SF2 Image Processing - Second Interim Report

Kingson Chan - ksc37

1 Introduction

In the previous interim report, we investigated the use of the Laplacian pyramid as a method of energy compaction in image processing, as well as comparing two different quantisation and coding schemes: the constant step size scheme and the equal MSE scheme. In this second interim report, we will be looking at more advanced techniques for energy compaction, namely the Discrete Cosine Transform (DCT), the Lapped Bi-orthogonal Transform (LBT), and the Discrete Wavelet Transform (DWT). We will again use the direct quantised image with step size 17 as the reference scheme throughout this report to compare their compression performance and visual features.

2 The Discrete Cosine Transform

The Discrete Cosine Transform (DCT) is a reversible linear transform process that operates on non-overlapping blocks of pixels. Each block of pixels is replaced by a block of transform coefficients of the same size. The 1-D N-point Type-II DCT is defined as follows:

$$y(k) = \sum_{n=0}^{N-1} C_{kn}\ x(n) \qquad \text{for } 0 \leq k \leq N-1 \tag{1}$$

$$\text{where } C_{0n} = \sqrt{\frac{1}{N}} \tag{2}$$

$$\text{and } C_{kn} = \sqrt{\frac{2}{N}} \cos\frac{k(n+\frac{1}{2})\pi}{N} \qquad \text{for } 1 \leq k \leq N-1 \tag{3}$$

The key advantage of the DCT compared to the Laplacian pyramid is that there is no expansion of the number of samples. The transformed image is the same size as the original image, which is more desirable for data compression.

In an 8-point Type-II DCT, the DCT analyses each 8x8 block of image pixels into a linear combination of sixty-four 8x8 basis functions, shown in fig 1. The basis at the top left corresponds to the DC component, with increasing horizontal and vertical frequency towards the right and the bottom. This gives us the corresponding DCT coefficients relating to each increasing frequency.

Fig 2 shows the 8-point DCT applied to the lighthouse image. The image is regrouped such that all the coefficients of a given frequency are placed in a small sub-image. It can be seen that as the frequency increases, the energies of the sub-images decrease. In fact, the energy of the DC component sub-image is 1.77×10^6 and decreases rapidly to only 483 in the highest frequency sub-image. The pixel distribution is very varied in the DC component sub-image but becomes less broad as frequency increases. Hence the entropy in higher frequency sub-images is low.

Similarly to the Laplacian pyramid, we will exploit the fact that higher frequency sub-images have less energy and lower entropy to compress the overall image. Using the lighthouse image with the 8-point DCT and quantised with a constant step size of 17, we find that only 9.75×10^4 bits are required if we consider the entropy of each sub-image separately, compared to 1.10×10^5 bits if we only consider the entropy of the overall transformed imaged. Interestingly, reconstructing from a quantised DCT transformed image resulted in a lower RMS error of 3.76 compared to 4.93 with direct quantisation of the image.

Using direct quantisation of step size 17 as the reference scheme, the quantiser step sizes for similar RMS error with 4-point, 8-point and 16-point DCT is found. The reconstructed images are shown in fig 3, along with a zoomed in region of the building roof and lighthouse edge. Since the DCT is applied block by block with no correlation between adjacent blocks, it creates block artifacts which is especially obvious in the 4-point image. There is heavy distortion around the edges in the reconstructed 16-point image, due to the loss of high frequency components in the DCT. The compression ratio is found by considering the entropy of each frequency sub-image. However, the analysis is slightly biased with larger transform sizes, as there are a greater number of smaller sub-images on which to calculate probability distributions. Consider the limiting case where the transform size is the size of the image. Each sub-image will only contain one pixel, meaning that the distribution is a peak at that pixel intensity and zero everywhere else. Hence the entropy is zero and wrongly suggests that zero bits are required to code the image. It might be better to use the same transform size to calculate sub-images entropy even when the actual transform size varies.

The 8-point DCT provides a suitable trade-off between sharpness and block artifacts, along with the best compression ratio of 2.99 amongst all three transform sizes, makes the 8-point DCT the best transform size for the Lighthouse image. Due to the fact that larger transform sizes tend to distort higher frequency content, images where there are many detailed edges might benefit from smaller transform sizes, whereas smoother images might benefit from larger transform sizes.

