Journal of Imaging Science and Technology® 51(3): 248–258, 2007.

Image Quality Comparison Between JPEG and JPEG2000. I. Psychophysical Investigation

E. Allen, S. Triantaphillidou and R. E. Jacobson
Imaging Technology Research Group, University of Westminster, School of Media, Arts and Design, Harrow Campus, Watford Road, Northwick Park, Harrow HA1 3TP, United Kingdom
E-mail: triants@westminster.ac.uk

Abstract. *The original JPEG compression standard is efficient at low to medium levels of compression with relatively low levels of loss in visual image quality and has found widespread use in the imaging industry. Excessive compression using JPEG however, results in well-known artifacts such as "blocking" and "ringing," and the variation in image quality as a result of differing scene content is well documented. JPEG 2000 has been developed to improve on JPEG in terms of functionality and image quality at lower bit rates. One of the more fundamental changes is the use of a discrete wavelet transform instead of a discrete cosine transform, which provides several advantages both in terms of the way in which the image is encoded and overall image quality. This study involves a comparison of subjective image quality between JPEG and JPEG 2000 to establish whether JPEG 2000 does indeed demonstrate significant improvements in visual quality. A particular focus of this work is the inherent scene dependency of the two algorithms and their influence on subjective image quality results. Further work on the characterization of scene content is carried out in a connected study [S. Triantaphillidou, E. Allen, and R. E. Jacobson, "Image quality comparison between JPEG and JPEG2000. II. Scene dependency, scene analysis, and classification", J. Imaging Sci. Technol. 51, 259 (2007)]. © 2007 Society for Imaging Science and Technology.*
[DOI: 10.2352/J.ImagingSci.Technol.(2007)51:3(248)]

INTRODUCTION

The JPEG compression scheme was developed as a standard for the compression of continuous-tone still digital images and, since its adoption in 1992, has become the most widely used method for the lossy compression of digital images.[1,2] JPEG is based upon the discrete cosine transform (DCT) of 8×8 pixel blocks or subimages, followed by quantization and encoding, which leads to certain characteristic artifacts at high compression rates.

JPEG 2000 has a similar structure in terms of stages in the compression algorithm to that of JPEG, but is based upon the discrete wavelet transform (DWT).[3,4] The wavelet transform has been researched for many years as an alternative for signal decomposition.[4–9] JPEG 2000 Part 1 was standardized in 2001,[10,11] with a number of features to improve flexibility in both compression options and applications in imaging.[12] It was also designed to produce superior rate distortion at low bit rates to that of existing standards.

A fundamental advantage of the DWT is that it can be applied using a filter bank, which simplifies the transform process.[4,5] The use of a pyramidal filter bank inherently provides the ability to encode the image at different scales. Additionally, the use of wavelets in compression minimizes the blocking artifacts inherent in schemes such as JPEG. However, there are certain other associated distortions. Both schemes are designed to be perceptually lossless at lower compression rates.

Because there is a level of error introduced by lossy compression, it is necessary to evaluate output image quality as part of an overall evaluation of the algorithm. Image quality may be defined as "The integrated set of perceptions of the overall degree of the excellence of the image."[13] As the definition implies, a true evaluation of image quality should produce results that correlate with the subjective impression of quality.

There are several methods of objective evaluation of distortion in lossy compression, such as mean-squared error (MSE) and peak signal-to-noise ratio (PSNR).[14,15]. A number of evaluations of JPEG 2000 have been performed using such measures.[4,11,12,16–18] These techniques are simple numerical measures and they do not always correlate well with perceived distortion.[19,20]

Methods that include the subjective perception of distortion correlate better with overall compression quality. These may be objective measures that incorporate some model of the human visual system[21,22] or can alternatively be based upon psychometric scaling techniques.

The differing architectures of the two compression algorithms lead to different types of errors. Both algorithms are *scene dependent*, meaning that they perform better on certain types of scenes than on others, producing higher compression ratios with less visible loss. In addition, the types of artifacts are more visible in some image areas than in others.[23] This work aims to compare the two compression schemes across a set of images of varying scene content to ensure that the effects of scene dependence are taken into account.

JPEG VERSUS JPEG 2000: COMPARISON OF ARTIFACTS

To gain an understanding of the issues affecting the image quality of the two algorithms, it is useful to compare their operations and the factors in their design that result in artifacts of a particular type.

Received Aug. 15, 2006; accepted for publication Jan. 13, 2007.
1062-3701/2007/51(3)/248/11/$20.00.

Table I. Comparison of compression stages in baseline JPEG and JPEG 2000.

