SF2: Second Interim Report

Nicholas Capel (nrjc2)

May 18, 2016

1 Introduction

In this report, I will investigate the results from the DCT, LBT and DWT energy compaction methods and compare them to the methods discussed in the first interim report.

2 Discrete Cosine Transform (DCT)

The Discrete Cosine Transform is a method of decomposing a signal into its elementary frequency components. In image processing, we deal with 2D signal blocks, and hence, the 2D DCT is constructed by $Y = (C * (C * X)')'$, where Y describes the transformed coefficients.

Perhaps this can be better illustrated by an image comprising the 2D basis functions (Figure 1a). The patterns describe the horizontal and vertical spatial filters that will be used to construct the filtered image, and the DCT essentially deconstructs an image into its constituent frequency basis components that are described by these filters. Observing sub-images obtained from the DCT (Figure 1b), it is observed that, as the image frequencies increase, the energies decrease. This is expected, as natural images have much higher energies at low frequencies than high frequencies. Furthermore, moving right and downwards in the sub-images corresponds to higher spatial frequencies in the x and y-components of the image respectively, which agree well with the earlier visualisation of the 2D basis functions.

In order to compress the image, we now quantise the sub-images. In this experiment, we do so at step size 17. When reconstructing the original image via direct quantisation, we get a RMS error of **4.8612**. When reconstructing from the DCT, we get a RMS error of **3.7568**. The RMS error from the DCT is lower than that of direct quantisation, which is unusual, but not unexpected. The DCT is analogous to a Fourier transformation, with high frequency components discarded. However, these high frequency components are not exceptionally important, and contribute little to the overall image. On the other hand, in the direct quantization, we force a rounding of all image values to the nearest step, and this can lead to greater loss of information.

Quantising the sub-images separately and together makes a huge difference to the entropy, and hence, the compression ratio that we can achieve.

Table 1: Entropies of sub-images quantised differently

quantised separately (dctbpp)	quantised together (dct)
9.7468×10^4	1.0963×10^5

The entropy of each image quantised separately is lower than that of the whole image quantised together. This makes sense because entropy is a measure of the number of possible states. When quantising the entire image at once, the number of possible states will definitely be higher than if each sub-image were to be quantised separately.

In the limit where you use quantise each pixel separately**,** the total entropy goes to zero. This is because the entropy of each individual pixel now becomes $\log(1) = 0$. This is not a realistic result. It does not make sense to encode each pixel separately, as we cannot exploit the correlations between pixels.

As compared to the original image, the DCT-compressed image has ***ringing artifacts*** that are introduced at the edges of sharp discontinuities. In the lighthouse image, this is most evident around the edges of the lighthouse and the roof (Figure 1c, 1d). These artifacts are a consequence of the DCT decomposing the image using cosine bases. On quantisation, the high frequencies components lose resolution, leading to ringing artifacts. Furthermore, upon closer inspection, the reconstructed image is split into multiple, identical squares, with discontinuities along the edges of those squares. These are known as ***blocking artifacts***. The sizes of those squares are exactly the same as the size of the DCT matrix used to filter the image, as the DCT is conducted block by block. In contrast, the directly quantised image has blocky artifacts introduced where originally smooth transitions are now turned into sudden jumps (Figure 1e).

An experiment was conducted, summarized in Figure 1f, where the size of the DCT base is modified while keeping the RMS error constant. The total entropy was measured from the function **dctbpp**, while keeping N constant at 8. This is because in real coding, the number of sub-images that can be quantised is fixed. It appears that as the size of the DCT basis matrix increases, the ringing artifacts become more noticeable (Figure 1c, 1d).

I conclude that a DCT of length **8** gives the overall best image, with a compression ratio of **3.1345**. It offers a good balance between compression ratio and the image quality. However, as most of the artifacts are created in the presence of sharp discontinuities, the number of these sharp discontinuities and their locations in the image might affect this choice of DCT length.

3 Lapped Bi-orthogonal Transform (LBT)

The Lapped Bi-orthogonal Transform (LBT) aims to remove the block discontinuities by generating smaller non-overlapping blocks. In this experiment, a LBT is implemented by using a POT (Photo Overlap Transform) combined with a DCT. In essence, an image is compressed by first prefiltering it with the POT to introduce overlaps between the blocks, then a DCT transform is applied on this result. As a consequence of this prefiltering, the major benefit is a suppression of the blocking artifacts in the quantised image.

The scaling factor, **s**, weighs the relative contributions between the POT filter, **Pf**, and the inverse POT filter, **Pr**. As the value of **s** increases, we can see that the amount of overlap between the bases increases (Figures 2a, 2b). This effect is also evident in the pre-DCT filtered images (Figures 2c, 2d).

An experiment was conducted to implement LBT with POT scaling factors between 1 to 2, while matching rms errors to the directly quantised image at step size 17. A graph of compression ratios was plotted against **s** values (Figure 2e). The optimal **s** value appears to be exactly $\sqrt{2}$, which gives a compression ratio (measured with N=16) of **3.898**. As the scaling factor increases, it seems that the reconstructed image becomes ***grainier***. At a high scaling factor, there is greater overlap between adjacent blocks. Hence, quantisation errors are more evenly spread throughout the image, leading to a more uniform grainy texture as compared to a lower factor, where the artifacts are more localised (Figures 2f, 2g).

