SF2 Image Processing - Final Report

Kingson Chan - ksc37

1 Introduction

A complete image compression system requires three key components to work in tandem: the input filtering process, the quantisation process, and the coding process. In the previous reports, we have explored various input filtering and quantisation processes in depth, comparing and contrasting their compression performance and visual qualities. In this final report, we will investigate the use of centre-clipped quantisers to improve compression performance and Huffman coding to efficiently encode the image into codewords. We will also investigate more advanced techniques, such as the use of frequency-dependent quantisation and the structural similarity index as an improved error measurement metric. A third flamingo image, shown in Appendix fig. 6, is introduced to compare the performance of various schemes alongside with the lighthouse and bridge images. Working in a team of two, a final end-to-end image compression system design is chosen and used to compress an unseen image to less than 5 kB.

2 Candidates for energy compaction schemes

We have previously investigated 4 different energy compaction schemes. Their respective optimal compression ratios using the lighthouse image is summarised in Table 1.

Energy compaction scheme	Compression ratio	Notes
Laplacian pyramid	1.568	2 layers and Equal MSE scheme
DCT	2.993	8x8 block size
LBT	3.597	4x4 block size, POT scaling factor of $\sqrt{2}$
DWT	3.159	5 frequency levels and equal MSE scheme

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

The subsequent processes that we will investigate are non-linear, so it is difficult to choose the perfect front-end at this stage. Hence, we will select 3 good front-end candidates and trust that one of them will lead to a near-optimum solution.

The Laplacian pyramid offers great visual quality, however it is not sufficient to compensate for it's weak compression performance. Therefore, in the rest of our investigate, we will focus on the 8x8 DCT, 4x4 LBT, and 5 level DWT.

3 Centre-clipped linear quantisers

In our previous investigations, we have only used uniform quantisers, where all steps have the same step size. However, we noted that the energy of high-pass images were much lower and a significant proportion of pixel intensities are near zero. We will exploit the fact that there is a peak in the image intensity probability distribution around zero to improve the compression ratio.

The compression ratio depends heavily on the proportion of intensities that are quantised to zero. As more pixels are quantised to zero, higher compression is achieved. By using a non-linear quantiser, we can adjust the width of the 'zero' step. The first rise of a quantiser determines the transition point from 'zero' to the 'first' level. In a uniform quantiser, the 'zero' step is normally centred on zero with the first rise equal to half of the step size. This allowed the 'zero' step to have the same width as other levels. By adjusting the first rise and hence widening the zero step, more samples will be coded as zero so the entropy of the data will decrease.

In fig. 1, we can see how the RMS error varies as a function of the number of bits for varying first rise to step size ratios with the lighthouse image. We find that widening the first rise size produced much better quantisation error for the same number of bits required compared to the uniform quantiser with $rise = 0.5 \times step$. The optimal value is found to be $rise = 0.8 \times step$, where only 7.3×10^4 bits are needed compared to 7.8×10^4 bits with the uniform quantiser to achieve the same RMS error. Similar results are found using the LBT and DWT, which is shown in the Appendix fig. 7 and fig. 8, and with the bridge and flamingo images. There is only marginal improvement using $rise = 0.8 \times step$ compared to using $rise = step$. Therefore, using $rise = step$ is a reasonable compromise for simplicity.

