Analysis of the variability in digitised images compared to the distortion introduced by compression

Sean Martin
British Library
Boston Spa
Wetherby, LS23 7BQ, UK
+44 1937 546716
Sean.Martin@bl.uk

Prof Malcolm Macleod
QinetiQ Ltd
St Andrews Road
Malvern, WR14 3PS, UK
+44 1684 543796
mdmacleod@iee.org

ABSTRACT
The paper evaluates the noise which is present in digitised images of very high quality and the noise/error which results when such images are compressed and then decompressed. The variations between pairs of captured images of identical material were compared and the two best pairs of images were identified. The variations between these pairs were then compared with the variations introduced by compression and decompression of those images. We found that even lossy compression can result in significantly lower variation than that between the best pairs of original images caused by imaging noise. We report the results of a qualitative questionnaire which are in good agreement with the quantitative assessment. The conclusions suggest that given the extent of noise in the imaging process the current practice of storing lossless master digitised images could be replaced by the use of more compact compressed images, arguably with no loss of quality.

Keywords
digitisation, camera, scanner, digitised image, camera noise, JPEG 2000, image compression, PSNR

1. INTRODUCTION
The motivation for the work described in this paper arose when some simple experiments were conducted in a digitisation studio. A particular item was imaged several times and the resulting images were examined visually. The extent of the differences between the images at a detailed level was surprising, and this prompted further investigation. This evolved into the structured process which is described in this paper. Meanwhile a review of the literature, summarized in Section 2, identified several papers which discuss noise in the imaging process and its consequences.

Three physical items were each imaged in colour with seven devices that produce images that can be compared automatically. Each item was imaged five times in rapid succession with the same device, a camera or a scanner, without moving the item, and without changing the background lighting. The variations between 210 pairs of these images were assessed. Additionally, two selected "best" pairs of images were subject to detailed further examination. One pair was produced by a top of the range camera, and the other by a regular production camera.

The differences between these pairs of images were compared with the variations caused by compressing one of the images in each pair in a lossy manner. It was found that for modest amounts of compression, the variations introduced by compression were less than the variations between the original lossless master images.

This quantitative assessment was complemented by a qualitative questionnaire in which images were compared by eye and respondents were invited to indicate which pairs of images had least or most differences. The qualitative assessment produced results in line with the quantitative assessment. The questionnaire also included examples where greater compression was applied and respondents were asked to indicate whether the resulting images were considered perfect, acceptable, marginal or unacceptable. It was found that modest compression could be applied without compromising the perceived visual quality of an image, even when the image has been greatly magnified.

This final part of the questionnaire produced an interesting result. In some cases the alternative lossless original master images were deemed merely acceptable, whereas compressed images with very small loss were deemed to be perfect.

These results, particularly the last one, question the need for retaining digitised images in a lossless manner. It is clear there are noise-induced differences between original high quality master images, while a mild level of compression can result in much less variation. When applied in an appropriate manner this could reduce storage costs, conservatively by 30-70% compared with storing lossless JPEG 2000 files. For bulk digitisation greater cost saving is possible. The choice might be dependent on the subject matter but a strong case is put forward that a minimum of the order of 30% compression is achievable with little reduction in perceived quality or value.

2. NOISE AND IMAGING
2.1 Noise in the Imaging Process
While noise in imaging is discussed widely in the literature, there has been limited attention regarding the extent and nature of noise that occurs in the imaging process and is thus present in digitised images.

(Liu, et al. 2008) [9] states that there are five primary noise sources in a camera with a CCD (charge coupled device) sensor. These are: fixed pattern noise (FPN), dark current noise, shot noise, amplifier noise and quantization noise. These arise in the successive processes by which photons cause electron activity, which is amplified and then digitised - noise is introduced at each stage. The paper discusses the statistics of noise and how noise can arise in the colour that is recorded - known as colour noise. (Faraji and MacLean 2006) [5] describe signal-independent noise and signal-dependent noise, and they characterise noise sources in a similar way to [9] including photon noise, FPN, amplifier noise and readout noise. They refer to an extensive discussion of noise in (Janesick 2001) [6] and they also note that at low light levels the noise is independent of the signal, at mid light levels the noise becomes signal dependent – arising from shot noise, photon noise and dark noise, typically with Poisson distributions. At high light levels FPN proportional to the signal dominates. (Chen, et al. 2009) [2] also characterise noise as FPN and random noise.

(Kurosawa, Kuroki and Akiba 2009) [8] establish that it is possible to identify that an image, or more specifically a series of images in a video, were taken by a particular camera. Distinctive FPN can be produced by individual "hot" pixels and the spatial arrangement of these pixels can be recognised in an image, and the camera thereby identified. The ability to identify the camera from its noise signature is analogous to identifying a gun from a bullet fired from it.

(McHugh) [10] gives an excellent tutorial on noise in digital cameras, and states that digital cameras produce three types of noise: random noise, FPN, and banding noise, noting that the latter is highly camera-dependent. The following example pictures from [6] are reproduced by permission:

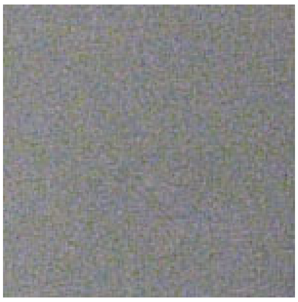

Figure 1: Example of random noise

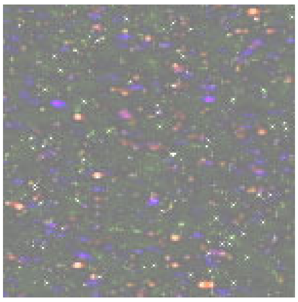

Figure 2: Example of fixed pattern noise

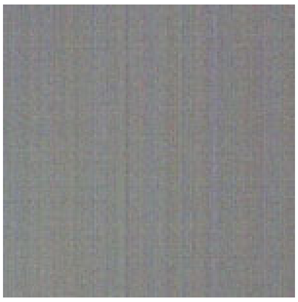

Figure 3: Example of banding noise

[10] also observes that noise is more prominent in darker regions, and that noise can comprise fluctuations in both colour and luminance, where, for example, chroma noise can be evident as colour superimposed on a grey portion of an image. Noise can be both fine- and coarse-grained in texture.