Using a scaling factor of $\sqrt{2}$, an experiment was conducted to investigate the performance of the LBT with different block sizes (Figure 2h), once again keeping the MSE equal to a direct quantisation at step size 17. It was discovered that a length **8** LBT block is ideal, yielding a compression ratio of **3.898**.

4 Discrete Wavelet Transform (DWT)

The Discrete Wavelet Transform is essentially an attempt to combine the best features of DCT and the Laplacian Pyramid. By applying a filter, we decompose an image into the coefficients of its associated wavelets, and by repeating this sub-sampling and filtering process, we create a binary filter tree. In this report, the Legall 5 and 3 tap pair are chosen as filters. This process is known as the DWT.

Applying the wavelet filter once in the x-direction, we obtain, **U** and **V** (Figure 3a), it is observed that **U** (8.231×10^7) has a much higher energy than **V** (3.5018×10^6). Applying this filter again in the y-direction, we have completed one layer of the binary tree. An image of the result can found in Figure 3b**.** **UU** displays a low pass filter, **VV** displays a high pass filter. **UV** implements horizontal edge detection, while **VU** implements vertical edge detection. The reason why **UV** and **VU** implement edge detections is simple. **UV** selects for low-frequency components in the x-direction, but high frequency components in the y-direction. Hence leading to horizontal edges being highlighted in the filtered image. **VU** detects vertical edges for similar reasons.

In general, the high pass images have a much lower energy than the low pass images, and hence require a multiplication factor to display clearly. Once again, this is an intrinsic property of natural images.

An experiment was conducted to investigate the compression ratios of equal MSE and equal step size schemes, and how they vary with the number of levels of DWT (Figure 3c). It seems that as the number of levels increases, the equal MSE compression ratios will increase, subject to diminishing returns. At each level, the sub-image in the top left is split up into **UU**, **UV**, **VU** and **VV**. **UV**, **VU** and **VV** have significantly lower energies and entropies than the low-pass image **UU**. Hence, we can encode **UV**, **VU** and **VV** using fewer bits of information without significant loss. By recursively repeating this process, we can further increase compression. However, information is not infinitely compressible, so there exists an upper bound that is approached asymptotically as the number of levels increase.

Inspecting the equal MSE images from the DWT of both the lighthouse and bridge (Figure 3d, 3e), it is seen that as the levels in the pyramid vary, the visual quality of the image does not change.

However, this trend does not hold true in the constant step size case, where a maximum compression ratio is reached at a level size of **2**. This is because when a constant step size is used, we are using the same number of bits to encode the high-pass and the low-pass images. However, the low-pass image has a much higher energy and entropy than the high-pass images. Thus when we quantise all sub-images using the same step size, significant detail is lost in the low-pass images when the binary tree is deep. Hence beyond a certain binary tree depth, the compression ratios begin falling again.

Inspecting the constant step size images from the DWT of the bridge (Figure 3f, 3g), it is seen that as the levels vary, the cloudy artifacts in the images seem to become more diffuse. The lighthouse images exhibit a similar trend, and thus for the sake of brevity, will not be included in the report.

In this section, experiments were done on two separate images, an image of a lighthouse and an image of a bridge. Universally, the image of the bridge is much harder to compress than the image of the lighthouse. This is because there are many more features / edges in the bridge image than in the lighthouse image, and this is most pronounced in the emptiness of the sky versus the foliage on the sides of the bridge. In other words, the image of the bridge has more energy in its high spatial frequencies than does the image of the lighthouse.

5 Comparison

Table 2: Comparison between compression schemes

Scheme	Optimal Compression Ratio	Artifacts
Pyramid	1.549	Loss of detail due to smoothing by **h** filter.
DCT	3.1345	Ringing introduced about edges. Blocky artifacts observed because DCT is applied separately to each image block.
LBT	3.898	Some grainy artifacts introduced by POT. Ringing artifacts around sharp transitions still exist, as in DCT.
DWT	3.092	Grainy artifacts introduced. Ringing artifacts around edges exist, but much less pronounced than DCT / LBT.

The results of the experiments are summarized in Table 2. The optimal compression ratios shown are the best compression ratios for any given input image and any given configuration, and thus represent the best-case scenario of that compression algorithm. It is seen that LBT exhibits the best compression ratios, and will thus be investigated in greater detail in the next report. DWT will be investigated as well, as although the compression ratios are not as high as two other schemes - from a qualitative point of view, the artifacts introduced using DWT are much less obvious and intrusive than the artifacts introduced by any other compression algorithm investigated in this report.