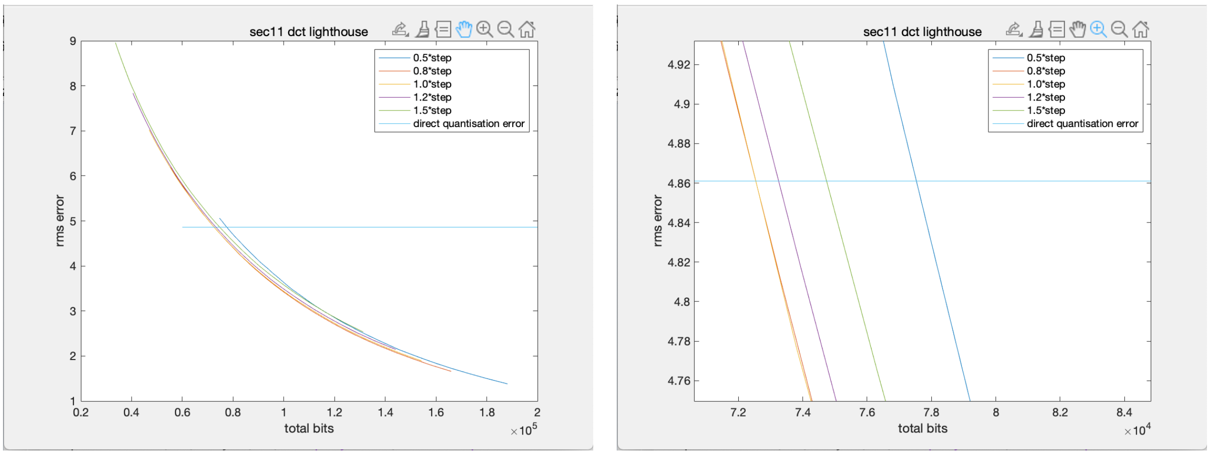


Figure 1: Left - Varying first rise to step size ratio for the DCT. Right - Zoomed in region around the reference RMS error.

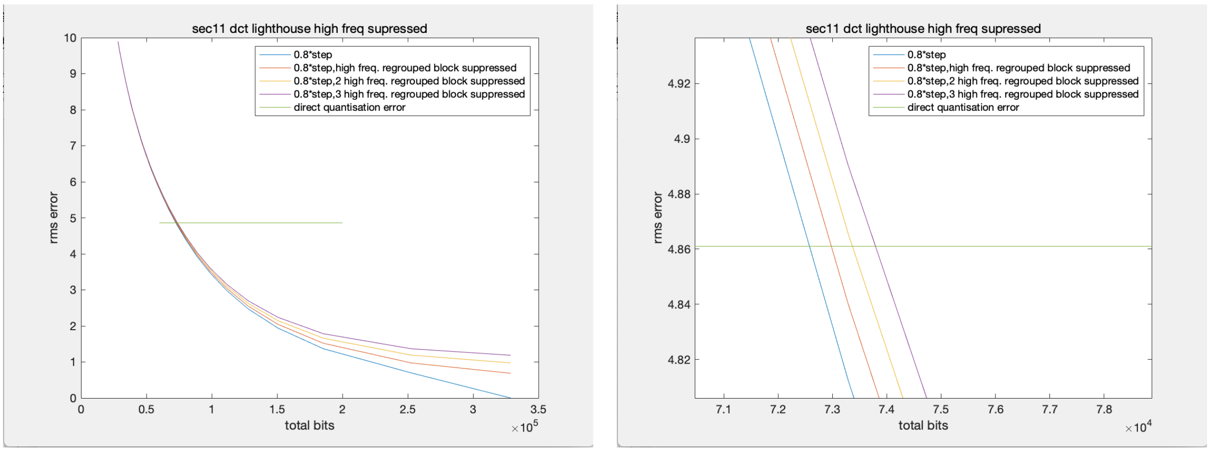


Figure 2: Left - Varying first rise to step size ratio for the DCT with highest frequency coefficients suppressed. Right - Zoomed in region around the reference RMS error.

Since most of the energy content is in the low-pass images, another technique is to completely suppress some higher frequency sub-images or DCT coefficients. We can effectively do so by setting a very large first rise step for the highest frequency sub-images. However, from fig. 2, we see that suppressing the highest frequency DCT coefficients resulted in an increased RMS error using the same number of bits. Again we find similar trends using the LBT and DWT, as shown in Appendix fig. 9 and fig. 10, and with the bridge and flamingo images. Therefore, we conclude that suppressing high frequency coefficients or sub-images is not an optimal choice.

