BIORTHOGONAL WAVELETS FOR INTRA-FRAME VIDEO CODING

Dr. T. Morris, Mr. D. Britch
Department of Computation, UMIST,
P. O. Box 88, Manchester, M60 1QD, U.K.
t.morris@co.umist.ac.uk dbritch@co.umist.ac.uk

Abstract: *The choice of filter bank in wavelet video compression is a crucial issue that effects both video quality and compression ratio. A series of biorthogonal, symmetric wavelet filters of differing length were evaluated by compressing and decompressing a number of standard video test sequences, using three sets of quantisation parameters. A video codec quality analyser (CQA) is used to assess the subjective quality of the coded stream with respect to the uncoded one. Experiments show that the familiar FBI wavelet offers the best compromise between high quality reconstruction and low bitrate, with the Mean wavelet being a suitable alternative for very-low bitrate applications.*

Keywords*:* video coding, wavelets.

1. Introduction

Much research effort has been expended in the area of wavelet-based compression, with the results indicating that wavelet-based approaches outperform DCT-based techniques (Li *et al.*, 1995; Shapiro, 1993; Antonini *et al.*, 1992; Albanesi *et al.*, 1992; Lewis *et al.*, 1992). However, it is not completely clear which wavelets are suitable for video compression. Wavelets implemented using linear-phase filters are generally advantageous for image processing because such filters preserve the location of spatial details.

A key component of an efficient video coding algorithm is motion compensation. However at the present time it is too computationally intensive to be used in software video compression. A number of low end applications therefore use motion-JPEG, in essence frame by frame transmission of JPEG images (Wallace, 1994), with no removal of interframe redundancy.

In considering the effectiveness of a video codec, two factors need to be considered. One is the how much compression is achieved and the other is the quality of the subsequently decoded data. The choice of wavelet effects both factors. The degree of compression is easily measured by comparing the volumes of the coded and raw data. The quality of the decoded data is much more difficult to quantify. It is well known that small errors randomly distributed over the whole frame reflect badly in objective measures, but they often correspond to invisible degradation of frame quality. On the other hand, relatively large errors concentrated in one region of a frame give a high PSNR while the subjective assessment is low because of a corruption of a particular region of the picture. This observation therefore calls into question the usefulness of objective measures.

In recent years, the disadvantages of both subjective and objective measures have motivated researchers to search for a better method of evaluating the quality of video sequences which have undergone lossy compression (Hosaka, 1986; Algazi *et al.*, 1992). A codec quality analyser (CQA) has been developed which aims to provide a consistent bias-free single quality measure, which correlates with a subjective assessment of video data (Morris *et al.*, 1999).

This paper investigates how biorthogonal wavelets effect the subsequent quality and size of the reconstructed data, using a wavelet based video codec that we have developed. Test sequences were compressed with the codec, and the results obtained indicate that the choice of wavelet greatly influences the quality of the compressed data and its size.

The paper is organised as follows. The compression algorithm is described in Section 2. We then present and discuss our experimental results in Section 3. Section 4 outlines the developments we intend to pursue in the future.

2. Algorithm

An overview of the algorithm is shown in Figure 1. The encoder takes a colour video sequence with arbitrary sized dimensions as input, and transforms it from RGB representation to YCbCr representation, so that the data is represented in a manner more suitable for compression. Zhang (Zhang *et al.*, 1997), has shown that luminance signals contain more than 60% of the total energy of the original signal, with the chrominance signals each having less than 20%. Therefore, the data is subsampled to a ratio of 4:2:2.

Each frame is then successively decomposed by a symmetrical wavelet transform (SWT). The wavelet coefficients are then quantised by a Lloyd-Max scalar quantiser, which outputs quantised wavelet coefficients. These coefficients are subsequently entropy encoded by zerotree coding and Huffman coding, producing a single bitstream of data that contains the compressed frame data, along with header information that allows the data to be decoded.

