

SF2 Image Processing - Second Interim Report

Jonty Page

Pembroke College, University of Cambridge

May 28, 2019

1 Introduction

In this report, three different energy compaction methods are explored and their relative merits discussed. These methods are the Discrete Cosine Transform (DCT), the Lapped Bi-orthogonal Transform (LBT) and the Discrete Wavelet Transform (DWT). Each method is assessed by comparing the compression ratio of each scheme with optimal parameters alongside the visual quality of the compressed images. The compression ratio compares the total bits required for a compressed image using the desired scheme with the bits required for a reference scheme (namely, in this case, direct quantisation with a step size of 17). The greater the compression ratio for the same RMS error and visual quality, the better the compression scheme.

2 The Discrete Cosine Transform (DCT)

The DCT is an energy compaction method which operates on non-overlapping blocks of pixels by a reversible transform process. The 1D DCT is very closely related to the Discrete Fourier Transform and it extracts data about the frequency content of the block. We can form a 2D DCT by performing the 1D DCT on both the rows and the columns of the image in either order since the transform is linear and separable.

To demonstrate the operation of the DCT, a 8-point 1D Type-II DCT matrix was used to perform a 2D DCT on each 8×8 non-overlapping block of pixels in an image. In the image produced by the transform, each block is replaced by an equivalent block of transform coefficients (increasing frequency in content from the top left pixel). In order to create a more meaningful image, the result can then be regrouped in order to place pixels relating to the same frequency together in a small sub-image as shown in *Figure 1*. As the frequencies increase, the energy of the associated sub-image decreases (energy is 1.7686×10^6 for lowest frequency sub-image and 482.7 for the highest frequency sub image). If there is no quantisation, the original image is perfectly reconstructable from the transformed image by carrying out the inverse transform.

Figure 2 shows the sixty-four 8×8 basis functions used in the above DCT transformation. The frequency and detail in each of the basis functions increases as we move from the top left-hand corner to the bottom right-hand corner hence when performing the transform, this gives us the corresponding filter coefficients relating to each increasing frequency band from the top left as was seen in *Figure 1*.

In order to compress the image, we must first quantise the transformed image. If we quantise the DCT image fairly coarsely (with a step size of 17) and then regroup the image as above, we obtain the image shown in *Figure 3*. The distribution of pixels in the transformed image becomes less broad as we move from the top left to the bottom right, hence the entropy contribution from each sub-image decreases if we consider each sub-image separately. The total number of bits if we consider each sub-image over a different distribution (separate contributions) is 9.7467×10^4 which is considerably less than if we consider the entire image over the same distribution giving 1.0963×10^5 . In the limit where we consider the entropy contribution from each pixel separately, the total number of bits will be 0 as there is only one value in the distribution for each pixel hence there is no uncertainty. *Figure 4* shows the reconstructed image from the quantised DCT at a step size of 17. The RMS error between the original image and the reconstructed image is 3.7568 compared to the RMS error between the original image and an imaged directly quantised from the original with a step size of 17 of 4.8612 - hence for the same quantisation step size, the DCT has a smaller RMS error.

In order to make comparisons between each energy compacting scheme we must first make the RMS error between the compressed image and the original consistent across all schemes. In order to do this we compare the RMS error to that of the directly quantised image and adjust the step size accordingly to ensure they are equal. It was found that the optimal number of steps required for equal RMS error with an 8×8 DCT was 23.69. The total number of bits required for the compressed image was 7.755×10^4 which gave a compression ratio (as defined in the introduction) of 2.942. This process was repeated for kernels of sizes 4 and 16 and the results are detailed in *Table 1*. It was found that the optimal kernel size was 8 for this image as it resulted in the greatest compression for the same image quality. *Figure 4* shows the reconstructed images for the sizes 4, 8, 16 kernels however visually, you can see in the size 16 image, some of the sharpness has been lost whereas in the size 4 image, blocks are clearly visible in the smooth areas of the image as a result of the small block size DCT. The size 8 kernel provides a trade-off between these two features and hence is the best choice kernel for this image. In other images where there are many detailed edges a smaller kernel size may be required or for images with large smooth surfaces, a larger kernel size could be used.