Compression Stage	Baseline JPEG	JPEG 2000
Preprocessing	• Conversion of RGB image to YCbCr and down-sampling of chroma channels.	• Image "tiling" (OPTIONAL) division of image into nonoverlapping image tiles (varying sizes).
	• Division of image into 8 × 8 pixel subimages.	• Reversible or irreversible color transformation (OPTIONAL).
Frequency Transformation	• Discrete cosine transform, resulting in 64 coefficients representing magnitudes of different frequencies for each sub-block.	• Reversible or irreversible discrete wavelet transform (for lossless or lossy compression, respectively). The image or image tile is decomposed into a number of "sub–bands." Each sub-band consists of coefficients describing horizontal and vertical frequency components at a particular resolution.
Quantization	• Coefficients are reordered using a zig-zag sequence though *each block*. Frequency coefficients from each block are quantized using visually weighted quantization tables, resulting in the highest frequency components and lowest magnitude components being removed. Results in a string of zero magnitudes for the highest frequency coefficients at the end of each block.	• *Sub-bands* of coefficients are quantized separately using a uniform scalar quantizer with the option of different quantizer step sizes for different sub-bands, based upon the dynamic range of the sub-band. Quantization step size will be 1 if lossless compression is required.
Entropy Coding	• Differential pulse code modulation (DPCM) of dc coefficients of *all blocks*.	• Sub-bands are divided into *precincts* and *code blocks*.
	• Huffman or arithmetic coding of *each block* of ac coefficients (those left after truncation).	• Each code block is input independently in raster order into the entropy coder.
	• Run-length coding of remaining string of zero magnitudes.	• Code blocks are coded by individual bit plane, using three passes on an arithmetic coder.

JPEG is a transform based compression method, meaning that a frequency transformation is applied to the image, to decorrelate the data before quantization and encoding. Baseline JPEG allows lossy compression rates of up to 100:1 (although achievable compression rate is very scene dependent). Perceptibity of information loss is minimized by the use of optimized quantization tables.

JPEG 2000 comprises a similar workflow to that of JPEG. The key differences in the operation of the two algorithms at each stage are summarized in Table I. Providing all the required features and functionalities has resulted in architecture more complex than that of baseline JPEG.

A key difference between the two encoders is the transform stage. Although both decompose the image into frequency coefficients, the arrangement of the coefficients is different. The output from the DCT stage in JPEG is an array consisting of blocks of 64 coefficients spatially arranged so that they relate to the magnitudes of frequencies in the *same spatial region* in the original image. At higher levels of compression this can result in *blocking artifacts*,[24] which arise as a result of coarse quantization in individual blocks of pixels and may be seen as one of the main causes of data loss and unrecoverable distortions in JPEG.[25] (See Figure 1.)

It is commonly assumed that the improvements so far demonstrated in quality comparison trials of JPEG 2000 against JPEG are due to the use of the wavelet transform instead of the DCT. However, Steingrímsson[26] suggests that

Figure 1. Test image "*ISO table*" displays clear blocking artifacts at a JPEG compression rate of 90:1.

Figure 2. Test image "*motorace*" at compression rate 80:1 displays severe smoothing artifacts from the JPEG 2000 algorithm.

Figure 3. Ringing artifacts around edges in test image "Chinatown" compressed at 60:1 (JPEG compression).

it is not the choice of the transform but the differences in the way in which the image is subdivided in the stages before entropy coding that might be the key to quality improvements.

JPEG 2000 images do not suffer from blocking artifacts unless the image has been tiled. The decomposition of an image into sub-bands using a wavelet transform results in a lower-resolution version of the original and high-frequency information in horizontal, vertical, and diagonal directions.[4] The sub-bands are quantized separately and further subdivided into *code blocks* before entropy coding. A code block is a rectangular section of a sub-band and a precinct consists of three groups of four code blocks, each group from the same position in each high-frequency sub-band at that decomposition level.[27] The input into the entropy coder is then the code blocks by bit plane, from a precinct scanned in raster order.

Where the values in a reconstructed block within a JPEG compressed image will be dependent upon quantization table and the frequencies present at that spatial location, the reconstruction from the JPEG 2000 image will be made up of code blocks from precincts in the same spatial position relative to the edge of a particular sub-band, from all the different sub-bands. Because the quantization step size is different in different sub-bands the errors will build up in a very different way, being much less uniform over a spatial location by comparison with JPEG blocks. "Smoothing" or "smudging" artifacts appear at higher levels of compression. These appear as a blurring of small regions within the image as shown in Figure 2.

The lossy versions of both compression schemes suffer from *ringing* artifacts. These artifacts are a result of abrupt truncation of high-frequency coefficients which affects the appearance of edges in particular and is evident as oscilla-

	S – MATRIX for image '*Lena*'								
	TIFF	JPEG20	JPEG40	JPEG60	JPEG80	JP2K20	JP2K40	JP2K60	JP2K80
TIFF	0.00	0.00	-1.64	-1.64	-1.64	0.52	-0.52	-1.64	-1.64
JPEG 20	0.00	0.00	-1.28	-1.64	-1.64	1.28	-0.25	-1.28	-1.64
JPEG40	1.64	1.28	0.00	-1.64	-1.64	1.64	1.64	0.00	-1.28
JPEG60	1.64	1.64	1.64	0.00	-1.64	1.64	1.64	1.64	0.52
JPEG80	1.64	1.64	1.64	1.64	0.00	1.64	1.64	1.64	1.64
JP2K20	-0.52	-1.28	-1.64	-1.64	-1.64	0.00	-0.25	-1.64	-1.64
JP2K40	0.52	0.25	-1.64	-1.64	-1.64	0.25	0.00	-1.64	-1.64
JP2K60	1.64	1.28	0.00	-1.64	-1.64	1.64	1.64	0.00	-1.28
JP2K80	1.64	1.64	1.28	-0.52	-1.64	1.64	1.64	1.28	0.00
Standard deviation	0.91	1.03	1.44	1.16	0.55	0.69	1.01	1.44	1.21
Mean Scale Difference	0.91	0.72	-0.18	-0.97	-1.46	1.14	0.80	-0.18	-0.77
Scale Value	**8.22**	**6.47**	**-1.64**	**-8.75**	**-13.16**	**10.28**	**7.19**	**-1.64**	**-6.97**

Figure 4. Scale value differences for image "*Lena*."