4 Huffman coding

We will now introduce the final coding process of an image compression system. A coding scheme converts an image into a vector of bits that can be reconstructed to the original image.

The Huffman code is generally a relatively efficient scheme for coding data. However, a major drawback is that it does not work well if the probability of any single event exceeds 50%. As we have seen, a large proportion of image coefficients will be quantised to zero after passing through an energy compaction process. By combining each non-zero coefficient with the number of preceding zero coefficients into a single event, we can ensure each individual event has probability well below 50%. This technique is used by the baseline JPEG specification and many other image compression systems. In the JPEG specification which uses 8x8 DCT transformations, the 63 AC coefficients are coded by scanning the 8x8 blocks in a zig-zag manner [1]. This places high frequency coefficients at the end so the end-of-block codeword can be used to effectively terminate the encoded sequence early, as high frequency

Image	Energy compaction method	Compression ratio	
		Default Huffman table	Custom Huffman table
Lighthouse	8x8 DCT	3.198	**3.291**
	4x4 LBT	2.720	**2.951**
	5-level DWT	2.875	**3.350**
Bridge	8x8 DCT	1.689	**1.743**
	4x4 LBT	1.526	**1.601**
	5-level DWT	1.654	**1.792**
Flamingo	8x8 DCT	2.259	**2.276**
	4x4 LBT	1.897	**2.078**
	5-level DWT	2.096	**2.389**

Table 2: Compression ratio comparison of JPEG default Huffman table and custom Huffman table

components are often quantised to zero.

We can adopt the LBT and DWT transformed images to look similar to DCT block transformed images such that we can apply the same zig-zag scanning strategy. The LBT provides compression benefits when coded several sub-blocks at a time, so we can make a 4x4 LBT look like a 16x16 DCT by regrouping intensity values 4 spaces apart to be adjacent together. Similarly, a 5-level DWT can be arranged into a 32x32 DCT block by grouping coefficients from the same square spatial area.

The coding process uses a Huffman table to determine the codeword used to encode each event. The JPEG specification provides a recommended Huffman table for luminance AC coefficients which the committee have determined to be suitable for typical applications. However, this does not guarantee that the assigned codeword lengths will be optimal according to the probability of each coefficient, as each image will be different. We can design our own Huffman table by considering the probabilities of each event by generating a histogram, hence guaranteeing an optimal Huffman coding table. However, in order for the receiving end to decode the transmitted image, the custom Huffman table needs to be sent along with the compressed image. This incurs a 1424 bits penalty for sending the required tables, which means that using the default Huffman table might actually be a better strategy overall.

Once again using the direct quantisation with a step size of 17 as the reference scheme, the compression ratios using the default and custom Huffman tables are found in Table 2. We find that the compression ratio is higher when using a custom Huffman table compared to using the JPEG default Huffman table for all test images and using all energy compression methods, even with the 1424 bits penalty for transmitting the custom table taken into consideration. Therefore, we conclude that it is nearly always beneficial to use a custom Huffman table to achieve the best compression ratio.

5 Structural similarity index

In our investigation so far, we have only used the RMS error as a quantitative measure of difference between the decoded and original images. However, the RMS error does not always reflect the actual visual quality of the image, as it only considers the absolute error between pixels. The structural similarity index (SSIM) is a perception-based model that considers pixel differences as perceived change in structural information [2]. The SSIM is a better metric for human perceived quality as it combines the visual impact based on luminance, contrast, and structure.

The luminance, contrast and structural terms are given by:

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{1}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{2}$$

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{3}$$

where μ_x, μ_y, σ_x, σ_y, and σ_{xy} are the local means, standard deviations, and covariance of the input

Image	Energy compaction method	SSIM
Lighthouse	8x8 DCT	0.760
	4x4 LBT	**0.784**
	5-level DWT	0.755
Bridge	8x8 DCT	0.942
	4x4 LBT	**0.944**
	5-level DWT	0.939
Flamingo	8x8 DCT	0.880
	4x4 LBT	**0.887**
	5-level DWT	0.876

Table 3: Comparison of SSIM index for different energy compaction methods

images x and y. $C_1 = (0.01 \cdot L)^2$ and $C_2 = (0.03 \cdot L)^2$, with L being the dynamic range of the pixel values.

