**Chapter 7: Wavelet Transform Methods.**

**7.1 Introduction.**

The basic form of the wavelet transform is given in equation 7.1 :

 (7.1)

where 

 is the one dimensional signal,

 is the mother wavelet or basis function,

and  is the transform coefficient for n, k.

The wavelet works by splitting the signal into different sized subbands which allows a strong definition in both the time and frequency domains. In a normal transform method the signal is represented by several basis functions over the same ‘time’ length as the signal. In a wavelet the signal is repeatedly split into a subbands where the length of the signal is contracted by 2. This gives a good representation of large scale features in a signal as well as good control of the local features.

Fixed length transforms, such as the DCT, suffer from ringing because high frequencies are not localised, this is solved in the wavelet methods. The methods developed for wavelet transforms were done in the early 1980’s by Daubechies [30] , Chui [53] and Jawerth and Sweldens [54]. The application of wavelet transforms will be examined in the next section, first using the Shapiro Embedded Zero-Tree Wavelet (EZW) method and then using the optimally quantised wavelet (expanding on ideas from chapter 5). Finally, the results from these methods will be compared with the DCT.

**7.2 Application of Wavelet Transform.**

The wavelet is applied to one dimensional signals by using two matched finite impulse response filters (HH, HL) . There are also two inverse filters (FH, FL), which reverse the transform. These are applied to the signal as shown in figure 7-1.



Figure 7-1. Diagram showing one scale of wavelet filters.

This is called one scale of a wavelet transform and is perfectly reconstructing, provided the calculations are of infinite precision and which ensures that the scaling of the transform is correct, as shown in equation 7.2:

 (7.2)

where *h(i)* are the filter coefficients of the Filters HL and HH.

To increase the number of scales, and hence the usefulness, of the wavelet transform the coefficients produced by the low pass filter are passed through the same filter bank, as shown in figure 7-2. This process can be repeated producing multiple scales, but in images the number of scales is usually limited to 4 or 5.



Figure 7-2. Diagram showing 3 scales of sub-band filtering.

To apply the wavelet filters to an image the filters are simply passed first in the x direction to form an image as shown in figure 7-3 (a) and then the filters are applied to the low pass x coefficients in the y direction to form figure 7-3 (b). Further passes of filters on the low low section of the image produce more complicated wavelet filter images 7-3 (c). The wavelet has two advantages over standard block methods:

* The coefficients localise any quantisation effects, hence there is no ringing.
* There are no block boundaries, since the filters stretch over the sides of the wavelet tree blocks produced.



(a) (b) (c)

Figure 7-3. Diagram of image subbands.

The wavelet filtering imposes an effective blocksize of , where n is the number of scales of wavelet filters applied. This effective block size is imposed since every coefficient in the lowest subband (LLLL) is related to other pixels in a  block as shown in figure 7-4.



Figure 7-4. Diagram showing the nature of wavelets and the relation between coefficients.

**7.3 Shapiro Embedded Zero Tree Wavelet (EZW) Compressor.**

The EZW compressor [36] is one of the best compressors available for wavelets at this time. Enhancements were made to the Shapiro method by Said and Pearlmann [37], but we are only interested in the basic EZW here. The EZW exploits the tree-like nature of the wavelet transform to good effect and uses embedded quantisation to represent the wavelet coefficients. Since it is embedded, the compressed data of an EZW can be stopped at any time during transmission and a picture can be reconstructed that is optimal for that bit rate with the EZW. It means that the EZW has useful applications in the scaleable bit rate field, where multiple users accept the same bit stream but decode it to different compressions.

The operation of a Shapiro coder is described in the following sections and the rate distortion performance it produces is calculated. Since the coefficient symbols produced by the EZW are low in entropy it was necessary to losslessly encode them using arithmetic compressor. Run length Huffman compression can be used but is not as efficient.

**7.3.1 Wavelet Tree Representation.**

Although the wavelet transform is not applied in blocks it can be considered to form collections of related pixels or ‘tree blocks’. Figure 7-4 shows how one pixel in the top subband relates to all the others in the other subbands. These pixels are related since they provide systematic improvements on the approximation of a ‘tree block’.

This relationship is exploited by the EZW system by using zero trees. The range of coefficient values that are allowed in each subband is limited by the action of the wavelet. The ranges of coefficients in the original image was 0🡪255 (256 range). After the first application of the wavelet filter, figure 7-3(b), all the values have a dynamic range of 512 although the high passed bands are zero centred (-128🡪127). This process of increasing the range of the filter subbands (controlled by equation 7.2) continues to the top subband which is usually 12 bits (in comparison with its original 8 bits). The effect of this image limiting means that a lot of embedded quantisation values are zero for the first few iterations of the EZW. This means that the symbols have a very low entropy which requires an arithmetic compressor.

**7.3.2 Embedded Quantisation.**

The EZW uses binary embedded quantisation, where all the quantisation steps are powers of two, and so binary operations can be used to implement them. The Shapiro method uses a binary threshold to quantise the coefficients. This threshold starts at half the size of the maximum coefficient in the wavelet transform and then is decreased in size, dividing by two, progressively increasing the quantisation, as shown in equation 7.3 :

 (7.3)

where .

