Object-based Intra-frame Wavelet Video Coding

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

Abstract

An object-based wavelet based video codec has been designed and implemented. The codec takes a colour video sequence of arbitrary size as input and performs intra-frame object-based compression on the sequence. Video sequences are segmented using an optic flow based algorithm, before being transformed with a symmetrical wavelet transform, prior to Lloyd-Max quantisation and entropy encoding. A video codec quality analyser (CQA) is used to assess the subjective quality of the coded stream with respect to the uncoded one in order to provide a single quality measure that correlates with a subjective assessment of the data. Experiments using standard test sequences show that a higher compression ratio is achieved than obtainable using a similar non object-based codec, at the same quality level.

1. Introduction

Object-based coding has been suggested as an alternative to conventional pixel-based coding [1], with segmentation performing the task of object identification. However, segmentation cannot be seen as a tool that is easily applied to generic scene content. Substantial gains can only be achieved if large parts of a scene contain regions that can be extracted from the remaining parts of the scene.

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 [2], 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 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 very 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 [3, 4]. Such a method has been produced and 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 [5].

This paper describes an object-based wavelet video codec that we have developed. It is designed to take a colour video sequence as input, and perform intra-frame object-based compression on the sequence. Test sequences were compressed with the codec, and the results obtained indicate that the object-based codec out-performs the non object-based codec in terms of compression ratio, at the same level of quality.

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 evaluates the algorithm and outlines its future development.

2. Algorithm

An overview of the algorithm is shown in Figure 1. The algorithm comprises five main processes: optic flow based segmentation, wavelet decomposition of the detected objects, scalar quantisation of the wavelet coefficients, zerotree coding of the quantised wavelet coefficients followed by lossless entropy coding of the zerotree data.

The encoder takes a colour video sequence with arbitrary sized dimensions as input, and codes the first frame using a non object-based approach [6, 7]. Horn and Schunk's optic flow algorithm [8] is then used to segment the remainder of the input sequence in order to separate the sequence into foreground and background objects. A minimum bounding box is used to obtain the area that the foreground object occupies, and then the bounding box is extended by five pixels on every side in order to store a small amount of surrounding background with the object. The extra pixels are stored for use in updating the background when the object moves. Along with the object, its location in the frame is also stored.

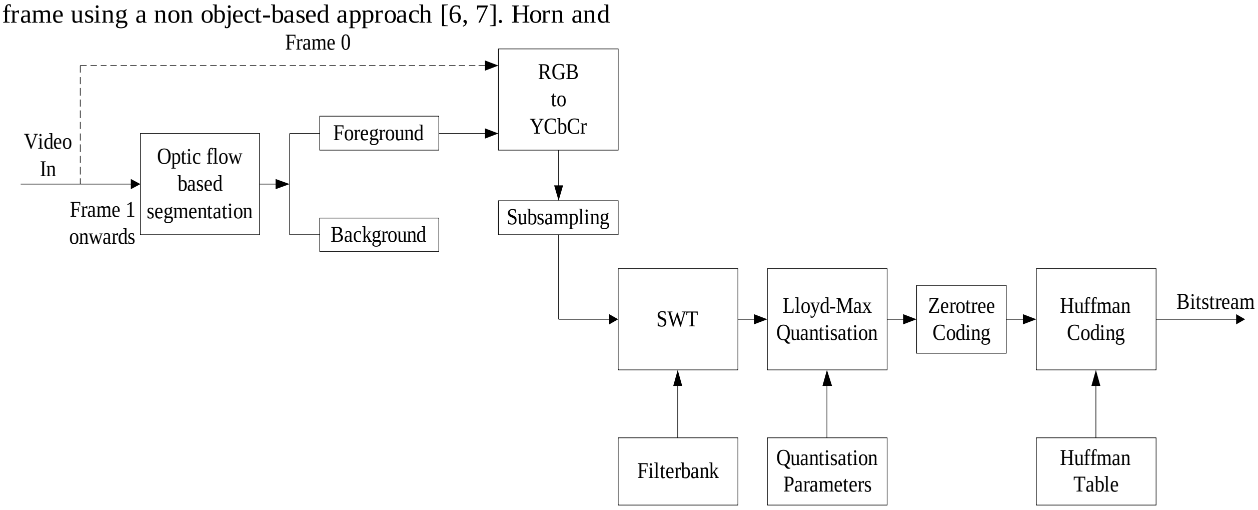

Figure 1. Simplified encoder diagram

The foreground object is then transformed from the RGB representation to the YCbCr representation, so that the data is represented in a manner more suitable for compression. Zhang and Zafar [9] have shown that luminance signals contain more than 60% of the total energy of the original signal, with the two chrominance signals each having less than 20%; therefore the data undergoes 4:2:2 subsampling.