The overall SSIM index is then given by

$$\begin{aligned} SSIM(X, y) &= l(x, y) \cdot c(x, y) \cdot s(x, y) \\ &= \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^@ + C_1)(\sigma_x^2 + \sigma_y^2 + C_2))} \end{aligned} \tag{4}$$

The SSIM is calculated using a sliding Gaussian window of size 11x11 and results in a value between -1 to 1, with 1 indicating identical input images and 0 indicating no structural similarity.

In Table 3, the SSIM index are shown for the 3 test images when they are are compressed using the 3 candidate energy compaction schemes to give the same RMS error. The RMS error when directly quantised with step size of 17 is used as the reference, with RMS error of 4.934 for the lighthouse image, 4.876 for the bridge image, and 4.888 for the flamingo image. The LBT gives the best SSIM value compared to the DCT and DWT. It is expected that the LBT performs better than the DCT, as the LBT is an improvement scheme based on the DCT which removes the blocking artifacts, hence improving structural integrity. It is surprising that the DWT performs worst out of all 3 energy compaction method. This is most likely because the test images all contain complex details, which the DWT tends to blur and smooth out.

We will continue our investigation of image compression techniques using both the RMS error and SSIM index as image quality metrics.

6 Frequency-dependent quantisation

In the quantiser schemes we have explored so far, the same quantisation table is used for all coefficients and sub-images, regardless of frequency. However, as we have noted in Section 3, the higher frequency components generally contain less energy and will incur only a low quality penalty if we completely suppress them. We can extend this idea to all frequency components, where the lower frequency coefficients are quantised using a smaller step size, and the quantisation step size increases progressively as frequency increases.

The JPEG standard provides a recommendation for luminance frequency-dependent quantisation table [1]. The recommended table is shown in Table 4 and is derived empirically based on psychovisual thresholding with test subjects.

We will attempt to compress the 3 test images to just below 40960 bits (5 kB), which is our final target of our image compression system. The RMS error and SSIM index when using a standard uniform quantiser and the frequency-dependent quantiser is shown in Table 5 and the reconstructed images are shown in fig. 3. We find that the frequency-dependent quantiser produces much better visual quality compared to the standard quantiser. The signature block artifacts of the DCT can be seen clearly in the standard quantiser image, but is drastically reduced in the frequency-dependent quantiser. The gradient and details on the shed and the lighthouse can be seen much more clearly and undistorted in the frequency-dependent quantiser image. Interestingly, the DCT is able to achieve a better RMS error performance compared to the LBT across all 3 images, whereas the SSIM is able to accurate reflect the

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Table 4: JPEG recommended luminance quantisation table

Image	Quality metric	DCT		LBT	
		Standard quantisation	Frequency-dependent quantisation	Standard quantisation	Frequency-dependent quantisation
Lighthouse	RMS error	**8.068**	9.424	9.122	9.259
	SSIM	0.629	**0.684**	0.617	0.613
Bridge	RMS error	**12.450**	12.779	13.340	13.310
	SSIM	0.699	**0.700**	0.677	0.687
Flamingo	RMS error	**11.502**	12.557	12.738	13.048
	SSIM	0.696	**0.720**	0.686	0.666

Table 5: RMS error and SSIM index comparison of DCT and LBT with and without frequency-dependent quantisation

(a) Standard quantisation

(b) Frequency-dependent quantisation

Figure 3: Comparison of standard and frequency-dependent quantiser using the 8x8 DCT on the lighthouse image

4	6	7	8	10	13	18	31
6	9	10	12	15	20	28	48
7	10	12	14	18	23	32	55
8	12	14	17	21	27	38	65
10	15	18	21	26	33	47	80
13	20	23	27	33	43	61	103
18	28	32	38	47	61	86	146
31	48	55	65	80	103	146	250