Using a threshold (t\_hold) is a fast way of implementing binary quantisation and the coefficients are compared to t\_hold and -t\_hold to cover negative numbers.

The data in the Shapiro method is handed in the form of bit planes that are produced after every threshold operation. This is demonstrated in figure 7-5 where five data items are split into bit planes.



Figure 7-5. Diagram showing five data items represented in binary. Its is important to see that there is no information about a data item, in a bit plane, until it becomes significant .

The innovative part of the Shapiro method is the compression of these bit planes. The bit planes are coded with the following symbols:

PSIG - positive significant coefficients

NSIG - negative signal

ISIG - isolated zero

ROOT - zero tree root.

PSIG and NSIG are used to describe when a coefficient is above the threshold and are relatively uncommon symbols. When a coefficient has been declared as significant, it is assured to be significant at every lower bit plane. When a coefficient is marked as significant, its sign is removed and after each proceeding bit plane has been coded the value of the significant coefficient is refined by transmitting the appropriate binary value. This can be illustrated in relation to figure 7-5. Coefficient A is marked as significant on bit plane 10 and in the proceeding bit planes (one bit per bit plane) the following refine data is passed:

001011011

The zero tree ROOT symbol is the core of the Shapiro method. Each coefficient in the wavelet filtered image is associated with coefficients in lower subbands. If a parent coefficient is zero tree there is a high probability that the children will be zero and hence a zero tree ROOT can be used to describe this as shown in figure 7-6. The ROOT symbol is very common and with the use of an arithmetic coder the Shapiro method is very efficient in rate distortion terms.



Figure 7-6. Diagram of Zero Symbols highlighting usefulness of ROOT.

ISIG represents an isolated zero, which can not be represented by a ROOT, it is not a very common symbol but it is critical to the operation of the method.

The embedded stream produced by the EZW can be decoded to any compression smaller that the target bit budget. The rate distortion produced is still optimal for the EZW method.

**7.3.3 Experiment: Calculate the rate distortion curves for the EZW coder**

The Goldhill test image was wavelet filtered to 4 scales using the UCLA 7/9 wavelet [55] and then the Shapiro EZW compressor was applied to the wavelet filtered image in order to compress it. The image was decompressed and the rate distortion curve produced is compared to JPEG in figure 7-7. It can be seen that the EZW outperforms JPEG at all compressions and the sections of Goldhill shown in figure 7-8 demonstrate (at 0.2 BPP) how much better the wavelet method looks. The EZW does not show any evidence of blocking artefact but a different type of artefact is present as shown in figure 7-9. This is a shot noise distortion of the picture and although it is not as visually displeasing as ringing it is still significant. It occurs because the EZW is forced to stop before completing an entire subbands hence some of the high frequency information is there while most of it is absent. This means that the artefact is not due to an error in the recovered image, rather it is an incomplete description of the detail.



Figure 7-7. Rate distortion curve comparing EZW with JPEG.



(a) JPEG (b) EZW

Figure 7-8. Sections of Goldhill comparing JPEG(a) to the EZW(b) at 0.2 BBP



Figure 7-9. Section from the top of the Goldhill (0.2 BBP) demonstrating shot noise.

1. **Optimal Quantisation of the Wavelet Transform**

The wavelet transform can be optimally quantised in much the same way as the DCT was in chapter 6. The block size (*N*) of the wavelet transform is dependant on the number of scales (n) the filters are applied to (*N = 2n*). In this section the processes described in chapter 5 are applied to wavelet filtered images. The optimal quantisation of the wavelet coefficients is tested in this chapter with three, four and five scales. The most common choice of scales for the Shapiro method is four although using five scales is not much different in rate distortion terms.

The coefficients generated by the wavelet transform need to be quantised for compression. In this case a linear quantiser was appropriate since multiple sets of quantisation need to be applied in the method and anything else would have been too slow. The aim of the compressor is to produce one quantisation table which can be used to quantise wavelet tree blocks (shown in figure 7-4) all over the image. This quantisation table should produce the optimal MSE for the exact number of bits required in the wavelet coder. The optimal compression method assumes that there is no connection between different coefficients in the wavelet transform but the EZW depends on these relations for its performance. It is important when using wavelets to exploit this relationship in order to improve the performance of the method. Unfortunately this is a contradiction for the optimal quantisation method.

The coefficients needed to be losslessly compressed after quantisation to produce the final compressed file. In the process to produce the multiple rate distortion curves it was necessary to apply a wide range of quantisation and a lot of data was produced. To avoid the overhead of losslessly compressing each data item the theoretical entropy was used to estimate the number of bits associated with a particular quantisation. The final set of quantised coefficients was compressed with an arithmetic compressor since the entropies indicated that a lot of symbols has less that one bit/symbol.