3 The Lapped Bi-orthogonal Transform

The type-II Lapped Bi-orthogonal Transform (LBT) attempts to reduce the block artifacts issue in the DCT by pre-filtering the image to create overlapping blocks before applying the DCT. This pre-filtering step is known as the Photo Overlap Transform (POT). This allows the DCT to take advantage of correlation between blocks to produce better end results.

A scaling factor s in the POT determines the degree of bi-orthogonality. A scaling factor of $s = 1$ represents equal contribution from both the forward and backward filter, while $1 < s < 2$ means that the forward filter is weighted more significantly compared to the backward filter. The effect of the scaling factor is shown in fig 4. There is more overlap between blocks for larger s, creating wiggly block artifacts.

Using an 8-point DCT, the quantiser step size which gives similar RMS error to the reference scheme is found for varying POT scaling factors. As shown in fig 5, the compression ratio peaks at around $s = 1.4$, approximately $\sqrt{2}$, corresponding to a compression ratio of 3.20. The effect of different scaling factor can be seen in fig 6. The visual differences between the images are subtle, with larger scale factors being slightly more pixelated in the sky.

Using the optimal POT scaling factor of $\sqrt{2}$, the performance of LBTs with different block sizes are compared. As shown in fig 7, the 4-point LBT achieved the best compression ratio of 3.60. From the zoomed in area, we can see that most of the block artifacts are gone compared to the pure DCT reconstructed images in fig 3. Similar to the pure DCT images, there are increased edge artifacts with larger transform sizes. We find that smaller block sizes produce the best quality for LBT, with better visual quality and compression ratio compared to larger block sizes.

4 The Discrete Wavelet Transform

The Discrete Wavelet Transform (DWT) attempts to combine the best features of the Laplacian pyramid and the DCT by analysing the image at a range of different scales with symmetrical filters while avoiding any expansion in the number of coefficients. By using band-pass filters in different parts of the frequency spectrum, we can split an image up as a binary wavelet tree. The DWT transforms the image into a set of 4 sub-images: one low-pass image and three high-pass images, corresponding to the vertical, horizontal and diagonal edges. Each sub-image is a quarter-size of the original image, so the total number of coefficients is preserved.

We will use the LeGall 5 and 3 tap filter pair, given by:

$$u_n = \frac{1}{8}(-x_{n+2} + 2x_{n+1} + 6x_n + 2x_{n-1} - x_{n-2}) \quad \text{and} \quad v_{n+1} = \frac{1}{4}(-x_{n+2} + 2x_{n+1} - x_n) \tag{4}$$

We can decimate by 2 by choosing even n only for u and v. The lowband outputs u_n are centred on the even samples, and the highband outputs v_{n+1} are centred on the odd samples. This property allows perfect reconstruction from u and v.

Fig 8 shows the LeGall filter pair applied to the lighthouse image. The low-pass image generated by u is shown on the left and the high-pass image generated by v is shown on the right. Both are decimated by 2 so is only half the width of the original image. As expected, the low-pass image has much higher energy content of 8.23×10^7 compared to only 3.50×10^6 in the high-pass image. Fig 9 shows the full set of DWT transformed sub-images. The top left sub-image is low-pass filtered in both the vertical and horizontal direction, while the other three sub-images are high-pass filtered in the vertical, horizontal and diagonal directions, showing their respective edge details. From the expressions for the LeGall highband filter v in equation (4), we see that the difference in neighbouring pixels is multiplied by a factor of $\frac{1}{4}$, hence the pixel intensity values are much smaller in the high-pass images and might need to be multiplied by a factor $k > 1$ to display clearly.

We will investigate the performance of both the equal step size and equal MSE quantiser scheme. Following the same principles as the Laplacian pyramid, impulse response measurements are taken and the step size ratios for each level and sub-image of DWT is chosen for the equal MSE scheme. We have included a second bridge image, shown in fig 10, which includes finer details and more features for comparison. Again using the direct quantiser of step size 17 as reference scheme, the compression ratio for different levels of DWT was found in fig 11. It is found that the compression ratio is much higher for the lighthouse image but the two images follow the same trends. Two levels DWT is optimal for the equal step size scheme with a compression ratio of 2.81 in the lighthouse image, whereas the compression ratio for the equal MSE scheme reaches an asymptote of around 3.16 in the lighthouse image as the number of levels increase. The asymptote in the equal MSE scheme compression ratio could be due to inaccuracies in the impulse measurement. As the size of the sub-images decreases, the impulse becomes closer to the edge of the sub-image and edge effects will dominate. Overall, the equal MSE scheme and constant step size scheme has similar performance for one and two levels, but the equal MSE scheme has much better compression performance from three DWT levels onwards.