3 The Lapped Bi-orthogonal Transform (LBT)

The LBT is an extension of the DCT which takes into account the correlation between neighbouring blocks in the image. The LBT can be represented as a Photo Overlap Transform (POT) which pre-filters the image, followed by a normal DCT. The POT can be performed easily through simple matrix multiplication with a coefficient matrix. The relationship between the forward POT matrix and the reverse POT matrix is defined by a scaling factor s where $1 \leq s \leq 2$. In order to find the scaling factor which maximises compression ratio for this image, an 8×8 LBT was performed using a range of different scaling factors between 1 and 2, the RMS error was matched to that of direct quantisation as before and the number of bits required for the compressed image was minimised - the results of this analysis are shown in *Table 2*. Although the optimal scaling factor was found to be 1.33 through the optimisation algorithm, the compression ratio was approximately constant for values of s such that $1.33 \leq s \leq 1.41$. The reconstructed LBT compressed images for a number of scaling factors are shown in *Figure 5* - we can see that as the scaling factor deviates from approximately $s \approx \sqrt{2}$ that the visual quality of the reconstructed image decreases with the image becoming more pixelated and a silhouette effect is more noticeable around sharp edges in the image.

As with the DCT, the basis functions for the POT are shown in *Figure 6* below for a variety of scaling factors alongside the pre-filtered image associated with each. The symmetry of the basis functions show us that edges are picked out by the pre-filtering and that this selection becomes more harsh as scaling factor increases.

Finally, the using a different block size for the LBT can have an impact on the quality of the compression hence this was tested by performing an LBT using a kernel size of 4×4 and 16×16 and comparing this to the optimal 8×8 kernel compression already analysed. *Figure 7* shows the reconstructed images for each of the kernel sizes with matched RMS errors as before. The results of analysis shown in *Table 3* show us that a kernel size of 4×4 achieves the greatest compression ratio for this image and is also produces the most crisp output image visually as there is very little silhouette effect around the edge of the lighthouse as with the other kernel sizes.

4 The Discrete Wavelet Transform (DWT)

The DWT is a third method of energy compaction which combines the best features of the DCT and the Laplacian Pyramid. It analyses the image at a range of different levels however it avoids any expansion in the number of coefficients. Each image is recursively split into three high-pass and one lowpass image. At each level, the low-pass image is split and the new lowpass image is placed in the top left-hand corner of the new image. *Figure 8* shows the result of performing a single layer 1D DWT, first in the horizontal direction and then again in the vertical direction to obtain a single layer 2D transform. After the first transform in horizontal direction, the energy of the low-pass image was found to be 8.2310×10^7 which is significantly greater than the energy of the high-pass image which was 3.5018×10^6 . By looking at the images which the 2D DWT has produced, each of the three high-pass images appear to pick out different features from the original image. One of them picks out the horizontal edges (bottom left), another the vertical edges (top right) and the final one picks out points upon which there is both a vertical and horizontal edge (bottom right).

If the transformed image is not quantised, perfect reconstruction is possible from the DWT transformed image however, in order to compress the image, quantisation is required. To quantise the DWT image we should quantise each sub-image according to a matrix of step-sizes with dimension $3 \times (N + 1)$ where N is the number of levels in the DWT. This matrix will be populated either with the same number in all elements (constant step-size quantisation) or a range of different number (equal mean squared error quantisation).

4.1 Constant Step-Size

Two test images will now be used - both the image of a lighthouse and also an image of a bridge, to assess how each scheme handles the different features within each image. In order to compare compression schemes, we must first optimise the step-size matrix such that the RMS error between the compressed image and the original is the same as the RMS error between the directly quantised image with a step size of 17 and the original. *Table 4* shows the optimal constant step-size which populates all elements of the matrix to give the same RMS error for each image. The tables also outline the various compression ratios for the different numbers of levels in the DWT in order to find an optimal DWT-depth. It was found that for both images, 3 levels was optimal with a compression ratio of 2.695 for the lighthouse image and 1.761 for the bridge image. The reconstructed images for each DWT in the table are shown in *Figure 9* in order to assess visual quality. As the number of layers in the DWT increases, the image quality also increases since the edges become sharper (especially seen in the bridge image) and the lighting is better. Hence a trade-off ought to be made between compression ratio and visual quality which suggests a DWT depth of 4 may be more suitable.