Table II. Peak-signal-to-noise-ratio values between original uncompressed images and compressed versions for the set of test scenes used in this investigation.

		PSNR (db)			
Compression Ratio		20:1	40:1	60:1	80:1
AFRICAN TREE	JPEG	44.0	42.3	41.4	40.6
	JPEG 2000	45.5	42.2	41	40.2
BIKE	JPEG	28.1	25.4	23.9	22.8
	JPEG 2000	30.1	26.5	24.6	23.6
BOATS	JPEG	31.8	28.4	26.9	25.7
	JPEG 2000	34.2	30	28	26.6
ISO CAFETERIA	JPEG	23.8	21.2	19.8	18.8
	JPEG 2000	25.4	21.9	20.4	19.5
CHINATOWN	JPEG	32.5	29.1	27.3	26
	JPEG 2000	35	30.5	28.2	26.6
FORMULA	JPEG	32.6	28.5	26.8	25.6
	JPEG 2000	36.7	31.1	28.3	26.6
ISO FRUITS	JPEG	32	29.2	27.7	26.5
	JPEG 2000	34.8	30.8	28.8	27.3
GLASSES	JPEG	36.5	32.9	30.6	29.1
	JPEG 2000	38.7	35.1	32.8	31.3
KIDS	JPEG	34.9	30.9	28.8	27.1
	JPEG 2000	38.3	33.3	30.8	28.8
LENA	JPEG	38.7	34.3	32.1	30.2
	JPEG 2000	41.7	37	34.4	32.3
LOUVRE	JPEG	32.2	29.4	27.7	26.8
	JPEG 2000	34.7	30.6	28.4	27.4
MOTORACE	JPEG	25.3	22.5	21	19.8
	JPEG 2000	27.5	23.6	21.8	20.6
SAULES	JPEG	25.4	23.4	22.4	21.9
	JPEG 2000	26.8	24	22.8	22.1
ISO TABLE	JPEG	32.3	28.3	26.2	24.9
	JPEG 2000	35.5	30.2	27.5	25.6
LEOPARD	JPEG	32	28.8	27.3	26.4
	JPEG 2000	34.8	30.8	28.9	27.8
YELLOW FLOWERS	JPEG	30.6	27.6	26.3	25.2
	JPEG 2000	34.4	30.1	28.1	26.5

tions or "ripples" around high-contrast edges,[24] as shown in Figure 3. The visual effects of ringing in JPEG 2000 are reduced because of the arrangement of sub-bands, meaning that they are less localised and the errors are distributed across the image. They tend therefore to be less noticeable than the smoothing artifact illustrated in Fig. 2.

In JPEG, however, because the errors from a block in a specific spatial area will affect the same area in the reconstructed image, ringing is much more visible. This can be identified as one of the key reasons that JPEG compresses text poorly. A side effect of ringing is that it can cause the image to appear sharpened. The density oscillations around an edge can appear similar to the slight overshoot or undershoot of density on either side of an edge as a result of the use of a sharpening filter.

Both algorithms suffer from color artifacts. These are caused by various factors, including the subsampling of chroma channels in the JPEG algorithm and the irreversible color transformation in JPEG 2000, as well as reconstruction errors from quantization. The visual effect of these errors is "color bleeding" which affects smoothly graduating areas.

ARTIFACTS AND SCENE DEPENDENCY

The level and visibility of the artifacts is dependent upon the level of compression predominantly; more compression will result in coarser quantization of frequency coefficients and therefore greater reconstruction errors. However, because the quantization is performed in frequency space, it is also dependent upon the frequencies contained within each image or image block. Therefore, the amount of error will also be dependent upon the frequency characteristics of the image, meaning that stimuli with different scene characteristics will have different levels of image quality at the same compression rate.

Triantaphillidou et al.[23] describe in detail different types of scene dependency affecting results in psychophysical studies of image quality. The difference in errors as a result of variations in scene content for lossy compression schemes may be classified as *scene dependency of the algorithm*. It is possible to provide some quantification of this type of scene dependency using simple error measures. This type of scene dependency is a result of the performance of the algorithm on the characteristics of a particular image.

However, there will be other types of scene dependency affecting the results of an image quality study as a result of the observer being part of the imaging chain. The first is *scene dependency resulting from an observer's quality criteria*, which may be viewed as the way in which the type of scene content (for example portrait, landscape) affects the observer's image quality judgements. Perhaps more important is *scene dependency due to a visibility of an artifact in some image areas compared to other areas*. Clearly, this depends upon several factors: the visual weight given to particular image attributes by the human visual system, the type of artifact and the content of the scene.