**7.4.1 Generation of the Rate Distortion Curves for the Equal Gradient method applied to the wavelet**

The equal gradient method described in chapter 5 discusses in detail how to finding the optimal number of bits or compression that should be applied to each coefficient set. The method uses a set of rate distortion curves (one for each coefficient set) but it does not deal with how the rate distortion curves are generated.

The application to the wavelet uses a linear quantisation method as in equation 7.4.

 (7.4)

where *cq* is the quantised wavelet coefficient,

*c* is the unquantised wavelet coefficient,

and *q* is the quantisation step.

It is necessary to generate a number of points on the rate distortion curves of each coefficient set in the wavelet transform. Although the separate subbands of the wavelet are very similar in nature, it was decided to treat each coefficient set separately and let the method decide they were similar by assigning a similar quantisation step. These points need to be distributed correctly between the quantisation limits so that information is not wasted.

The quantisation limits are:

*qn*  which is the quantisation step that completely truncates the all the data

and *q0*  which is the quantisation step that integerises the already integer wavelet (*q=1*).

Quantisation is not a linear process but the compressions produced are more linearly spaced if the quantisation values are chosen to vary as in equation 7.5:

 (7.5)

where *qi* is the quantisation step,

*i=0,1,2…n,*



and max is the maximum coefficient value in the set.

In this work n=10 was found to be satisfactory, by experiment, as described in chapter 5.

When the quantisation steps have been calculated they are applied to the appropriate wavelet coefficient set and a set of quantised coefficients are generated. The error in each quantisation step is evaluated as the sum of the squares of the errors between the quantised and unquantised coefficients. The number of bits required to store the quantised coefficients set is calculated from the theoretical entropy (equation 7.6) :

 (7.6)

where *pdd(x)* is the probability density distribution of symbol *x.*

The equal gradient method is then passed to a set of ‘support points’ to describe the rate distortion performance of each coefficient set. It returns the number of bits required to compress each coefficient set at the optimal point. Since this needs to be converted to quantisation steps to be useful, the data from the rate distortion curve calculations is used to linearly interpolate the number of bits to the correct quantisation.

**7.4.2 Experiment: Comparisons of Optimal Quantisation of the Wavelet** **Transform at different scales.**

The process described in chapter 5 and the previous section was implemented for the wavelet transform using the UCLA 5/7 wavelet filters. The method was applied with three, four and five wavelet transform scales, corresponding to 8x8,16x16 and 32x32 block sizes. An arithmetic lossless coder was used to compress the quantised coefficients and it was initialised from a store of PDFs to improve its performance.

The rate distortion curves are shown in figure 7-10 for the 3 different scales. It is clear that 4 scales is best for the wavelet transform method. The 3 scale implementation fails at about the same point as JPEG for much the same reason.



Figure 7-10. Rate distortion performance curves showing 3,4 and 5 wavelet transform scales using the optimal quantisation method.

1. **Comparison of optimally quantised wavelet to JPEG.**

The optimally quantised wavelet is compared to JPEG and EZW in figure 7-11. The new method is an improvement of JPEG but does not perform as well as the EZW. This is probably because the wavelet coefficients are assumed not to be related when optimally quantising the wavelet as so it is not as effective as would have been.

The wavelet transform packs all the important information in to the lowest frequency subbands of the image but the details are represented by the smaller fluctuations in the larger sub-bands (LH,HL,HH). The optimal quantisation method works well if there are enough bits to represent all the subbands effectively but as soon as a subband is turned off (or has a particularly high quantisation and is almost turned off) the method has no way of efficiently representing these large quantities of zero coefficients.

Wavelet methods are not just a matter of efficient transform, quantisation and lossless coding stages and a lot more work has to be put into the relationship between coefficients. Unfortunately the optimal quantisation method only works effectively when there is no connection between the coefficients in a block and so it would be difficult to pursue this method any further with wavelets since it would lose most of its advantages (predictable compression and optimal rate distortion prediction).



Figure 7-11. Rate distortion performance of the optimally quantised wavelet (4 scales) compared to JPEG and EZW.

1. **Summary**

This chapter looked at the wavelet transform and compared an implementation of the EZW to JPEG. It shows that the EZW is a superior compressor than JPEG and why it is necessary to exploit the structure of the wavelet transform to make wavelet compressors work effectively. The optimal quantisation of the wavelet transform has been shown to be less effective than the EZW and this is primarily because of the contradiction that exists between the wavelet transforms structure and the optimal quantisation method. Although this is a negative result it does highlight the fact that the wavelet transform itself is not the key to improved compression of the EZW. The improvements to the EZW come primarily from an enhanced source coding and not the wavelet transform. This is demonstrated further by Monro and Laurance [56] in which an embedded source coding method is used to good effect with the DCT.

The performance of the EZW and the optimal quantisation DCT are shown in figure 7-12. Even though the EZW method is very effective it still does not compete with the DCT when it is used correctly. This is an important result since it shows that DCT’s are a viable compression method and not only wavelet based methods have a future.



Figure 7-12. Comparison of rate distortion performance of Optimally quantised DCT, JPEG and EZW.