Table 6: Fu (2015) proposed novel luminance quantisation table

Image	Quality metric	DCT		LBT	
		JPEG recommended quantisation table	Fu (2015) novel quantisation table	JPEG recommended quantisation table	Fu (2015) novel quantisation table
Lighthouse	RMS error	9.424	0.947	**9.259**	9.627
	SSIM	0.684	**0.689**	0.613	0.658
Bridge	RMS error	12.779	**12.647**	13.310	13.461
	SSIM	**0.701**	0.698	0.687	0.666
Flamingo	RMS error	12.557	**12.170**	13.048	13.353
	SSIM	0.720	**0.726**	0.666	0.695

Table 7: RMS error and SSIM index comparison of DCT and LBT using JPEG recommended quantisation table and Fu (2015) novel quantisation table

improved visual quality between the images. This confirms that the SSIM is a better quality metric when judging visual quality and is able to more accurately reflect human perception.

The quantised LBT images are shown in Appendix fig. 11, and does in fact visually appear much worse than the DCT images, as expected from Table 5.

Following literature review, we find that Fu Q. (2015) proposed a novel quantisation table that is more suitable for biorthogonal transformed images [3]. They proposed a novel luminance quantisation table, shown in Table 6, which can be used directly in baseline JPEG as well without compatibility problems. We again perform the same analysis as before using the new luminance quantisation table. From Table 7 and fig. 4, we find that the proposed quantisation table provides visual quality benefits for the lighthouse and flamingo images. There is a large improvement when used with the LBT and only a small improvement when used with the DCT. This is because the novel quantisation table is optimised for biorthogonal transformed images. There is a small SSIM degradation for the bridge image. This is likely due to the high structural complexity in the image making the image difficult to compress in general, resulting in noise in the metrics. The bridge image compressed with LBT using the two different quantisation tables is shown in Appendix fig. 12, which shows that there is practically no visual difference between the two images.

7 Final scheme

Working in a pair with Bryan Zheng, we tested the various techniques and strategies explained above. Bryan was mostly responsible for the tests relating to the centre-clipped linear quantiser and researching various alternative error measures. I was mostly responsible for implementing the Huffman coding functions and investigating the usage of frequency-dependent quantisers with the DCT and LBT.

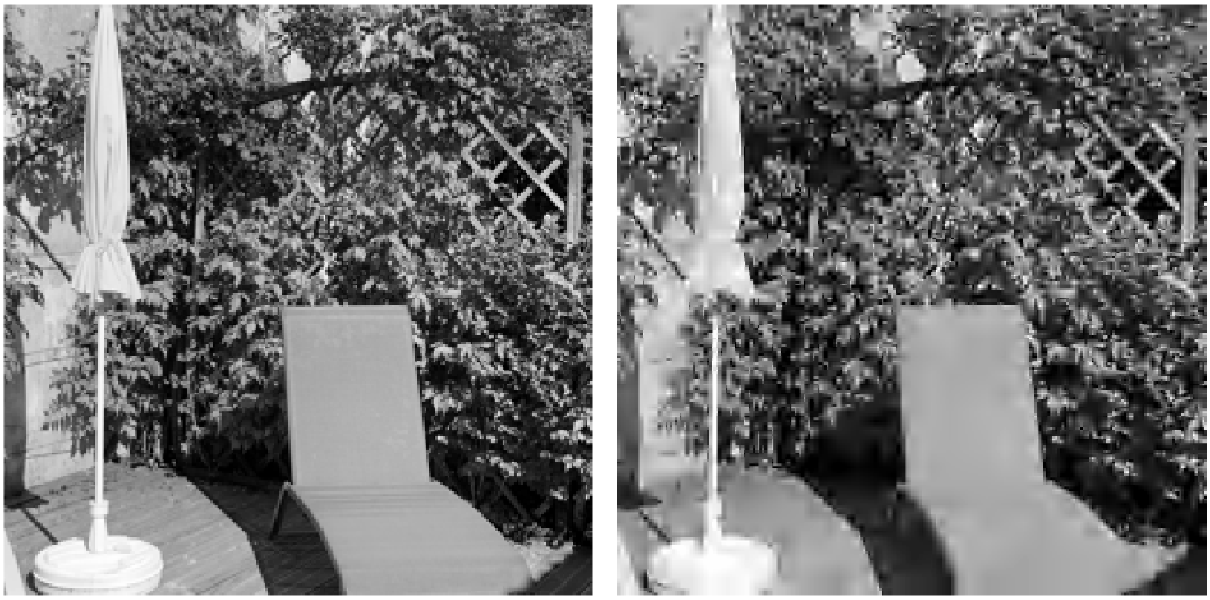
After performing the various tests, we determined that the 5-level DWT using an equal MSE quantisation scheme is the best energy compaction strategy. The 5-level DWT offered the highest compression ratio out of all 3 candidate schemes whilst still maintaining a reasonable visual quality. The DWT often provided images that were the most pleasing to look at, with few artifacts and distortions.