In summary, the susceptibility of a scene to a particular artifact will clearly influence the results of an image quality

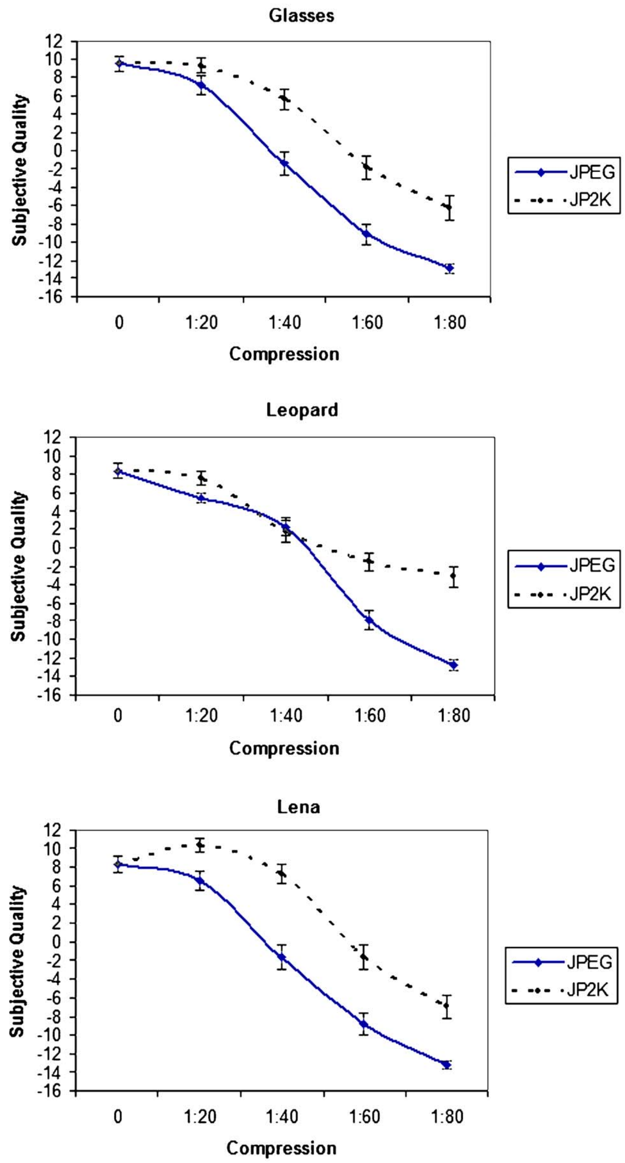

Figure 5. Results Group 1.

study, in both objective and subjective assessments. It might be hypothesized that in a comparison of the image quality of JPEG and JPEG 2000, scenes with large areas of smoothly graduating tone will show more obvious blocking artifacts at high levels of compression, therefore such scenes might be expected to produce poorer results for JPEG. Meanwhile scenes containing many edges and straight lines, for example those typical in architectural images might suffer more visual degradation from the smoothing artifacts of JPEG 2000. These images may also suffer ringing artifacts.

SUBJECTIVE IMAGE QUALITY ASSESSMENT

This investigation aims to provide a comparison of the subjective image quality of JPEG versus JPEG 2000 in relation to scene content.

Thurstone's Law of Comparative judgement[28] assumes that the discriminal process (the process by which observers make judgements of samples) is a random variable with a probability density function following a Gaussian or normal distribution on the perceptual attribute scale (or "ness" according to Engeldrum[29]); in this case the image quality scale. Thurstone postulated that the proportion of times that a stimulus is judged greater than another stimulus may be viewed as an indirect measure of the distance between the two stimuli on the "ness" scale being evaluated. Normalizing the difference between two mean scale values by dividing by the standard deviation of the probability density function describing the values produces results in terms of z values. The z scores may then be used to generate an interval scale of image quality.[29]

Interval scales provide numerical values for a perceptual attribute against the physical properties of the image, in this case, compression rate. Distances between values on the scale are proportional to distances in perceived image quality, allowing predictions of differences between samples.[29] They may be compared in terms of relative magnitudes of differences.

A psychophysical experiment was performed using a paired comparison in which observers were asked to select an image from a pair displayed on screen, based upon their preferred image quality. Ten observers, six male and four female, with some experience of visual assessment of images carried out the tests. All had normal color vision. Each uncompressed image was compressed to the same range of compression rates using both algorithms. All of the compressed images for a particular scene were then compared with all the others from the same dataset. The dataset also included the original uncompressed TIFF version of the scene. The total number of unique pairs was 36 per scene.

TEST IMAGES

Sixteen original images were used in the investigation. Twelve were selected from a Kodak Photo-CD collection, three from the ISO 12640:1997 standard image set and the final one was "*Lena*," an image commonly used in compression quality investigations. All images are included in Appendix 1. The images were selected to cover a range of image content and characteristics.