The signal to noise ratio (SNR) is a useful and universally used way of comparing the relative amounts of signal and noise in any electronic system; high ratios will have very little discernible noise whereas the opposite is true for low ratios.

The literature concerning noise and images arises from a wide range of disciplines, including astronomy with low light levels, and medical imaging, such as (Belbachir and Goebel 2006) [1] which discusses noise in the incoming photon stream. Many of the cited papers discuss schemes for reducing noise, and they therefore discuss the sources of noise and models for it.

The concept of noise in a camera image is for some an abstract notion. However, it may be helpful to relate it to the hiss heard in an old audio recording. The hiss is noise – if the record is replayed then the hiss could be different, although the symphony may sound the same. The value is in the symphony, whereas it is rarely of any value to record and reproduce faithfully the hiss that occurred on one particular occasion.

2.2 Image compression and noise

Image compression is an important technology for reducing the amount of storage required to hold images, or the communication capacity required to transmit them.

There is a widespread opinion in the library and archive community that it is vital that images be stored losslessly. The noise in such an image would also be preserved. However, as is well known, random noise is inherently difficult to compress. This can lead to a significant proportion of a lossless compressed digitised image file being used to reproduce exactly the noise in the image.

If instead lossy compression is used, then the decompressed image will differ from the original. Since the compression is lossy it is likely that it will fail to encode the noise completely, as it is difficult to compress. However, with a low degree of compression, and hence loss, it is likely that the signal in the image will remain almost intact other than a small amount of distortion that is introduced. Provided the power (or extent) of the distortion that is introduced is less than the power of the original input noise, it can be argued that the decompressed image has exactly as much quality (SNR) as the original. This hypothetical consideration is not exactly what occurs, but it demonstrates the argument that the artefacts in an image reconstructed after compression may represent no loss of quality compared to an original master image with the unavoidable noise present in it.

To explore this hypothesis requires a detailed comparative analysis of input noise and the noise resulting from compression. It is that which is the goal of this paper.

3. EXPERIMENTAL WORK

3.1 Introduction

Initial experiments were conducted in which the same item was imaged several times in a manner designed to be as close to identical as possible. The resulting images were compared and there were visually obvious significant differences between the magnified images. These experiments led to the development of a systematic process for characterising the noise in an image, as described below.

Each original image created in this process is a lossless master image file and therefore 'authentic', but the images are different from each other because of the presence of noise. We developed a method whereby we could compare (a) the variations between these lossless master files with (b) the variations, usually called degradation, introduced by compressing a master file in a lossy manner.

The full detail of the process is now described.

3.2 Method of Imaging

Three separate physical samples were selected and the same samples were imaged multiple times with different cameras and scanners that are, or were, in regular use in the digitisation studios in two national libraries, the Norwegian National Library and the British Library. The three samples that were imaged are approximately A3 in size.

A sample was placed under a camera or scanner and imaged multiple times in quick succession with no deliberate change in the background lighting conditions in the digitisation studio.

Each sample was imaged in this way with ten different cameras or scanners. However, the images from three devices were later discarded since their images were too variable to make detailed comparisons practicable. This left sets of high quality images from seven devices which were appropriate for detailed inspection. These included one scanner and six cameras, all in regular production use. Two were automated page turning machines, and each of those had two standard professional Phase One backed digital cameras. Images taken with these devices were designated as N01-N04. A top of the range Hasselblad specialist camera was designated as N05. The scanner was designated as N06, and a separate Phase One backed digital camera was designated as B07. An N indicates that the images were taken at the Norwegian National Library, and B indicates that the images were taken at the British Library. As will be seen later there is broad consistency of quality between the best pairs of images produced by these devices.

However, these sets of images also showed detectable variations and so we undertook more experimentation on the manner in which the images were taken. The nature of the variations is discussed later.

We experimented by taking multiple images with a longer (ten second) delay between them – to see if the action of imaging introduced a small vibration that caused a wobble in the image. However, the resulting variations were similar to those in an original set of five images.

We conjectured that there might be a lensing effect arising from density variations in the air flow between the camera and the item. We therefore set up an experiment where one half of the item had air blown across it with a fan while the other half had no air flow. We found that the two halves of the image had similar variations to an original set of five images, and we could detect no differences caused by the difference in air flow.

We emphasise again that each image we used in the experiments described below is an example of what an archivist would regard as a 'valid master file'.

3.3 Quantitative Assessment of Images

Within a set of five images of the same item taken with the same camera or scanner, ten pairwise comparisons are possible. (The first image is compared with four others, the second image with three others, and so on.)

As explained above, three separate items were imaged, and seven devices used to produce sets of images for automated quantitative assessment.

There were therefore 210 pairwise comparisons of images available. An immediate impression was that for each pair of images there are significant variations between them, despite each image being an authentic master image.

The experiments assessed the similarity between a pair of images. However visual inspection showed that there were often small, but quite noticeable, lateral shifts between images, and this greatly complicated the comparison process. The comparison thus had to be preceded by aligning the two images to obtain the best correlation score.

A simple hill climbing technique proved effective for correcting shifts that were small compared with the size of features in the image. This worked well on the images from the selected seven devices. Reference was made earlier to devices that delivered images sufficiently different to make comparison difficult – in one case because the observed shifts were large, for example 70-100 pixels, whereas a typical feature might only be ten or so pixels across. A simple hill climbing algorithm was no longer effective since it stopped at intermediate local maxima and failed to find the overall best fit. In another case the device produced images whose width and height dimensions were so significantly different as to make comparison difficult.