4.2 Equal-MSE

In order for the quantisers in each layer of the DWT to contribute to the mean squared error (MSE) equally, the impulse response for each layer had to be found. This was achieved by adding a single impulse to each sub-image of the DWT in turn, reconstructing and recording the energy of the reconstructed image for each impulse. In order for equal-MSE to be satisfied, the ratio of quantisation step-sizes used for each sub-image should be inversely proportional to the square-root of the energy of the impulse responses for each sub-image. The ratio of step-sizes for each sub-image in an n -level DWT was found and used to create a matrix of step-sizes for each n -level DWT. As above, in order to compare the schemes the RMS errors of direct quantisation and compression must match hence the step-matrix was optimised for each n -level DWT whilst keeping in the correct ratios defined above. *Table 5* outlines the optimal step-size for the high-high-pass image (bottom right) in the first layer of the DWT (from this all other step-sizes can be found using the ratios) for each n -level scheme. The total number of bits and compression ratio is also shown with the optimal being $n = 5$ for both images (the lighthouse image having a compression ratio of 3.101 and the bridge 1.919). Finally, again in order to assess visual quality, *Figure 10* shows the reconstructed images for each DWT in the table. The difference in visual quality for the Equal-MSE scheme does not change much as the depth of the DWT decreases from 5 to 3, hence we should choose for the optimal compression ratio, hence a DWT depth of 5 is likely to be most suitable.

5 Conclusions

All three of the energy compaction schemes discussed above achieve a good compression ratio for a small trade-off in visual quality reduction, however, since the DWT combines the advantages of both the DCT and the Laplacian Pyramid, it consistently matched or outperformed the other compression schemes in both compression ratio and compressed image quality for a number of different images. Also, the quantising the DWT according to the Equal-MSE criterion provided an even greater compression ratio for very little reduction in image quality, hence improving the performance of the DWT even further. For this reason, the DWT is likely to be a better choice of compression scheme than the DCT or the LBT for a random image.

Appendix A The Discrete Cosine Transform - Figures

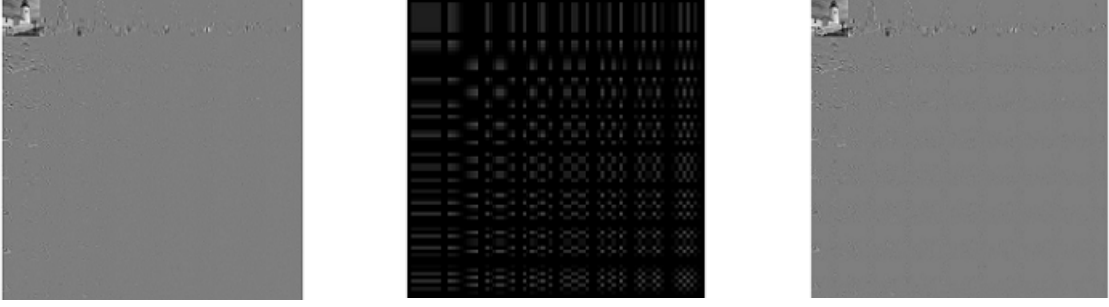


Figure 1: Regrouped DCT Image Figure 2: 8×8 basis functions for Figure 3: Regrouped DCT Image using 8-point 1-D Type-II Matrix DCT quantised with a step size of 17



Figure 4: DCT Reconstructed Images for different kernel sizes

Kernel Size	Optimal DCT Step Size	DCT Total Bits	Compression Ratio
4	23.90	8.6621×10^4	2.646
8	23.69	7.755×10^4	2.942
16	22.34	7.9153×10^4	2.882