The data set included:

- A range of different scenes, such as portraits, natural scenes, architectural.
- Scenes containing smoothly graduating tones, in which blocking distortions might be highly visual at higher compression rates.
- Scenes containing text, which might be susceptible to the ringing distortions inherent in JPEG.
- Highly chromatic scenes and some with a very low chromatic content.
- Scenes containing a large amount of fine detail, which might be particularly susceptible to JPEG 2000 smoothing artifacts.

The majority of images were color, although two gray scale images were also used, including "*Lena*." The Kodak Photo-CD images were opened at a resolution of 512 by 786

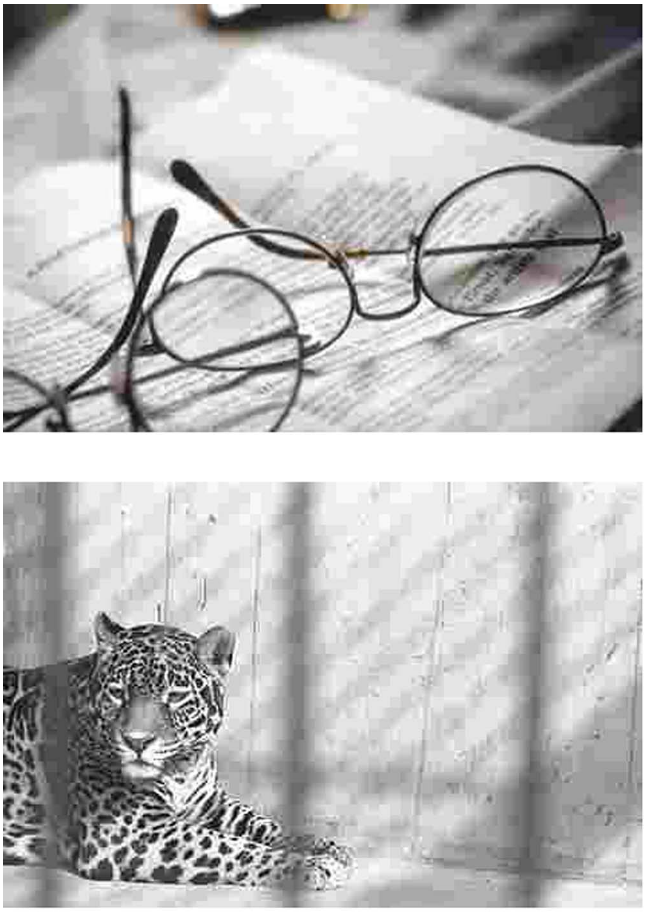

Figure 6. Images "*glasses*" and "*leopard*" show severe blocking and ringing artifacts at JPEG compression ratio 80:1.

at 72 dpi in CIELAB 16 bit per channel color space. They were converted to sRGB color space, downsampled to 317 by 476 pixels and finally saved as TIFFs. The original images were all the same size, approximately 445 Kb, to be displayed at 100% resolution. The selected size allowed two images to be displayed side-by-side on screen without any effects from further interpolation by the graphics card for the display.

IMAGE COMPRESSION

It was necessary to set a maximum compression rate based upon JPEG rather than JPEG 2000 because of the more limited compression capabilities of JPEG. After initial tests the images were compressed at intervals: 80:1, 60:1, 40:1, and 20:1. This set of compression rates was selected to cover a range which might conceivably be used in everyday imaging across a range of applications, particularly consumer imaging applications and the web. For each scene there were therefore four compressed images for each algorithm and one uncompressed image. These gave a total of 36 unique pairs per scene and 576 comparisons in total.

The JPEG compressed images were processed using Advanced JPEG Compressor v4.1, a stand-alone software by Winsoftmagic Development.[30] The software compresses using baseline JPEG standard compression, while allowing specification of output file size, quality setting or compression rate. The JPEG 2000 images were compressed using Lurawave SmartCompress 3.0, developed by Algo Vision Luratech GmbH.[31] Default settings were used in both methods of image compression.

PSYCHOPHYSICAL DISPLAY

The images were displayed on a 15 in. NEC Multisync M500 monitor, with a Matrox Graphics MGA Millenium graphics card adapter at screen resolution of 1024×768 pixels. To ensure correct color rendition, the monitor was characterized,[15] before being calibrated to the sRGB standard.[32–34] The viewing environment was also set to sRGB reference conditions.

The paired comparison software was written in Visual Basic™ 6 and run on an IBM compatible HP Vectra VA platform. Images were displayed side-by-side in the center of the screen in a random sequence.

RESULTS AND INTERVAL SCALE GENERATION

According to Thurstone, the relationship between the z values and scale values for samples A and B is defined by[29]

$$S_A - S_B = z_{A-B}\sqrt{\sigma_A^2 + \sigma_B^2 - 2\rho\sigma_A\sigma_B}, \qquad (1)$$

where $S_A - S_B$ is the difference between scale values, z_{A-B} is the z value produced, σ_A and σ_A are the standard deviations of the observers' responses for the two samples, and ρ is the correlation between the two samples. In the case V solution to this expression, it is assumed that the variances are equal and that there is zero correlation between samples, which simplifies the expression to[29]

$$S_A - S_B = z_{A-B}\sigma\sqrt{2}. \qquad (2)$$

To prevent inaccurate scale values, Engeldrum suggests $p=1$ is substituted by $1\text{-}1/(2n)$ and $p=0$ by $1/(2n)$, where n is the number of observers. An example of the scale value differences and scale values for image "*Lena*" is shown in Figure 4.