A simple PSNR (peak signal to noise ratio) was used as the correlation metric, though other metrics are possible and have been reported to produce better comparisons between digitised images. As is customary, the PSNR is expressed in logarithmic (deciBel, dB) units, which give the best correspondence with the perceived quality.

It was often also found that the lateral shift was not constant and could vary by a small amount across the image – this is a form of spatial distortion between a pair of images. Often there is a slow progression, with the lateral shift slowly changing or drifting across the compared images. We also observed one case where the extremities of the compared images diverged – in effect there had been a small change in the magnification.

The method of comparison took the lateral shift into account by considering a portion, or tile, from each image in turn and then aligning and correlating each pair of tiles independently. A PSNR metric was then calculated for the entire image as an aggregate of the metric from each optimally aligned pair of tiles. A typical tile size used was 400 x 400 pixels. A tile size of 100 x 100 was also tested and produced similar results.

The lateral shifts were usually not an exact whole number of pixels. The alignment technique was therefore extended such that once there was optimum alignment based on shifts of a whole number of pixels between a pair of tiles, each tile was then expanded by interpolation, and then aligned to an accuracy equivalent to a fraction of a pixel in the original image. Early experimentation showed that a bi-linear interpolation was as effective as bi-cubic interpolation, and so bi-linear interpolation was used for this further analysis.

With this enhancement, each of the 210 pairs of images was aligned to 0.25 pixels.

3.4 Quantitative Assessment Results

Table 1 shows a summary of the comparison of pairs of images. The first column cites the identity of the imaging device, referred to as N01-N06 or B07. The six remaining columns record for each of the three sample documents A, B and C the PSNR results. "Av" is computed by averaging the PSNR values from the ten pairwise comparisons (not in dB form) and then converting the average to dB. "Max" is the maximum within the set. A standard colour coding has been applied to help highlight particular scores where red indicates a low score and blue a high score.

Table 1: PSNR in dB comparing images without shifting

Sample	A Av	A Max	B Av	B Max	C Av	C Max
Device						
N01	30.790	36.742	31.400	37.024	36.742	36.942
N02	32.712	36.431	36.874	36.925	36.390	36.656
N03	30.083	37.042	32.009	38.277	32.048	37.916
N04	30.095	37.373	30.391	37.823	36.123	37.348
N05	41.449	42.128	42.317	43.000	42.197	42.508
N06	29.851	31.286	28.011	31.335	29.669	30.961
B07	19.479	33.842	22.878	38.862	16.687	36.347

We see that:

Device N05 has consistent and relatively high scores. For each of the three samples the maximum for N05 is only a little greater than the average – this indicates that the ten pairwise comparisons are quite consistent. The N05 scores are also consistent across the three samples A. B and C.

By contrast device B07 shows much greater difference between the average and maximum scores; for example an average of 16.687dB and a maximum of 36.347dB for sample C. This indicates considerable variation between the individual scores, as will be confirmed later.

Devices N01 to N04 are all supplied by the same manufacturer. The consistency of their images falls between those for B07 and N05, with a greatest difference between average and maximum of around 6dB and the least being only 0.05dB.

Table 2 shows the results of comparing matching tiles from pairs of images. The tiles were processed independently within each pair of images in the manner previously described, where the tiles from the different images were aligned for best fit to the nearest pixel and an aggregate PSNR value was derived for the entire image.

Table 2: PSNR in dB comparing images after shifting

Sample	A Av	A Max	B Av	B Max	C Av	C Max
Device						
N01	30.811	36.742	31.400	37.024	36.742	36.942
N02	32.712	36.431	36.874	36.925	36.390	36.656
N03	31.020	37.042	33.780	38.277	33.395	37.916
N04	32.424	37.373	30.494	37.823	36.123	37.348
N05	41.449	42.128	42.317	43.000	42.197	42.508
N06	29.857	31.286	29.478	31.335	29.850	30.961
B07	25.861	33.842	29.048	38.862	24.171	36.347

We see that:

Devices N02 and N05 have identical results in tables 1 and 2 indicating that all their pairs of images are already aligned - there are no lateral shifts between them.

For device B07 the average scores increase from table 1 to table 2 – this demonstrates that the shifting algorithm is able to improve the alignment between some pairs of images. However, for B07 and the other four devices the maximum scores remain unchanged indicating that the pairs of images which generated them were already optimally aligned.

The information from table 2 is summarized in table 3 which records three scores for each of the devices. These are the average of the averages for the three items A, B and C, the average maximum for the three items, and finally the overall maximum value.

Table 3: Average and Maxima from Table 2

Device	Average of Averages	Average Maximum	Maximum of Maxima
N01	32.984	36.903	37.024
N02	35.325	36.671	36.925
N03	32.732	37.745	38.277
N04	33.014	37.515	37.823
N05	41.988	42.545	43.000
N06	29.728	31.194	31.335
B07	26.360	36.350	38.862

The last column shows quite consistent results. Devices N01 to N04 are similar, with maximum scores of 36.9 to 38.3dB. Device B07, which as noted earlier showed considerable variations, had a slightly better maximum score of 38.9dB. Device N05 shows the best results at 43.0dB. All these devices were cameras, whereas device N06 was a scanner. It has a noticeably lower score of 31.3dB.

As reported earlier, visual inspection of pairs of images showed that there was still a discernible shift between pairs of images even though the images were aligned to the nearest pixel. As explained we therefore interpolated pixels and repeated the alignment of tiles within an image, to the nearest interpolated pixel. Table 4 presents a summary of the results when interpolating and shifting were applied to each pair of images.

The results in table 4 show small improvements when compared with the results in table 2. Identical results would not be expected since the basis of comparison has changed. The images for device N05 show increases of around 1.5dB between the two tables for both the average and the maximum scores. The corresponding scores for other devices show greater increases in the range 2-6dB.