The reconstructed images using various levels of DWT are shown in fig 12 for the bridge image and fig 13 for the lighthouse image. For the bridge image, there is practically no visual difference between the different quantiser schemes and with different DWT levels. This is mostly because the bridge image has much more complex edges and geometry, which the highband filter in the DWT is unable to pick out well. This also explains the lower compression ratio overall using the DWT for the bridge image compared to the lighthouse image. The lack of visual quality difference means that for the bridge image, the optimal configuration is to use the one with the best compression ratio, which is the equal MSE scheme with five DWT levels. The difference between quantiser schemes and DWT levels is more pronounced in the lighthouse images. In the constant step size scheme, the sky is more patchy in the two DWT levels image, whereas the sky is less blocky but has some streaky artifacts in the five DWT levels image. The visual difference is even larger with the equal MSE scheme. Using only one DWT level, the sky is very pixelated, whereas with three DWT levels, the sky is much smoother and there are less edge artifacts. The better visual quality and compresssion ratio makes the equal MSE scheme with three DWT levels the best option for the lighthouse image.

5 Conclusion

We have investigated the use of 4 different energy compaction methods so far: the Laplacian pyramid, DCT, LBT, and DWT. The LBT and DWT builds upon the concepts from the Laplacian pyramid and the DCT, so it is expected that the LBT and DWT will outperform it's predecessors in visual quality and compression performance. Indeed we find that is true from table 1. The LBT offers a better compression ratio of 3.60 compared to 3.16 with the DWT. However, we find that the reconstructed LBT image has slightly worse visual quality, with some block artifacts and edge distortions. The reconstructed DWT image is smoother and has less artifacts, but is missing some finer details in the image. Similar compression ratio trends are found with the bridge image using the LBT, but there is very little visual differences using the LBT compared to the DWT.

In conclusion, the LBT and DWT are both excellent schemes for energy compaction. The key differences in visual quality might affect the optimal choice of compaction scheme depending on the image. Images with high amount of complex details will benefit from the higher compression ratio of the LBT, whereas images with smooth gradients will benefit from the reduced artifacts of the DWT. For a random unseen image, the DWT with equal MSE quantiser scheme is likely to be a better choice of compression with it's reasonably high compression ratio for very little reduction in image quality.

Energy compaction scheme	**Optimal compression ratio**
Laplacian pyramid	1.568
DCT	2.993
LBT	3.597
DWT	3.159

Table 1: Comparison between different energy compaction schemes for the lighthouse image