This year's competition photo, shown in fig. 5, is set in a garden and consists of a lawn chair and a patio umbrella. The image consists of both highly complex textures by the plants along the background, and smooth simple textures by the lawn chair in the foreground. This makes it a particularly tricky

(a) JPEG recommended quantisation table (b) Fu (2015) novel quantisation table

Figure 4: Comparison of JPEG recommended quantisation table and Fu (2015) novel quantisation table using the 4x4 LBT on the lighthouse image

image to compress. Unfortunately, our chosen scheme was only able to achieve an RMS error of 19.65 and SSIM of 0.65 when compressed to just below 5 kB in size. This meant that we were placed fourth in the design competition.



(a) Original (b) Compressed with chosen scheme

Figure 5: Left - Original competition image. Right - Compressed using the chosen 5-level DWT scheme.

8 Conclusion

In this design project, we investigated the 3 key components that make up a complete image compression system: the input filtering process, the quantisation process, and the coding process. A final compression scheme was chosen consisting of suitable strategies from each process through extensive testing and exploration of their qualities and weaknesses. The final compression scheme was entered into the design competition. It was a shame that our team's chosen compression scheme did not perform very well compared to other designs. Retrospectively, there are a number of obvious gaps in our final compression system and would benefit from further research and investigation, including strategies such as using further optimised coding processes using Arithmetic coding or using differential coding for DC coefficients, and more careful consideration of error measures. The zig-zag scanning strategy is not designed for the DWT and is likely the cause for poor practical compression ratios. However, this reflects the highly complex nature of modern image data compression systems and the difficulty of designing such a system that is expected to work well on vastly different images and processed on many devices.

References

[1] CCITT (1992). T.81: Information technology - Digital compression and coding of continuous-tone still images - Requirements and guidelines.

[2] Wang, Zhou; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), pp. 600-612.

[3] Fu, Q.; Jiang, B.; Wang, C.; Zhou, X. (2015). A Novel Deblocking Quantization Table for Luminance Component in Baseline JPEG. *Journal of Communications*, 10(8). Available at: https://pdfs.semanticscholar.org/a07a/d800b5a6f405a5349da2db338c90ce958904.pdf [Accessed 27 Mar. 2020]