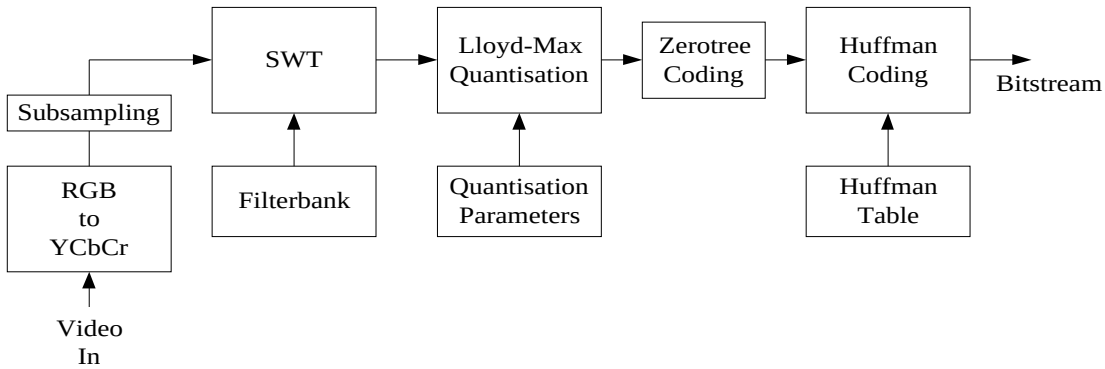


Figure 1 Simplified encoder diagram

The decoder simply performs the relevant processes in reverse order, to produce the reconstructed frames.

2.1. The Wavelet Transform

The DWT is implemented using a two-channel perfect reconstruction linear phase filter bank (Strang *et al.*, 1997). Symmetric extension techniques are used to apply the filters near the frame boundaries; an approach that allows transforming images with arbitrary dimensions. A detailed treatment of symmetric wavelet transform methods is given in (Brislawn, 1996). Five iterations of the SWT are carried out per frame, resulting in a sixteen subband decomposition.

2.2. Quantisation

Uniform quantisation is a simple but effective quantisation solution, assigning equally spaced reconstruction levels on a fixed interval. An optimum quantiser (Max, 1960) is achieved by

ensuring that the reconstruction levels are denser in regions of high probability. The optimum reconstruction level for a decision boundary is at the centroid (with respect to the source probability distribution) of the decision boundary, and forms the basis of the Lloyd (Lloyd, 1982) algorithm for quantiser optimisation. Conditions for optimal decision boundaries b_i ($b_1 = -\infty$, $b_{L+1} = \infty$) and optimal reconstruction levels r_i are (Jayant *et al.*, 1984):

$$b_i = \frac{r_{i-1} + r_i}{2} \qquad i \in \{2,3,\ldots,L\} \tag{1}$$

$$r_i = \frac{\int_{b_i}^{b_{i+1}} x f_x(x)dx}{\int_{b_i}^{b_{i+1}} f_x(x)dx} \qquad i \in \{1,2,\ldots,L\} \tag{2}$$

Given an initial estimate of the reconstruction levels, improved estimates of decision boundaries and reconstruction levels are calculated by iteratively applying equations (1) and (2); this converges to a global optimum (in the MSE sense) for the Uniform probability density function.

It is well known that to achieve good results, a separate quantiser should be designed for each scale, taking into account both the properties of the Human Visual System and the statistical properties of the scale's coefficients. Experimentation found that good results could be obtained by quantising subbands 0 to 3 with one set of parameters, subbands 4 to 12 with another, and discarding subbands 13, 14, and 15 altogether.

Rather than outputting a set of indices, the quantiser outputs a set of quantised wavelet coefficients, which are approximations to the original wavelet coefficients. Therefore, no inverse quantisation is required during the decoding process.

2.3. Entropy Encoding

Following scalar quantisation of the subbands within a frame, the quantised wavelet coefficients output by the quantiser are entropy-encoded by embedded zerotree coding (Shapiro, 1993) and Huffman coding.

Along with the compressed frame information, the bitstream contains information that consists of video sequence dimensions, parameters for the scalar quantisers and tables for the Huffman coder.

2.4. Codec Quality Analyser

CQA is designed to give a subjective measure of the degradations introduced into a video stream by the coding process. This is achieved by measuring the subjectively important differences between the original and compressed video streams. Full details of these measurements can be found in (Morris *et al.*, 1999). The measurements are then combined using a neural network to give a single value that reflects the subjective quality of the video stream. A value of 1 indicates a sequence with an objectionable amount of degradation, while a value of 5 indicates a sequence with no perceptible degradation.

3. Results and Discussion

Comparative tests were carried out on different biorthogonal wavelets using the proposed coding scheme on various test sequences. Video sequences of different resolutions were used, all at 30 fps and 24 bits per pixels. Each sequence was compressed using three different quantisation parameters (Test1 = 5 levels, Test2 = 20 levels for subbands 0 to 3, and 15 levels for subbands 4 to 12, Test3 = 40 levels for subbands 0 to 3, and 30 levels for subbands 4 to 12). In each test, the quantiser was trained using a maximum of 20 iterations and a threshold of 0.01, meaning that when the distortion between iterations drops below 1%, the training algorithm terminates. Table 1 shows the coding results for the Claire sequence (360x288 pels, 168 frames) using the proposed method, together with the quality rating given by CQA.