Appendix

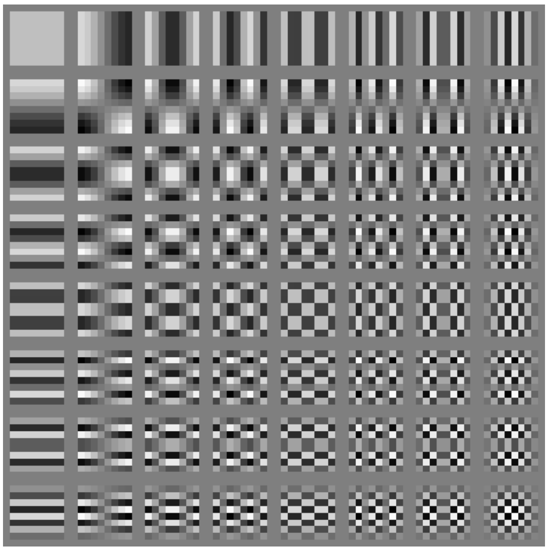


Figure 1: Sixty-four 8x8 DCT basis functions

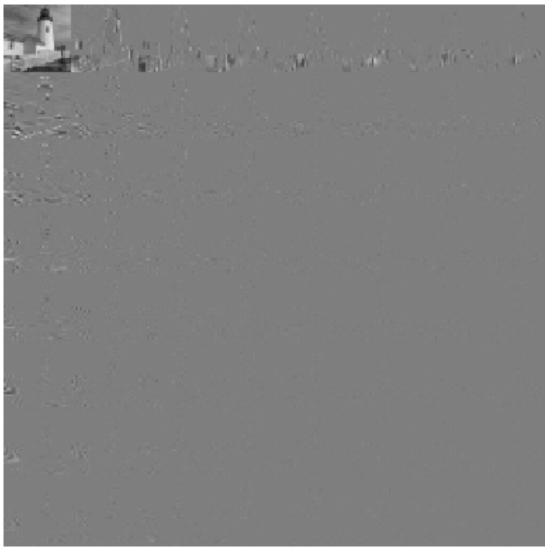


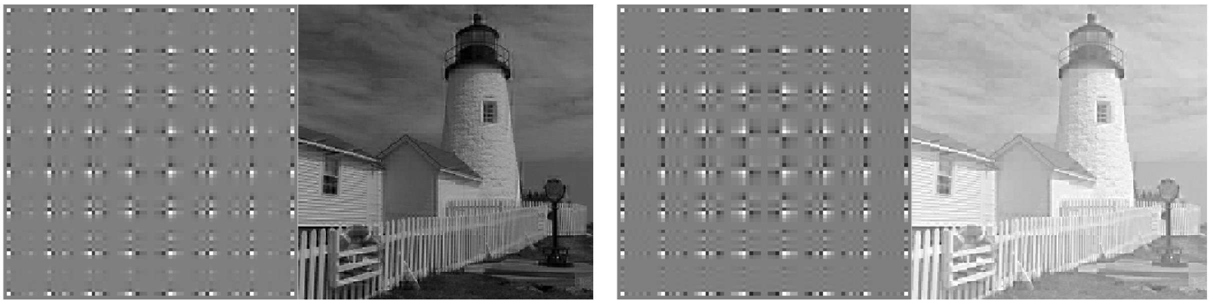
Figure 2: Regrouped 8x8 DCT image

(a) 4x4 DCT
Compression ratio = 2.69

(b) 8x8 DCT
Compression ratio = 2.99

(c) 16x16 DCT
Compression ratio = 2.92

Figure 3: Top - Reconstruction images using various DCT sizes. Bottom - Close examination of roof and lighthouse edge area



(a) s = 1 (b) s = 1.5

Figure 4: POT bases (left) and pre-filtered image (right) for different POT scaling factors, s

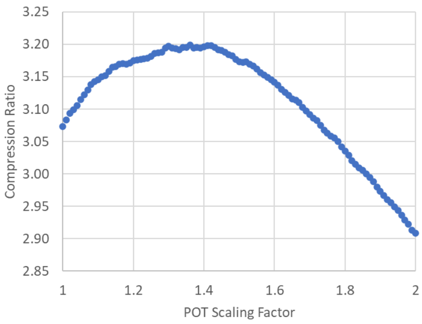


Figure 5: Compression ratio for various POT scaling factors

Figure 6: Reconstructed LBT with POT scaling factor $s = 1$ (left), $s = 1.5$ (middle), $s = 2$ (right)

(a) 4x4 LBT
Compression ratio = 3.60

(b) 8x8 LBT
Compression ratio = 3.42

(c) 16x16 LBT
Compression ratio = 2.96

Figure 7: Top - Reconstruction images using various LBT block sizes. Bottom - Close examination of sky and lighthouse edge area

Figure 8: Row filtered with LeGall filters

Figure 9: 2-D filtered with LeGall filters

Figure 10: Original bridge image

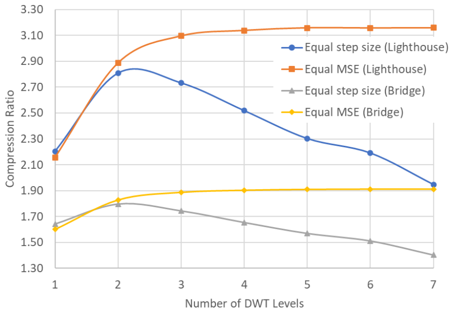


Figure 11: Compression ratio for varying DWT levels and quantisation scheme in the lighthouse and bridge image

N = 2
Constant step size
Compression ratio=1.80

N = 5
Constant step size
Compression ratio=1.57

N = 3
Equal MSE
Compression ratio=1.89

N = 5
Equal MSE
Compression ratio=1.91

Figure 12: Reconstructed bridge image using various DWT levels and quantisation scheme

N = 2
Constant step size
Compression ratio=2.81

N = 5
Constant step size
Compression ratio=2.30

N = 1
Equal MSE
Compression ratio=2.16

N = 3
Equal MSE
Compression ratio=3.10

Figure 13: Zoomed in reconstructed lighthouse image using various DWT levels and quantisation scheme