ERROR MEASURES

Additional to the subjective investigation, values for peak-signal-to-noise ratios (PSNR) were calculated between each original image and all compressed versions. PSNR is defined as

$$\text{PSNR} = 20 \log\left[\frac{2^k - 1}{\text{RMSE}}\right], \qquad (3)$$

where $k=$number of bits per pixel and RMSE is root mean square error. The results for all images are shown in Table II and average results for most scenes in Figure 15.

DISCUSSION: SUBJECTIVE RESULTS

The results are presented as plots of interval scale values against compression ratio. Scenes producing similar trends in subjective image quality have been grouped for clarity. Error bars indicate ±1 standard deviation in distribution of responses from observers.

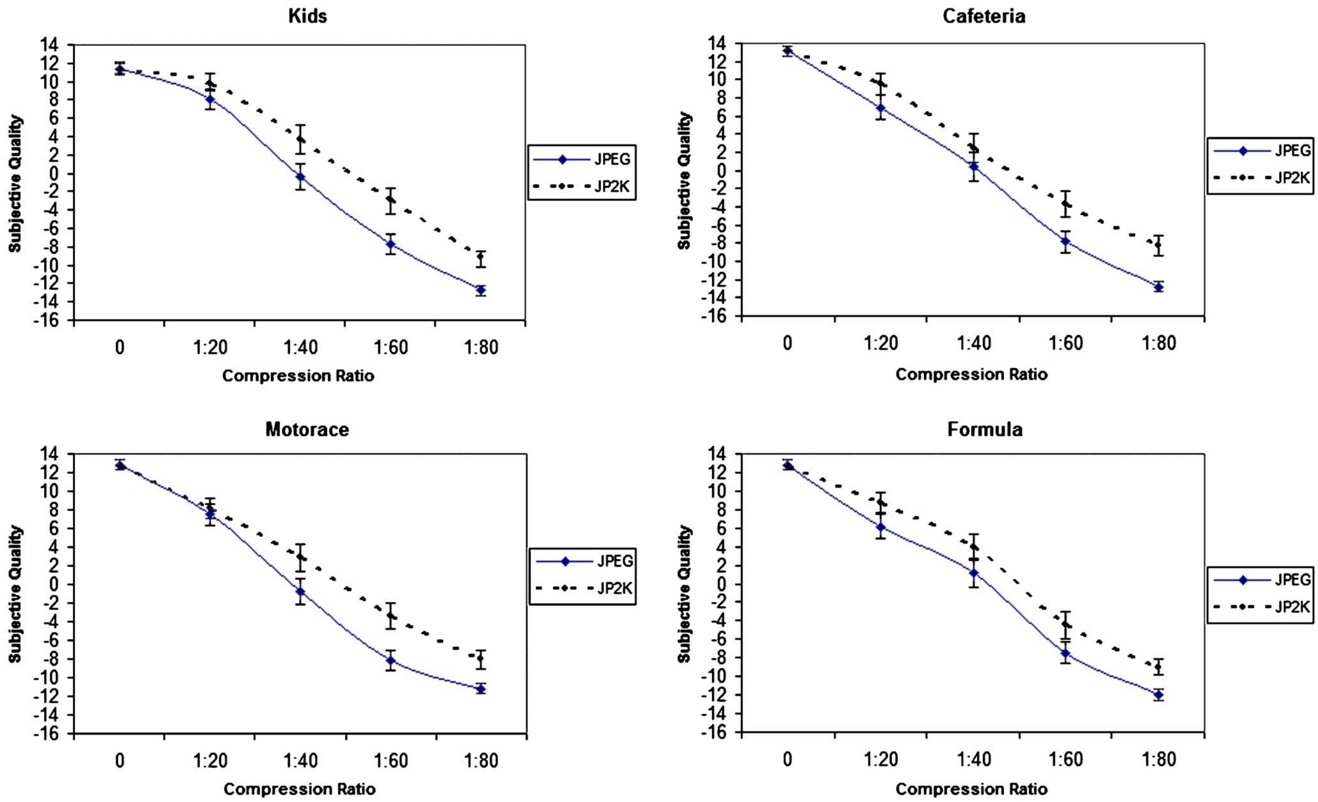

Figure 7. Results Group 2.

a)

b)

Figure 8. Comparison of artifacts at the highest compression rates on the Formula image: (a) JPEG 80:1; (b) JPEG 2000 80:1.

The first two groups show clear quality improvement for JPEG 2000 over JPEG across most of the range, with significantly better performance at higher compression ratios.