6 Conclusion

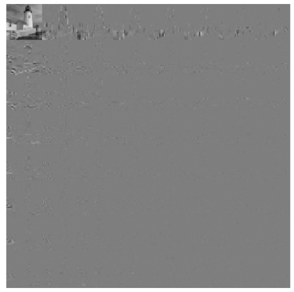
In this report, I have investigated the DCT, LBT and DWT, and compared their performance to the Laplacian pyramid. I have also identified further areas of investigation.

A Appendix

Figure 1: Discrete Cosine Transform Graphs and Images

(a) Filtered Bases for 8×8

(b) DCT sub-images for lighthouse



(c) DCT-compressed image (N=8)

(d) DCT-compressed image (N=16)

(e) Direct quantisation (step 17)

(f) Compression Ratio vs DCT base size

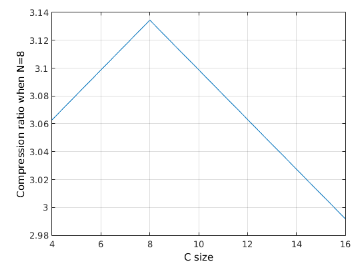
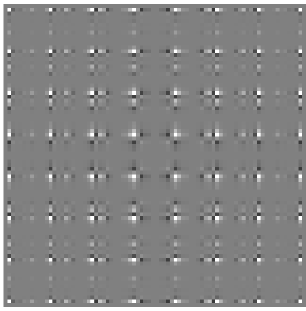
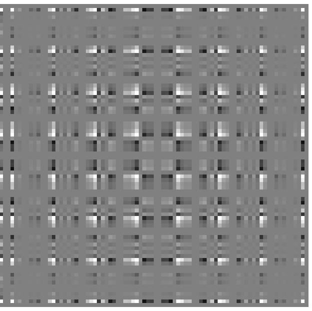


Figure 2: LBT Graphs and Images

(a) Filtered Bases with s=1.0



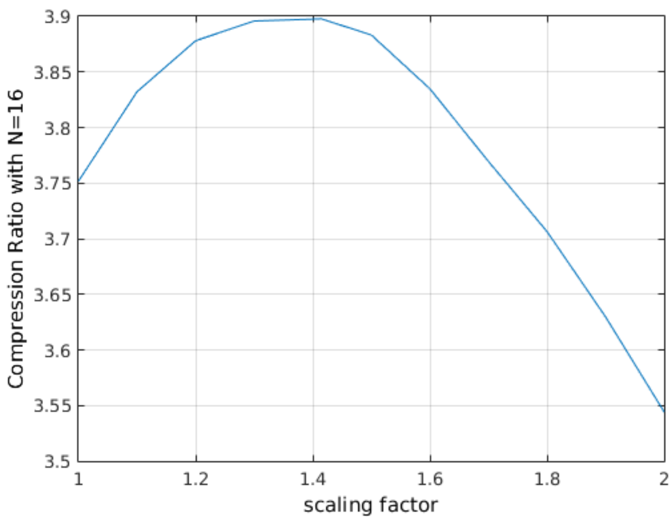
(b) Filtered bases with s=2.0



(c) Prefiltered lighthouse with s=1.0

(d) Prefiltered lighthouse with s=2.0

(e) Compression ratio vs size of **s**



(f) Prefiltered lighthouse with s=1.0

(g) Prefiltered lighthouse with s=2.0

(h) Compression Ratios vs Block Size

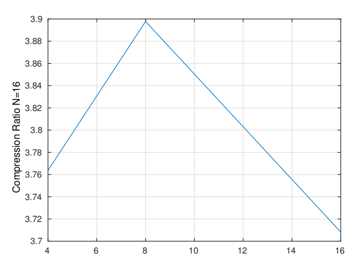
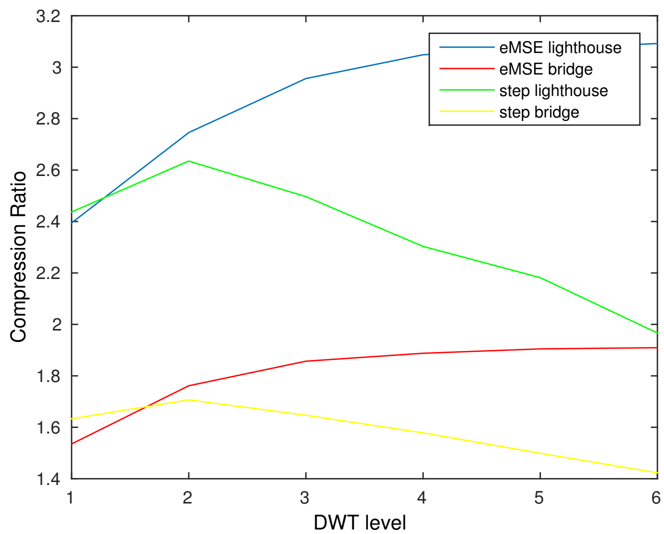


Figure 3: DWT Graphs and Images

(a) [U V]

(b) [UU VU; UV VV]

(c) Compression Ratio vs DWT level



(d) Lighthouse eMSE N=6

(e) Bridge eMSE N=6

(f) Bridge constant step N=1

(g) Bridge constant step N=6