Table 1: Comparison of DCT Compression Ratios for different kernel sizes

Appendix B The Lapped Bi-orthogonal Transform - Figures

Scaling Factor, s	Optimal DCT Step Size	LBT Total Bits	Compression Ratio
1	23.37	7.5548×10^4	3.020
1.25	25.16	7.3043×10^4	3.123
1.5	26.23	7.3069×10^4	3.122
1.75	26.53	7.5785×10^4	3.010
2	26.44	7.9834×10^4	2.857
1.33 (Optimal)	25.57	7.2757×10^4	3.135

Table 2: Comparison of LBT Compression Ratios for different scaling factors, s



Figure 5: LBT Reconstructed images for different scaling factors

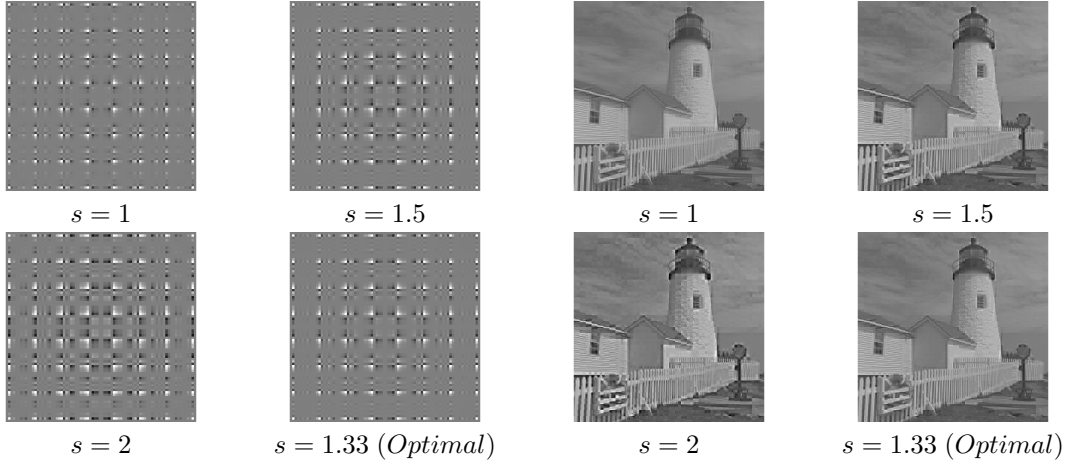


Figure 6: POT basis functions and associated pre-filtered images for different scaling factors



Figure 7: LBT Reconstructed Images for different kernel sizes

Appendix C The Discrete Wavelet Transform - Figures

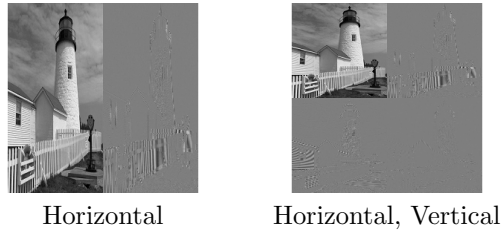


Figure 8: Single Layer DWT in 2D

	<i>Lighthouse Image</i>			<i>Bridge Image</i>		
No. Levels	Optimal Step Size	Total DWT Bits	Compression Ratio	Optimal Step Size	Total DWT Bits	Compression Ratio
3	7.63	8.4895×10^4	2.695	6.13	1.3462×10^5	1.761
4	6.56	9.1910×10^4	2.489	5.51	1.4215×10^5	1.668
5	5.61	1.0127×10^5	2.259	5.04	1.4969×10^5	1.584

Table 3: Comparison of DWT Compression Ratios for Constant Step-Size Quantisation with different Depths



Figure 9: DWT Reconstructed images for Constant Step-Size

	<i>Lighthouse Image</i>			<i>Bridge Image</i>		
No. Levels	Optimal Step Size	Total DWT Bits	Compression Ratio	Optimal Step Size	Total DWT Bits	Compression Ratio
3	8.70	7.4925×10^4	3.053	6.35	1.2458×10^5	1.903
4	8.70	7.3985×10^4	3.092	6.35	1.2386×10^5	1.914
5	8.68	7.3783×10^4	3.101	6.35	1.2354×10^5	1.919

Table 4: Comparison of DWT Compression Ratios for Equal-MSE Quantisation with different Depths



Figure 10: DWT Reconstructed images for Equal-MSE