Table 4:PSNR in dB comparing images with shifting and interpolation to 0.25 pixel

Sample	A Av	A Max	B Av	B Max	C Av	C Max
Device						
N01	35.485	38.234	35.804	38.500	38.182	38.418
N02	36.128	37.898	38.354	38.413	37.818	38.141
N03	35.671	38.437	37.115	39.854	37.265	39.478
N04	36.014	38.924	36.040	39.371	37.502	38.894
N05	42.944	43.724	43.757	44.536	43.698	44.070

N06	34.445	34.941	34.670	34.909	34.260	34.570
B07	30.718	35.947	32.183	40.211	27.075	37.162

The best overall individual match was obtained with device N05 and item B where the ten individual comparisons without interpolation between pairs of images are shown in Table 5. These images are designated N05B1-N05B5.

Table 5: PSNR in dB comparing pairs of images using device N05 and Item B

Images	N05B2	N05B3	N05B4	N05B5
N05B1	42.774	42.432	41.711	41.228
N05B2		42.890	42.192	41.724
N05B3			42.747	42.470
N05B4				**43.000**

Table 5 shows that the overall best match pair was between N05B4 and N05B5. This pair was used in later qualitative assessments. As noted earlier these pairs of images are already aligned and so shifting produces identical results. Table 6 shows the corresponding results after interpolation.

Table 6: PSNR in dB comparing pairs of images with interpolation for device N05 and Item B

Images	N05B2	N05B3	N05B4	N05B5
N05B1	44.260	43.865	43.074	42.540
N05B2		44.408	43.617	43.098
N05B3			44.243	43.931
N05B4				44.536

The best match for a standard Phase One backed digital camera was obtained with device B07 and item B. Table 7 shows the ten individual comparisons between pairs of images. The five images are designated as B07B1-B07B5. It is worth noting that some of the other image pairs show significant differences, such as the pair B07B1 and B07B2 which has a remarkably low PSNR of 14.2dB.

Table 7: PSNR in dB comparing pairs of images for device B07 and Item B

Images	B07B2	B07B3	B07B4	B07B5
B07B1	14.168	16.265	16.301	16.130
B07B2		22.401	22.037	22.683
B07B3			30.500	**38.862**
B07B4				29.433

Table 8 shows the results after shifting by integer pixels (i.e. without interpolation). As noted earlier the best match in this set is between B07B3 and B07B5 and its score is not improved by shifting. The score for the poorest image pair (B07B1 and B07B2) has improved but is still significantly below the best value. Table 9 shows the results after shifting and interpolation.

Table 8: PSNR in dB comparing pairs of images with shifting for device B07 and Item B

Images	B07B2	B07B3	B07B4	B07B5
B07B1	25.834	27.667	28.625	27.317
B07B2		28.493	25.330	28.417
B07B3			30.500	**38.862**
B07B4				29.433

Table 9: PSNR in dB comparing pairs of images with shifting and interpolation for device B07 and Item B

Images	B07B2	B07B3	B07B4	B07B5
B07B1	29.529	32.200	33.629	31.097
B07B2		33.230	27.409	34.101
B07B3			30.775	40.211
B07B4				29.645

4. ASSESSMENT OF COMPRESSION

4.1 Analytical work

The previous section identified two sets of 'most similar' images; they were of item B, from devices N05 and B07. These sets were N05B and B07B, and in each set there are five images. Each image was next encoded into a set of JPEG 2000 files with various degrees of compression.

JPEG2000 is becoming increasingly used within the archival community. It supports lossless (reversible) compression using an integer based encoding and also lossy (irreversible) compression using floating-point encoding. The latter can be configured to minimise the loss – where perfect computation would incur no loss but floating point calculations are subject to round off error and this does cause loss. This technique is colloquially known as “minimally lossless”. If lossless integer encoding is taken as a baseline, then minimally lossless encoding typically introduces variations at around 50dB PSNR but with a reduction in file size of 30-40% compared with a lossless JPEG 2000 encoding.

Each image in both sets was encoded in a range of ways: lossless, minimally lossless, and then with a series of lossy compression factors designated as G2 to G12, indicating progressively increasing compression. Each compressed image was compared with the original using PSNR and a compression ratio was derived from the size of the image files. The baseline chosen for the compression ratio was the size of a lossless JPEG 2000 file. There was a particular reason for this. An organisation wishing to store lossless files could choose to use the TIFF format; however, JPEG 2000 offers a lossless format. Those experiencing cost pressure are likely to choose the latter and hence this is an appropriate baseline for determining the additional cost saving in adopting lossy compression. (It should be noted that there are concerns about the ability of JPEG 2000 to retain colour space information; however, when the effect of noise is taken into account it could be argued that a camera is not able to produce a sufficiently accurate colour to make this relevant.)

The compression ratio of a lossless JPEG 2000 file is thus deemed to be 1.0. (A JPEG 2000 lossless file is typically 30-40% smaller than an uncompressed TIFF file.)

Kakadu software was used to encode the images using the British Library JPEG 2000 encoding profile. However the tool used to derive PSNR and compression ratio used the Leadtools software library to decode the JPEG 2000 images.

Table 10 shows the average PSNR and average compression ratio for each way of encoding the images in each of the two sets N05B and B07B. When two images are identical then the PSNR between them is defined by the PSNR algorithm as infinity; that shows that compression was lossless.