In group 1 (Figure 5), "*Lena*," "*Glasses*" and "*Leopard*" show the best subjective quality for JPEG 2000 compared to JPEG across the majority of the compression range. All demonstrate little perceptible quality loss until a compression ratio of 40:1 for JPEG 2000 compression, and from 40:1 show only a gradual loss in quality. The results for the JPEG compression of these images however produce steeper curves, showing more quality loss, the scale values at the highest compressions around −13. The difference between the results from the two algorithms is more marked at the bottom of the range, indicating that JPEG 2000 performs better at higher compression rates. Image "*Leopard*" shows anomalies at a compression ratio of 40:1; at this point only indicating that JPEG is slightly preferred, however the error bars indicate that the distribution of responses is large and therefore the preference might alter with a larger number of observers; the remainder of the range again indicates a preference for JPEG 2000.

These scenes have certain common characteristics. Images "*Lena*" and "*Leopard*" are the two gray scale images from the data set and "*Glasses*," although an RGB image, is very low in chroma. The lack of chromatic information in the images means that they will not suffer from color artifacts; however this does not explain the reason for such a difference in results, as both suffer from these artifacts. An

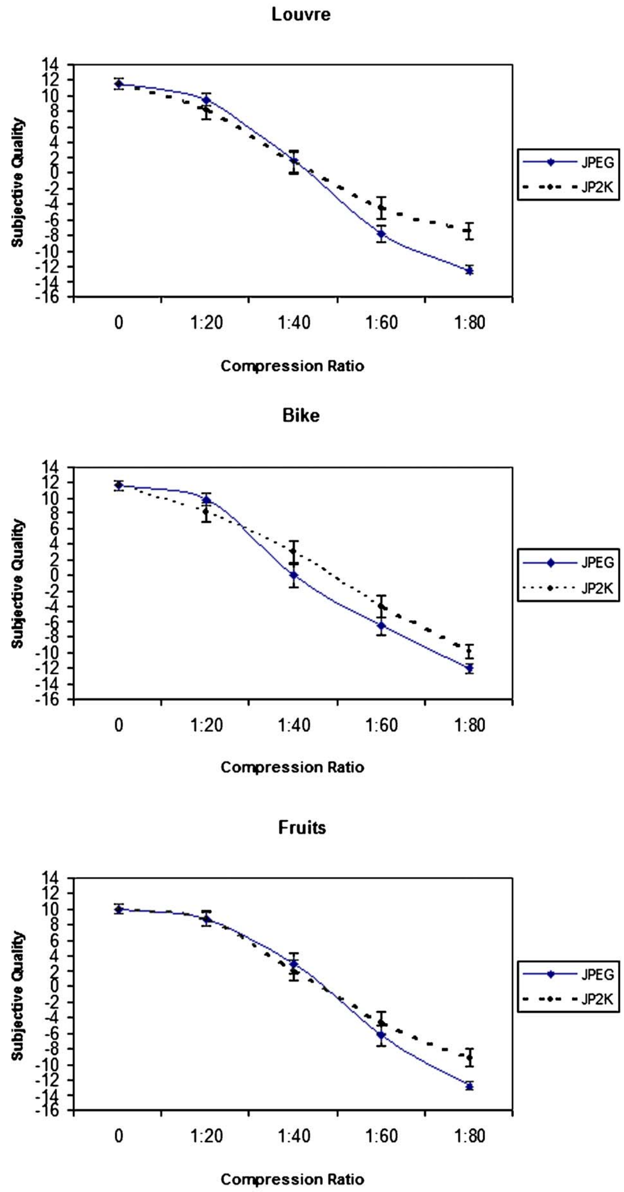

Figure 9. Group 3.

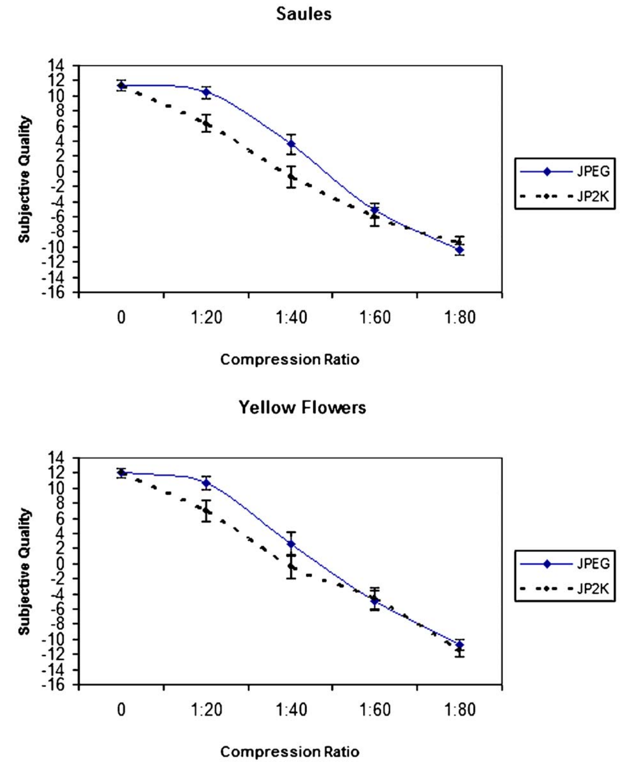

Figure 10. Group 4.

explanation might be the large areas of smoothly graduating tone in all three images and also, in "*Glasses*" and "*Leopard*" fine detail and high-contrast edges. Figure 6 shows these two images at the highest JPEG compression rate and demonstrates that blocking is highly visible in areas of low frequencies, while ringing is problematic around the edges.