Table 1 Compression and quality results for Claire

Name	Taps		Compression Ratio			CQA		
	H_0	G_0	Test 1	Test 2	Test 3	Test 1	Test 2	Test 3
Brislawn	10	10	221:1	14:1	14:1	1	4	5
FBI	9	7	242:1	21:1	16:1	1	5	5
UCLA	13	11	201:1	17:1	16:1	1	4	5
Bi Haar	2	6	60:1	18:1	16:1	1	2	3
Lazy	1	1	23:1	16:1	13:1	1	2	2
Mean	10	10	428:1	17:1	17:1	1	4	5
5/3	5	3	61:1	17:1	14:1	1	4	4
Bi-Coiflet	17	11	167:1	17:1	15:1	2	5	5

The surprise result obtained here is that the Mean wavelet (Munro *et al.*, 1997) produces an extremely high compression ratio when using a low number of reconstruction levels. This makes this wavelet extremely desirable for very-low bitrate applications. Furthermore, the results indicate that the number of taps in the filter have a huge influence on compression ratio and quality. Of the wavelets tested here, those with a small number of taps produced a relatively low compression ratio, while the wavelet with the largest number of taps produced the highest quality reconstruction overall.

Figure 2 shows a frame of Claire compressed using three different wavelets, all using the same quantisation parameters (5 reconstruction levels, 20 iterations and a threshold of 0.01). As expected the Bi-Haar wavelet introduces a significant amount of blocking. Although the Mean wavelet produces the highest compression ratio, the resulting quality is noticeably lower than that produced by the FBI wavelet, which produces the best compromise between reconstruction quality and bitrate.

4. Conclusions and Future Work

This paper has shown that the choice of wavelet greatly influences the size and quality of the reconstructed data. We have identified the familiar FBI wavelet as producing the highest quality reconstruction for low compression ratios, although the Mean wavelet is a suitable alternative for very-low bitrate applications. Furthermore, the FBI wavelet allows codec performance to degrade gracefully, with none of the blocking which is present in MPEG coding schemes.

Future research efforts will concentrate on two items. Firstly, we aim to investigate how different orthogonal wavelets effect the compression ratio and quality of the decompressed data. Secondly, work is in progress on a segmentation algorithm for locating objects within a

scene. It is envisaged that this segmentation algorithm will be combined with the video codec to yield an increased compression ratio.

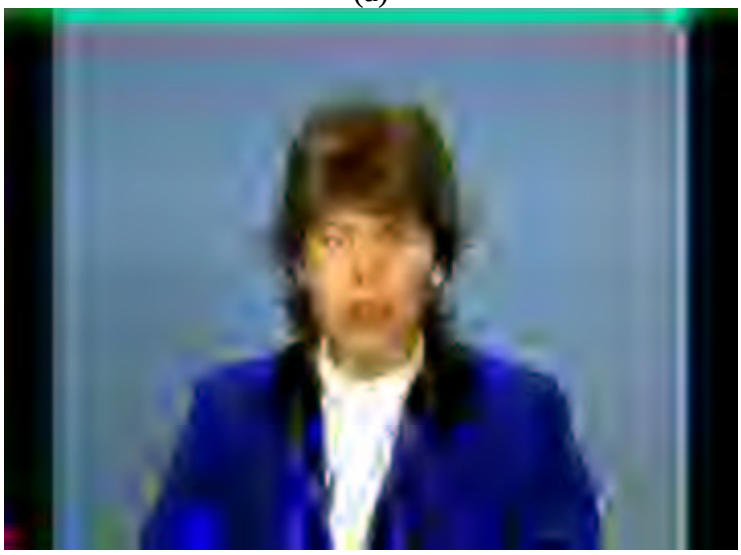
5. Acknowledgements

D. Britch is supported by the UK Engineering and Physical Sciences Research Council (EPSRC). Thanks go to Brendt Wohlberg for his assistance with the Lloyd-Max quantiser.