Appendix

Figure 6: Original flamingo image

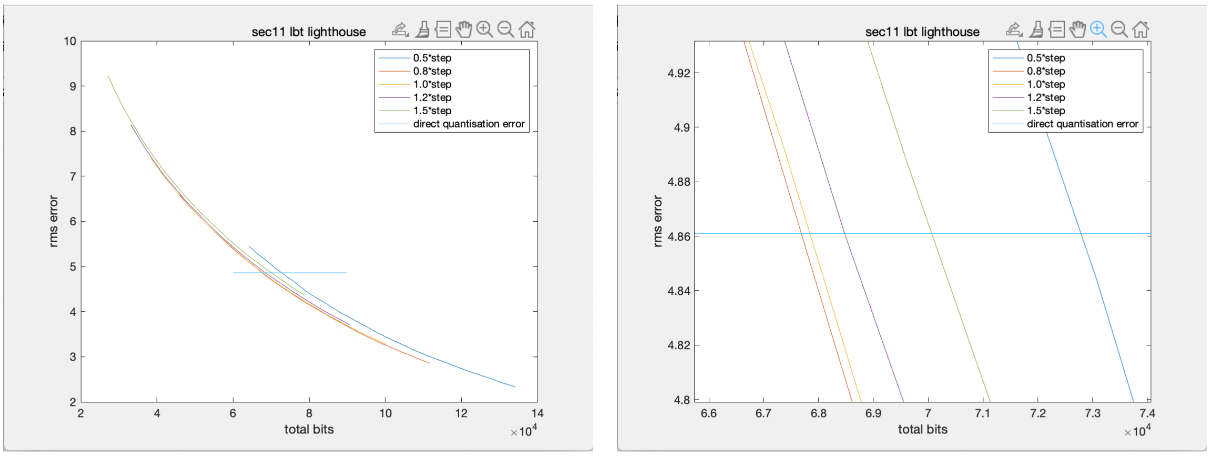


Figure 7: Left - Varying first rise to step size ratio for the LBT. Right - Zoomed in region around the reference RMS error.

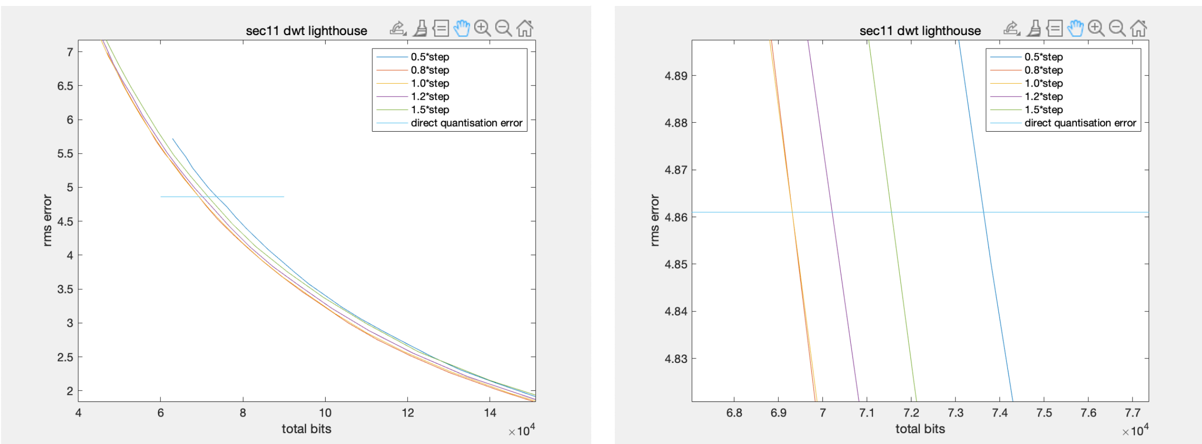


Figure 8: Left - Varying first rise to step size ratio for the DWT. Right - Zoomed in region around the reference RMS error.