Table 10: Compression ratio and PSNR for N05B & B07B

	Image Set N05B		Image Set B07B	
Compression designation	Compres-sion ratio	PSNR dB	Compres-sion ratio	PSNR dB
lossless	1.00	Infinity	1.00	Infinity
minloss	1.70	50.477	1.59	49.906
G2	1.68	50.255	1.57	49.709
G3	2.24	46.224	2.60	43.745
G4	2.64	44.476	3.06	42.153
G5	3.20	42.836	3.71	40.044
G6	4.26	41.341	4.94	37.685
G7	5.59	39.576	6.48	36.220
G8	7.46	37.231	8.64	34.299
G9	9.94	35.135	11.51	31.952
G10	14.90	32.412	17.26	29.417
G11	19.83	31.566	22.97	28.535
G12	29.73	29.798	34.44	26.738

Table 3 summarised the average and maximum scores for all the devices. The PSNR of the best overall match between a pair of images was recorded there for device N05 as 43.00dB. This lies between the two highlighted rows for device N05 in Table 10.

The PSNR of the best overall match for a standard Phase One backed digital camera was recorded for device B07 as 38.86dB. This lies between the two highlighted rows for device B07 in Table 10.

For device N05 this indicates that a compression ratio of 2.64 produces less variation from an original image than was measured as the best match between a pair of master images as a result of image capture noise. Similarly for device B07 a compression ratio of 3.71 produces less variation than has been measured as the best match between master images.

Visual inspection of the images also confirms that encoded images with less compression than the highlighted amounts have noticeably less variation than the best matching original master files. This forms the subject of the qualitative investigation which is described later.

The minimally lossless images have PSNR values around 50dB. For N05 this is 7.47dB better than the best matched pair of original images, and for B07 this is 11.05dB better. As PSNR is a logarithmic measure this means that the variations introduced by minimally lossless compression are small compared with the variations between these best pairs of original images.

For N05 the root mean square (RMS) variations introduced by minimally lossless compression are 42% of the variations between the most similar original images, and for B07 only 28%.

The information in table 10 is shown in Figure 4 where the two lines characterize the PSNR with increasing compression for the images N05B4 and B07B3. The images for device N05 show a shallower decline than for device B07.

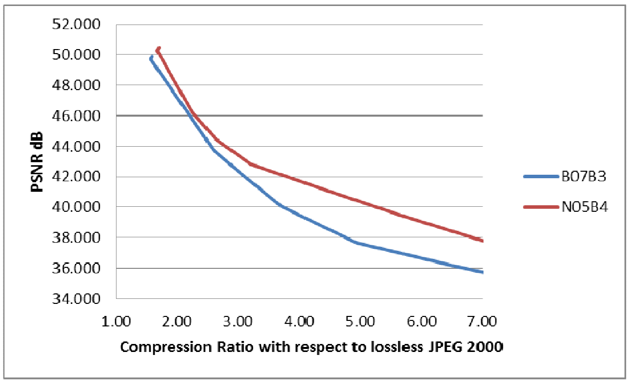

Figure 4: Compression Ratio and PSNR for N05B4 & B07B3 at low compression ratios

The information recorded in table 10 is derived by comparing a compressed file with the original from which it was derived. However, in the set there are a total of five images, and hence there are four alternative master files with which a compressed file can be compared.

So compressed versions of an original image, N05B4, were compared with that original but also with the four alternative master images, N05B1-3, and N05B5. The comparison is in terms of differences as measured by PSNR in dB and the compression ratio with respect to a lossless JPEG 2000 file. The results are shown in table 11.

Table 11: Comparison of compressed versions of image N05B4 with alternative master images for device N05 and Item B

Compression designation	Compression ratio	N05B4 original	N05B1 master	N05B2 master	N05B3 master	N05B5 master
lossless	1	infinity	41.81	42.29	42.83	43.08
minloss	1.70	50.89	41.48	41.98	42.49	42.69
G2	1.68	50.70	41.45	41.96	42.47	42.67
G3	2.24	46.42	40.87	41.31	41.74	41.91
G4	2.64	44.60	40.43	40.83	41.21	41.36
G5	3.20	42.92	39.86	40.20	40.53	40.66
G6	4.26	41.40	39.16	39.45	39.73	39.84
G7	5.59	39.62	38.12	38.35	38.57	38.66
G8	7.46	37.26	36.42	36.57	36.71	36.78
G9	9.94	35.15	34.69	34.79	34.87	34.92
G10	14.90	32.42	32.23	32.28	32.33	32.35
G11	19.83	31.57	31.43	31.47	31.51	31.53

G12	29.73	29.80	29.72	29.75	29.77	29.78

The information from table 11 is also shown in Figure 5. It can be seen that PSNRs of comparisons of compressed versions with the corresponding original start above 50dB and drop fairly rapidly with increasing compression. The best match between master images is between N05B4 and N05B5 at 43dB. Compression by up to a factor of three results in better PSNR than that.

It can be seen from Figure 5 that the PSNRs of comparisons of compressed versions with different master files also drop off with increasing compression, but much more slowly. It had been anticipated there might have been a plateau up to compression by a factor of 3 before this drop off, but it is evident there is an immediate drop off. This can also be seen in comparing the two rows for lossless and minimally lossless compression in table 11. These show a small reduction despite the small changes incurred in using minimally lossless compression.

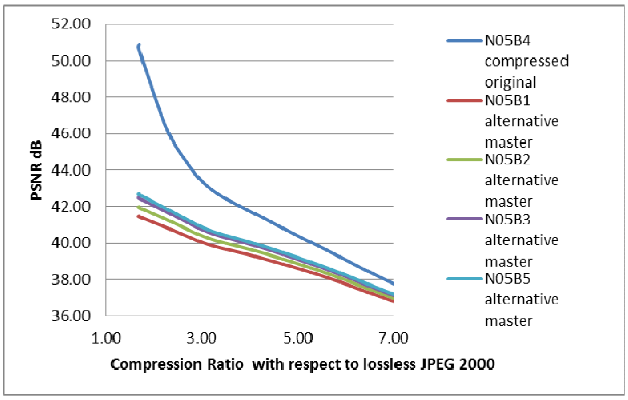

Figure 5: Comparison of compressed versions of image N05B4 with alternative master images for device N05 and Item B

The results in tables 5 and 11 were produced by different tools and there are some minor differences in the results which can be attributed to round-off differences when decoding the images. Kakadu and ImageMagick were used for table 11, whereas Leadtools was used for table 5.