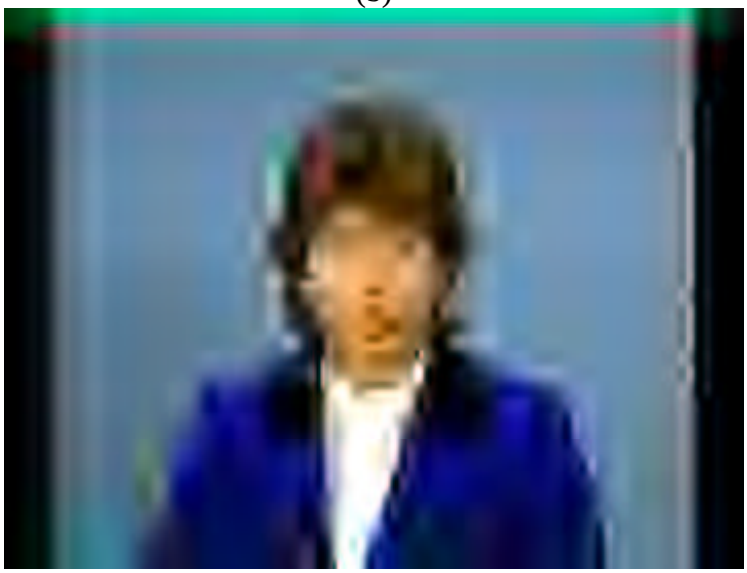
6. References

1. Albanesi, M. G., de Lotto, I., and Carrioli, L. (1992) "Image compression by the wavelet decomposition", Signal Processing, Vol. 3, No. 3.
2. Algazi, V. R., Kato, Y., Miyahara, M., and Kotani, K. (1992) "Comparison of image coding techniques with a picture quality scale", Proc SPIE Applications of Digital Image Processing, XV: 1771, 1992.
3. Antonini, M., Barlaud, M., Mathieu, P., and Daubechies I. (1992), "Image coding using wavelet transform", IEEE Trans. Image Processing, Vol. 1, No. 2.
4. Brislawn, C. (1996), "Classification of nonexpansive symmetric extension transforms for multirate filter banks", Applied and Computational Harmonic Analysis, Vol. 3, pp. 337-357.
5. Hosaka, K. (1986), "A new picture quality evaluation method", PCS'86, Tokyo, Japan, April 1986.
6. Jayant, N. S., and Noll. P. (1984), Digital Coding of Waveforms, Prentice-Hall Signal Processing Series, Prentice-Hall, Englewood Cliffs, NJ, USA.
7. Lewis, A. S., and Knowles, G. (1992), "Image compression using the 2D wavelet transform", IEEE Trans. Image Processing, Vol. 1, No. 2.
8. Li, W., and Zhang, Y. Q. (1995), "Vector-based signal processing and quantization for image and video compression", Proceedings of IEEE, Vol. 83, Feb 1995, pp. 671-681.
9. Lloyd, S. P. (1982), "Least squares quantisation in PCM", IEEE Trans. Information Theory, IT-28, pp. 129-136.
10. Max, J. (1960), "Quantising for minimum distortion", IEEE Trans. Information Theory, IT-6, pp. 7-12.
11. Monro, D. M., and Sherlock, B. G. (1997), "Space-frequency balance in biorthogonal wavelets", IEEE Intl. Conf. on Image Proc., Vol. 1, pp. 624-627.
12. Morris, T., Angus, K., Butt, R., Chilton, A., Dettman, P., and McCoy, S. (1999), "CQA – Subjective video quality analyser", British Machine Vision Conference, 1999, Nottingham, UK.
13. Shapiro, J. (1993), "Embedded image coding using zerotrees of wavelet coefficients", IEEE Trans. Signal Processing, Vol. 41, pp. 3443-3463.
14. Strang, G., and Nguyen, T. (1997), Wavelets and Filter Banks, Wellesley-Cambridge Press, Wellesley, MA, USA.
15. Wallace, G. K. (1994), "The JPEG still picture compression standard", Comm ACM, Vol. 34, No. 4, pp. 30-44.
16. Zhang, T. Q., and Zafar, S. (1997), "Wavelet-based video compression", in Video Data Compression for Multimedia Computing: Statistically Based and Biologically Inspired Techniques, ed. by H. L. Hua, Kluwer Academic Publishers, Netherlands, Chapter 1.

(a)



(b)



(c)

Figure 2 Comparison of wavelet performance on the sequence "Claire": (a) Bi-Haar, 60:1 (b) FBI, 241:1, (c) Mean, 428:1. The resolution is 360 x 288 pixels per frame.