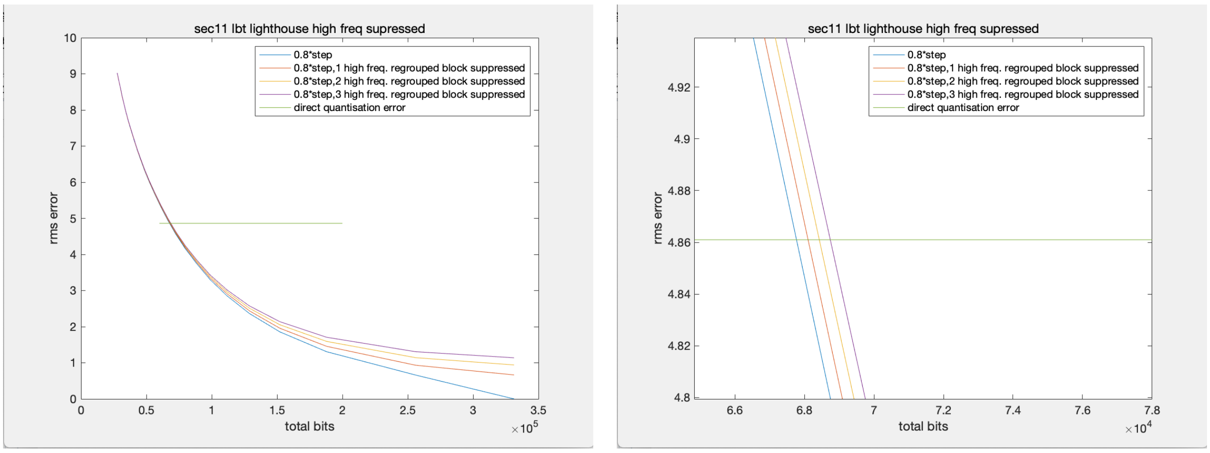


Figure 9: Left - Varying first rise to step size ratio for the LBT with highest frequency coefficients suppressed. . Right - Zoomed in region around the reference RMS error.

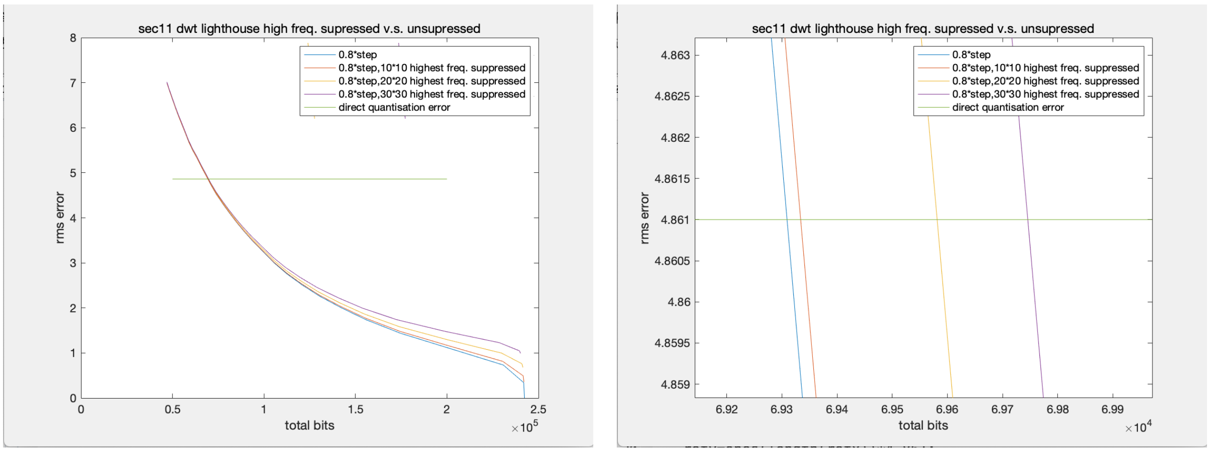


Figure 10: Left - Varying first rise to step size ratio for the DWT with highest frequency sub-images suppressed. . Right - Zoomed in region around the reference RMS error.

(a) Standard quantisation

(b) Frequency-dependent quantisation

Figure 11: Comparison of standard and frequency-dependent quantiser using the 4x4 LBT on the lighthouse image

(a) JPEG recommended quantisation table

(b) Fu (2015) novel quantisation table

Figure 12: Comparison of JPEG recommended quantisation table and Fu (2015) novel quantisation table using the 4x4 LBT on the bridge image