As noted earlier, camera N02 produced consistent sets of five images. The same process was repeated using the five images taken with device N02 and item B. The results are shown in Figure 6.

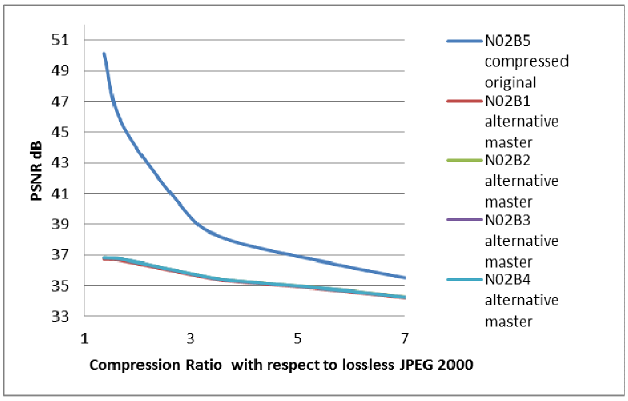

Figure 6: Comparison of compressed versions of image N02B5 with alternative master images for device N02 and Item B

The general patterns of figures 5 and 6 are clearly similar where the four lines for the alternative master files have lower PNSR values, but they are much closer in figure 5.

The comparisons in Figure 6 with individual master images were averaged and these are shown in Figure 7. Also shown are the average PSNR values when compressed versions of image N02B5 are compared with the corresponding compressed versions of the other master files. For example, the G3 compressed version of NO2B5 is compared with each of the G3 compressed versions of the other master files, and their PSNR values are then averaged.

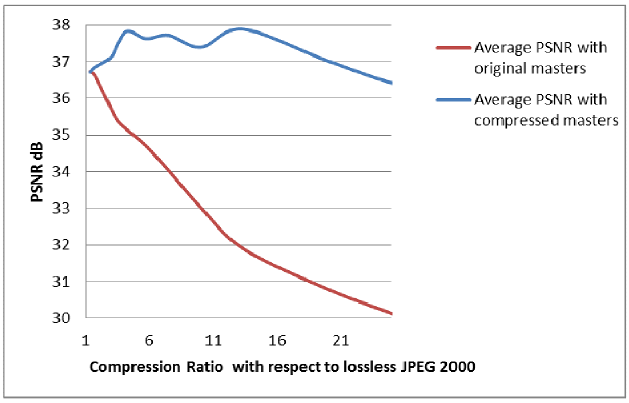

Figure 7: Comparison of compressed versions of image N02B5 with compressed alternative master images and master for device N02 and Item B

Figure 7 shows a remarkable result that the PSNR values rise indicating that the compressed versions are "less different" than the master images from which they were derived. This supports the hypothesis in section 2.2 that predominantly noise is being removed with low levels of compression. With a level of compression above 14 then the signal (or quality) of the image is also being removed. This effect is worthy of further investigation.

4.2 JPEG 2000 Encoding Artefacts

There are several publications which discuss three types of artefacts that can arise when encoding an image using JPEG 2000 at low bit rates and hence with a modest or high degree of compression. These are ringing, colour bleed and tiling artefacts. Ringing and colour bleed can both arise when there is a rapid change such as with a sharp edge or a colour boundary in an image.

(Fang and Sun) [3][4] discuss a technique for reducing the extent of ringing effects that arise from the wavelet compression in JPEG 2000 when encoding at low bit rates. These are visible spurious oscillations or ringing artefacts such as shadows that can occur when there are sharp edges in an image. They show how the technique can be applied and provide examples with levels of compression that result in PSNR values in the range 21-33 dB.

(Nasonov) [11] discusses a method for estimating the extent of ringing in an image and they note that it is as a result of a cut-off of high-frequency information in the encoded image.

(Punchihewa) [12] discusses a technique to evaluate colour bleeding artefacts which result when there is a leakage of colour across distinct colour boundaries in an image.

(Hashimoto et al) [7] discusses techniques for reducing tiling artefacts that can arise in the JPEG 2000 encoding process. Tiles of the image are analysed separately and artefacts can occur at tile boundaries. These can be quite conspicuous especially at low bit rates.

(Qin et al) [13] proposes a post-processing method that can significantly reduce the tiling artefacts in low bit JPEG 2000 images.

These three types of artefact are all described as arising when encoding at low bit rates and comparatively high levels of compression. The levels of compression discussed in the previous section are much smaller than those discussed in the literature. These effects might in principle still be present even at low levels of compression; however they have not been detected in the compressed images produced in this work.

5. QUALITATIVE ASSESSMENT

5.1 Introduction

The preceding sections described a quantitative assessment which identified two best pairs of images for the top of range camera N05 and also from B07, one of the standard production cameras. The PSNR of the former pair was 43.00dB and the PSNR of the latter pair was 38.86dB. The degradation resulting from progressively greater compression was also assessed for both of these best match pairs of images. Within each of these series of compressed images two neighbouring images were identified: one with a PSNR value just greater, and one with a PSNR value just lower than the PSNR between the best match pair of master images.

We now report on a qualitative assessment which used a questionnaire, in which the relationship between the variations between these best match pairs and the neighbouring lossy compressed versions of one of the original master files formed a central part.

Two further types of assessment question were included in the questionnaire, regarding (a) the 'suitability for envisaged use' of a range of compressed images, and (b) a comparison between minimally lossless images and alternative lossless master images.

5.2 Questionnaire Design

There were three groups of questions, with ten questions overall.

The first group comprised four questions to compare the best match pair of original images against compressed versions of one image from each pair, as follows.