The foreground objects in each frame are then successively decomposed by a symmetrical wavelet transform (SWT). The wavelet coefficients are then quantised by a Lloyd-Max scalar quantiser that 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.

The decoder extracts the information needed in the decoding process from the compressed data. An entropy decoder reconstructs the quantised wavelet coefficients which are then inverse transformed before being upsampled and converted from YCbCr format to RGB, to produce the reconstructed frame.

2.1. The Wavelet Transform

The DWT is implemented using a two-channel perfect reconstruction linear phase filter bank [10]. Symmetric extension techniques are used to apply the filter bank near the frame boundaries; an approach that allows transforming frames with arbitrary dimensions. A detailed treatment of symmetric wavelet transform methods is given in [11]. Five iterations of the SWT are carried out per frame, resulting in a sixteen-subband decomposition.

The filter bank chosen was the familiar biorthogonal 9/7 filter constructed by Cohen, Daubechies, and Feauveau [12]. This filter bank was selected based on its ability to produce reconstructed images with superior quality to those generated by other filters [13].

2.2. Quantisation

Uniform quantisation is a simple but effective quantisation solution, assigning equally spaced reconstruction levels on a fixed interval. An optimum quantiser [14] 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 algorithm for quantiser optimisation [15]. Conditions for optimal decision boundaries b_i ($b_1 = -\infty$, $b_{L+1} = \infty$) and reconstruction levels r_i are [16]:

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

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

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. Through experimentation it was found that excellent 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. Figure 2 shows the location of the subbands within a frame.

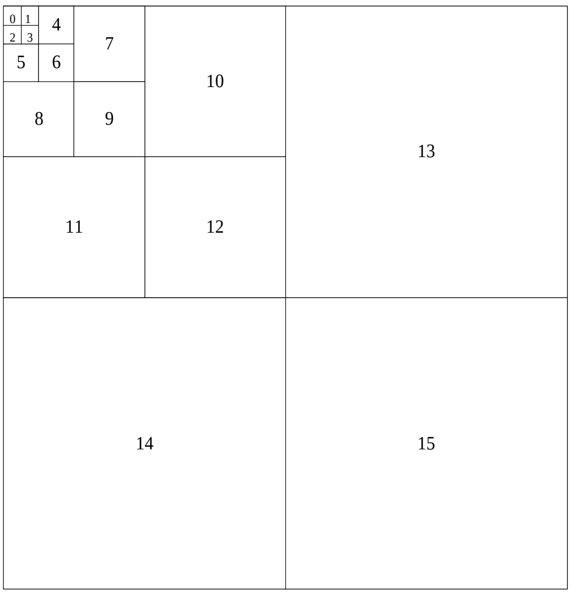

Figure 2. Frequency support of the subbands

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 [17] 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 [5]. 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 between the proposed object-based coding scheme and the non object-based coding scheme [6, 7] on various test sequences. Video sequences of different spatial resolutions were used, all at 30 fps and 24 bits per pixels. 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.

In Table 1, coding results for five video test sequences are presented in terms of the compression ratio that the non object-based wavelet codec achieved, together with the quality rating given by CQA.

Table 1. Wavelet results (non object-based)

Sequence	Resolution	Ratio	CQA
Claire	360 x 288	44:1	4
Missa	360 x 288	22:1	4
Susie	720 x 480	27:1	4
Mobile	720 x 576	19:1	4
Football	720 x 486	21:1	4

Table 2 shows the coding results for the same sequences using the proposed object-based method.

Table 2. Wavelet results (object-based)

Sequence	Resolution	Ratio	CQA
Claire	360 x 288	82:1	4

Missa	360 x 288	32:1	4
Susie	720 x 480	28:1	4
Mobile	720 x 576	146:1	4
Football	720 x 486	24:1	4

The results show a significant improvement over the non object-based approach. In all five video sequences the compression ratio has increased, with the quality of the decoded stream remaining the same.

Figure 3 shows frames from the Claire, Missa and Mobile sequences coded using the non object-based approach, and the proposed object-based approach. The frames coded with the non object-based method are of obvious lower quality, with degradation becoming evident in the frames. However, the frames coded using the proposed object-based method have no perceptible degradation.

