291	0.82367	0.30291	-0.05765	
-0.02044	0.05770	-0.11055	0.83752	0.23576	
0.49324	0.64137	-0.19910	0.09490	-0.03042	half pel
-0.13267	0.63675	0.62111	-0.18000	0.05481	
0.06809	-0.20582	0.64721	0.60582	-0.11531	
-0.04088	0.11541	-0.22111	0.67504	0.47153	
0.27970	0.78098	-0.09038	0.04371	-0.01401	three quarter pel
-0.08794	0.34668	0.80025	-0.08399	0.02500	
0.04778	-0.12267	0.33441	0.79355	-0.05308	
-0.02979	0.07447	-0.12732	0.35939	0.72325	

Table 1 The Block Based Quarter Pel Precision Phase Shifting Filters

4 SIMULATIONS AND RESULTS

A simulation was performed comparing the effectiveness of the bi-linear, the minimum drift and the five tap block based phase shifting filters. Fifty eight frames of the sequence "Mobile and Calendar" was coded at 4 MBits/s having an Intra Frame repetition rate equal to 12 and Prediction Frame repetition rate equal to 3 (M=3 and N = 12). The coder simulated was the two layer, open loop, frequency scalable coder of the test model [1]. Objective results for the decoded low resolution image sequence using different phase shifting filters are presented in table 2.

Quarter Pel Filters used in low resolution decoder	Low Resolution PSNR
Bi-linear filters	27.90 db
Minimum drift filters	30.60 db
Five tap block based filters	30.33 db

Table 2 Decoded Low Resolution Objective Performance for the Sequence Mobile and Calendar.

5 CONCLUSION

This paper has presented the design procedure used in optimising quarter pel precision phase shifting filters for use in a non-embedded open loop frequency scalable coder. The filters maximise the quality of a decoded low resolution video sequence. Results have been presented comparing different quarter pel phase shifting filters. It has been shown that the performance of the block based shift varying filters proposed in this paper is significantly better than the bi-linear filters proposed in [1], and approach that of the more complex minimum drift filters.

6 ACKNOWLEDGEMENTS

The permission of the Director - Applied Research & Development, Telecom Australia, to publish this paper is acknowledged.

7 REFERENCES

[1] Test Model Editing Committee, "Test Model 5", ISO-IEC/JTC1/SC29/WG11, MPEG 93/457 March 1993.

[2] D.J. Goodman, "Embedded DPCM for Variable Rate Transmission," *IEEE Transactions on Communications*, VOL. COM 28, NO. 7 pp. 1040-1046, July 1980.

[3] R. Mokrys and D. Anastassiou, "Insight into the problem of Error Drift in Frequency Scalability MPEG/VIDEO", ISO-IEC/JTC1/SC29/WG11, MPEG 92/466 Sept. 1992.

[4] A.W. Johnson, T. Sikora and T. K. Tan, "Scalable Coding Employing Block Based Quarter Pel Precision Scale_4 Interpolation Filters", ISO-IEC/JTC1/SC29/WG11, MPEG 93/120 January 1993.

[5] A. W. Johnson, J. P. Princen, M. Chan and M. Biggar, "Rational Value Decimation Using the DCT for Layered Coding", *1991 Picture Coding Symposium*, Tokyo, Japan, 4-6 September 1991, pp. 53-56.

[6] K. N. Ngan "Two-dimensional transform domain decimation techniques", *Proc. 1986 IEEE Int. Conf. Acoust. Speech Signal Processing*. pp. 1001-1004, April 1986.

[7] M. Iwahashi, "Motion Compensation Technique in Scaling Decoder for Drift Reduction", ISO-IEC/JTC1/SC29/WG11, MPEG 92/556, November 1992.