For device N05 one image of the best pair, N05B4, was designated as the original, and three alternative images were presented. Two of these were compressed lossy images and the other was the other lossless master file. The three images were the G3 and G4 lossy compressed images derived from N05B4 and master image N05B5.

Responses were sought indicating which image was least different and which image was most different from the original.

A very similar second question used a different small sample from the images from N05B5 and the G4 and G5 compressed versions from N05B4. The combined result of these two questions enabled us to relate the perceived difference between N05B4 and N05B5 to those from three lossy compressed versions of N05B4 – G3, G4 and G5.

Two further questions repeated this process with the other best match pair B07B3 and B07B5, relating the difference between them to those from three lossy compressed versions of B07B3 – namely G4, G5 and G6.

The second group of questions assessed the suitability of a range of compressed images for envisaged use. Image samples were taken from different types of content: a western manuscript, a music manuscript, and an eastern manuscript. The first two samples were presented at normal full resolution and the last at a magnification of 20. A lossless image was designated as the original and six alternative images were presented. One of these was the lossless original and the remaining five were the progressive more compressed lossy images G7 to G11 derived from the original. Responses were sought on whether the images were considered perfect, acceptable, marginal or unacceptable.

The third group of questions compared minimally lossless images with alternative lossless master images. The three questions again sought responses on whether the images were perfect, acceptable, marginal or unacceptable. The first image in this group was at normal full resolution and was based on the best overall set of images taken with device N05 with Item B. A sample taken from N05B4 was designated as the original and six alternatives were offered. These were the lossless original N05B4, a minimally lossless compressed version of N05B4, and lossless samples from N05B1-3 and N05B5.

The remaining questions in this group were both at magnification 60. One of these questions used samples from five alternative images: one was the original and the remaining four were all minimally lossless images derived in four different ways using kakadu. (These arise from whether a 'precise' flag is used during encoding, and separately if the same flag is used in decoding).

The final question was similar to the first question in the group except that this is at magnification 60. As before, a sample taken from N05B4 was designated as the original. There were six alternatives: the lossless original, a minimally lossless compressed version, and the other four were lossless samples from N05B1-3 and N05B5.

5.3 Questionnaire Responses

A survey questionnaire has been conducted comprising the ten questions described above. As responses were not mandatory their number varied. There were between 146 and 175 responses for each of the questions in the first group, and between 128 and 134 responses for each of the questions in the remaining groups. The conclusions are as follows:

The responses from the first group of questions were in line with the quantitative analysis.

Regarding the overall best match pair of images N05B4 and N05B5 the results from the questionnaire show that the difference between these images is comparable with the difference between N05B4 and its G5 lossy compressed version.

As noted in table 10 the PSNR comparing N05B4 and N05B5 is 43.00dB, and this lies between the PSNR values for G4 and G5 compression which are 44.48dB and 42.84dB respectively. These are relatively small differences in PSNR; hence the distinction between the images is not great and this does lead to a spread of responses:

- 90% indicate that the least overall change is from the lossy compressed version G3, and a further 4% with G4.
- 52% indicate that the most overall change is from N05B5, while 42% indicate that it is from G5.

Two questions have partial overlap when comparing only G4 and N05B5: with one question 72% indicated that G4 had least change while 21% indicated that N05B5 had least change, and with the other question 21% indicated that G4 had most change while 74% indicated that N05B5 had most change

These lead to the conclusion that the difference in this best match pair is comparable with the difference between the original and the G5 lossy compressed version.

Regarding the overall best match pair of images B07B3 and B07B5 the results from the questionnaire show that the difference between these images is comparable with the differences between B07B3 and its G5 and G6 lossy compressed versions.

As noted in table 10 the comparison PSNR between B07B3 and B07B5 is 38.86dB, and this lies between the PSNR values for G5 and G6 compression which are 40.044 and 37.685 respectively. The responses were as follows:

- 96% indicate that the least overall change is from the lossy compressed version G4.
- 82% indicate that the most overall change is from G6 and 14% indicate that the most change is from B07B5.

Two questions have partial overlap when comparing only G5 and B07B5: with one question 71% indicate that G5 has less change than B07B5 while 26% indicated B07B5 had least change, and with the other question 64% indicated that B07B5 had most change while 33% indicated that G5 had most change.

These lead to the conclusion that the difference in this best match pair is comparable with the differences between the original and the G5 and G6 lossy compressed versions.

The questions in the second group were of similar structure except for a change in magnification.

At the original magnification compression of G8 or lower is typically considered perfect and G10 is considered acceptable or perfect. These correspond to compression ratios around 6 and 13 respectively.

However, at a magnification of 20 these drop to G2 and G3 respectively, where G2 is considered perfect and G3 is considered acceptable. These correspond to compression ratios around 1.8 and 2.3 respectively.

The questions in the third group investigated two different comparisons:

1. One pair of questions used the same master files, and included an original, a minimally lossless compressed version, and four lossless alternative master files. However, the two questions are at different magnifications.
2. The other pair of questions are both at magnification 60. One question comprises an original and different minimally lossless compressed images. The other question comprises an original, a minimally compressed image and four lossless alternative master images.

Regarding the first comparison:

- At the original magnification 84% considered the original to be perfect while 90% considered the minimally lossless version to be perfect. Between 4% and 5% considered that the alternative master files as perfect, 53% to 66% as acceptable, 23% to 30% as marginal, and 5% to 12% as unacceptable.
- At magnification 60 93% considered the original to be perfect, and 73% considered the minimally lossless version to be perfect. Between 1% and 3% considered that the alternative master files as perfect, 32% to 56% as acceptable, and 41% to 66% as marginal or even unacceptable.

Regarding the second comparison at magnification 60:

- 90% to 92% considered the minimally lossless versions to be perfect.
- Between 2% and 3% considered that an alternative master file was perfect, and the remaining 97% to 98% indicated that these were acceptable (32-56%), marginal (30-47%) or unacceptable (11-19%).