In the Susie and Football sequences, the improvement in compression ratio is negligible, but in the remaining three sequences the results show the substantial benefits that object-based coding can produce.

4. Conclusions and Future Work

This paper has presented a baseline object-based wavelet colour video codec, which out-performs a similar non object-based codec in terms of compression ratio, at the same subjective quality level. The results show the benefits that object-based coding can achieve when the scenes contain suitably sized regions that can be extracted from the remaining parts of the scene.

Future research efforts will concentrate on a number of items. Firstly we aim to increase the compression ratio itself by replacing the scalar quantiser with a vector quantiser.

Secondly, to allow comparison of this video codec with others in existence it will be necessary to modify the codec so that video sequences can be coded at user specified bit rates.

Thirdly, in order to demonstrate the power of wavelet-based coding it will be necessary to implement this work as a high-performance real-time video codec.

Finally, although it was possible to accurately detect the boundaries of objects located in the video test sequences, it is unlikely that this would be possible in video sequences more representative of the real world. Therefore, a substantial amount of research needs to be performed into the topic of motion segmentation.

5. References

[1] T. Sikora, "MPEG digital video-coding standards", *IEEE Signal Processing Magazine*, September 1997.

[2] G. K. Wallace, "The JPEG still picture compression standard", *Comm ACM*, Vol. 34, No. 4, pp. 30-44, April 1994.

[3] K. Hosaka, "A new picture quality evaluation method", *PCS '86*, Tokyo, Japan, April 1986.

[4] V. R. Algazi, Y. Kato, M. Miyahara, and K. Kotani, "Comparison of image coding techniques with a picture quality scale", *Proceedings of the SPIE: Applications of Digital Image Processing*, XV: 1771, 1992.

[5] T. Morris, K. Angus, R. Butt, A. Chilton, P. Dettman, and S. McCoy, "CQA – Subjective video quality analyser", *British Machine Vision Conference*, Nottingham, UK, 1999.

[6] T. Morris, and D. Britch, "Biorthogonal wavelets for intra-frame video coding", *1st International Workshop on Image and Signal Processing and Analysis*, Pula, Croatia, June 14th-16th, 2000.

[7] T. Morris, and D. Britch, "Intra-frame wavelet video coding", *2nd International Symposium on Video Processing and Multimedia Communications*, Zadar, Croatia, June 28th-30th, 2000.

[8] B. K. P. Horn, and B. G. Schunk, "Determining optical flow", *Artificial Intelligence*, Vol. 17, pp. 185-203, 1981.

[9] T. Q. Zhang, and S. Zafar, "Video Data Compression for Multimedia Computing: Statistically Based and Biologically Inspired Techniques", Chapter 1, ed: H. L. Hua, *Kluwer Academic Publishers*, 1997.

[10] G. Strang, and T. Nguyen, "Wavelet and Filter Banks", *Wellesley-Cambridge Press*, Wellesley MA, 1997.

[11] C. Brislawn, "Classification of nonexpansive symmetric extension transforms for multirate filter banks", *Applied and Computational Harmonic Analysis*, Vol. 3, pp. 337-357, March 1996.

[12] A. Cohen, I. Daubechies, and J. C. Feauveau, "Biorthogonal bases of compactly supported wavelets", *Communications on Pure Applied Mathematics*, Vol. 45, pp. 485-560, 1992.

[13] J. D. Villasenor, B. Belzer, and J. Liao, "Wavelet filter evaluation for image compression", *IEEE Transactions on Image Processing*, Vol. 4, No. 8, pp. 1053-1060, August 1995.

[14] J. Max, "Quantising for minimum distortion", *IEEE Transactions on Information Theory*, IT-6, pp. 7-12, 1960.

[15] S. P. Lloyd, "Least squares quantisation in PCM", *IEEE Transactions on Information Theory*, IT-28, pp. 129-136, March 1982.

[16] N. S. Jayant, and P. Noll, “Digital Coding of Waveforms: Principles and Applications to Speech and Video”, *Prentice-Hall*, NJ, USA, 1984.

[17] J. M. Shapiro, “Embedded image coding using zerotrees of wavelet coefficients”, *IEEE Transactions on Signal Processing*, Vol. 41, pp. 3445-3462, 1993.

Figure 3. Object-based coding and non object-based coding: Claire compressed at 82:1: (a) non object-based, (b) object-based, Missa compressed at 32:1: (c) non object-based, (d) object-based, Mobile compressed at 146:1: (e) non object-based, (f) object-based.