Minimally lossless images were deemed (mostly) to be perfect, whereas alternative master files were deemed to be only acceptable, marginal, or even unacceptable.

6. CONCLUSIONS AND DISCUSSION

6.1 Conclusions

Several conclusions may be drawn from the quantitative and qualitative assessments:

1. Images taken with even with a top of range camera show considerable variability despite all efforts to minimise difference in conditions. When compared with an original image at high magnification, alternative original images were considered merely acceptable, whereas at the same magnification, minimally lossless compressed vesrions of the original image were considered perfect.
2. The RMS variations in a minimally lossless compressed image are of the order of 30% - 40% of the variations between original lossless master images.
3. A minimally lossless file is typically 30-40% smaller than a lossless JPEG 2000 file.

Depending on a use case, which may be related to the type of content, an image may be compressed by a factor of between 3 and 6 compared with a lossless JPEG 2000 file and still be considered perfect at a magnification up to 20.

These conclusions, especially the recognition that there is considerable variability in the original images, question the need for images to be losslessly retained.

6.2 Discussion

The conclusions raise the question about how the value in an image arises. It could be associated with the image itself, perhaps because this image was taken by a famous person on a particular occasion. Or, more often, the value is in the subject of the image, such as a manuscript. Especially in the latter case, the conclusions suggest it might be appropriate to store the image in a slightly lossy compressed manner, especially if there is cost pressure on storage or transmission of the images.

Retaining images in a minimally lossless manner does reduce storage costs but appears to reduce the inherent value by a rather small, indeed we would argue negligible, amount. There may be concern that OCR may work less well and this should be investigated. However, today's OCR tools work with high quality images, and this paper has shown there is considerable variability in these. If OCR is compromised by minimally lossless compression then it would be highly likely that it would work with only a low proportion of quality digitized images. This clearly is not the case.

Depending on the type of content, for example with bulk digitisation, it would seem prudent to apply more compression since a modest amount of compression can be applied without visual degradation of the image.

Some people express a belief that future tools will be developed that will reduce the noise in an image and thereby improve the quality in images that have already been taken. This could be done already if multiple images were taken of the same item, but this is not the standard process in today's cost efficient digitisation studios. There seems no greater reason to believe that the noise in single images of an item could in future be reduced better than the artefacts arising from slightly lossy compression – indeed the latter is arguably slightly more deterministic and therefore easier to tackle. Reliance on future improvements is thus questionable.

In terms of a business case, a baseline can be proposed based on the value and cost of a certain level of compression. An option can also be proposed to provide additional value but at additional cost by applying less compression but requiring more storage. Our results suggest that very little additional value is obtained in moving from minimally lossless to lossless but this would increase storage costs by roughly a half.

This work started with experimentation and over time it has helped establish a process for evaluation of noise in the digitisation process compared with the effects of lossy compression. Manufacturers are continually producing new camera models, so it would seem prudent to repeat these experiments periodically to provide a baseline assessment of the quality and repeatability of cameras as this changes over time.

7. ACKNOWLEDGEMENTS

We are grateful to Kjetil Iversen and his staff at the National Library of Norway and to Andrew Austin and his staff at the British Library for taking multiple images with several cameras and scanners. Ken Tsang developed software which analysed the sets of images.

8. REFERENCES

[1] Belbachir A. N. and Goebel P., "Medical Image Compression: Study of the influence of noise on the JPEG 2000 Compression performance," in 18th International Conference on Pattern Recognition, 2006.

[2] Chen L., Zhang X., Lin J. and Sha D., "Signal-to-noise ratio evaluation of a CCD camera," Optics and Laser Technology, vol. 41, pp. 574-579, 2009.

[3] Fang J. and Sun J., "Efficient Embedded Ringing Artifact Reduction for Low Bit-rate JPEG2000 Images," in Signal Processing, 8th International Conference , 2006.

[4] Fang J. and Sun J., "Ringing Artifact Reduction for JPEG2000 Images," in Proceedings Third International Conference on Intelligent Computing, ICIC 2007, Qingdao, China, 2007.

[5] Faraji H. and MacLean W. J., "CCD Noise Removal in Digital Images," IEEE Transactions Image Processing, pp. 2676-2685, 2006.

[6] Janesick J. R., Scientific Charge-Coupled Devices, Bellingham,WA: SPIE, 2001, pp. 605-719.

[7] Hashimoto M., Matsuo K. and Koike A., "JPEG2000 encoder for reducing tiling artifacts and accelerating the coding process," in International Conference on Image Processing, ICIP Proceedings., 2003.

[8] Kurosawa K., Kuroki K. and Akiba N., "Individual Camera Identification Using Correlation of Fixed Pattern Noise in Image Sensors," Journal Forensic Science, vol. 54, no. 3, pp. 639-641, May 2009.

[9] Liu C., Richard S., Kang S. B., Zitnick C. L. and Freeman W. T., "Automatic Estimation and Removal of Noise from a Single Image," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 2, pp. 299-314, 2008.

[10] McHugh S., "Cambridge in Colour," [Online]. Available: http://www.cambridgeincolour.com/tutorials/image-noise.htm.

[11] Nasonov A. and Krylov A., "Scale-space method of image ringing estimation," in 16th IEEE International Conference Image Processing (ICIP), 2009; Pages: 2793 - 2796, 2009.

[12] Punchihewa A., "Objective evaluation of colour bleeding artefact due to image codecs," in 5th International Conference on Visual Information Engineering, VIE 2008. Pages: 801 - 806 , 2008.

[13] Qin X., Yan X.-L., Yang C.-P. and Ye Y., "Tiling artifact reduction for JPEG2000 image at low bit-rate," in IEEE International Conference on Multimedia and Expo, ICME '04. 2004; Pages: 1419 - 1422 Vol.2 , 2004.