Of interest within this group is the "*Lena*" image. This image was preferred to the uncompressed image at a JPEG 2000 compression of 20:1. This is not surprising, when considered in the context of scene dependence. Biederman[35] observed that portraits were often preferred in terms of quality when slightly blurred, and the artifacts of JPEG 2000 blur the image, therefore this result confirms the preference.

The images in group 2 (Figure 7), "*Kids*," "*Formula*," "*Motorace*," and "*ISO Cafeteria*" show improved perceived quality for JPEG 2000 compared to JPEG across the whole range of compression ratios, however the improvement is not as marked as it is in group 1, with quality decreasing at a relatively constant rate for both algorithms. The range across the quality scale is large for both algorithms, indicating a greater loss in image quality at high compression ratios than observed in the first group. These images all have a high chromatic range and contain key areas of fine detail, where loss of high frequencies is more noticeable. All four images also contain text or numerical data. The high frequencies within the images may account for the large loss in quality, as both algorithms remove these frequencies and therefore the blocking and smoothing artifacts are highly apparent. In addition, both algorithms produce ringing. Fundamentally, these scenes will compress poorly in any transform based lossy compression scheme. The difference between the results from JPEG and JPEG 2000 indicates that blocking artifacts are more bothersome than smoothing artifacts in these scenes. This is illustrated in Figure 8, which shows the results for the "*Formula*" image at a compression ratio of 80:1. It is clear from this image that the ringing artifact is more evident in the JPEG version. As text and numerical data tends to become a focal point, artifacts in these areas are more noticeable and this may be another reason for the difference.

The next few groups of scenes produce more ambiguous results, shown in Figures. 9 and 10. The curves shown in Fig. 9 show similar results, in that the subjective quality is very similar for both algorithms at low compression rates, with a slight preference for JPEG over JPEG 2000, but the reverse is true at higher compression rates. Scenes "*ISO Fruit*" and "*Louvre*" have a slightly smaller quality range for JPEG 2000 compression than JPEG than the scene "*Bike*,"

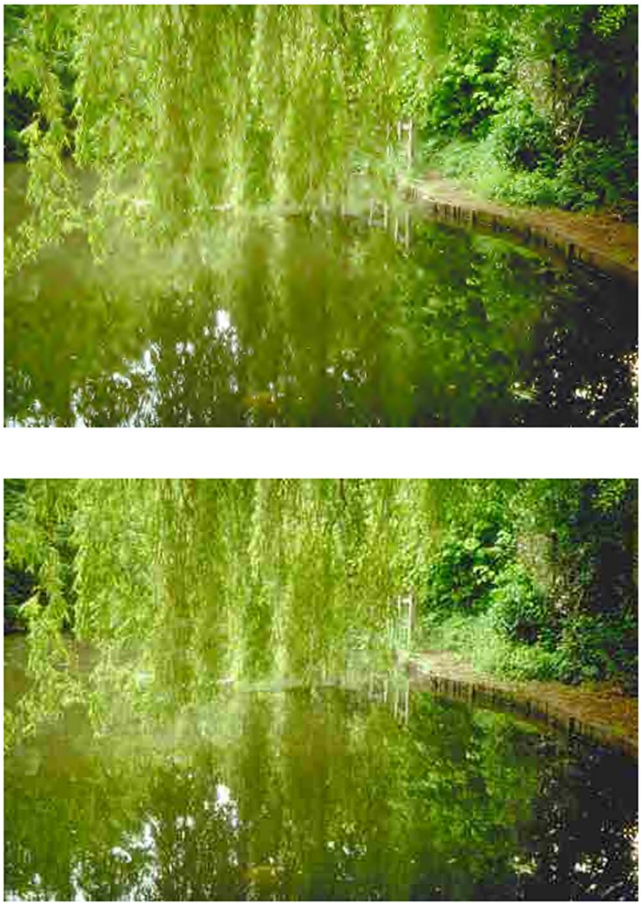

Figure 11. Compression ratio 40:1. Smoothing is clearly evident in the JPEG 2000 image (top); however, the fine detail within the image appears to mask blocking artifacts produced by JPEG (bottom).

meaning that there is less quality loss at higher JPEG 2000 compression rates. The results are very close for both algorithms, but the slight preference for JPEG at low compression suggests that the sharpening effect from JPEG ringing artifacts might be preferable to the blurring caused by JPEG 2000. As compression is increased the perceived quality decreases and JPEG 2000 is preferred. At high compression rates blocking becomes more visible and ringing more severe. These scenes contain both flat areas and high frequencies; therefore the effects of either might be less preferable or more noticeable than the smoothing of JPEG 2000.

The results for group 4 are unexpected. In both of these scenes, JPEG produces much better subjective quality than JPEG 2000 across most or all of the compression range. The curves (Fig. 10) are extremely similar. The quality range for both compression algorithms is almost identical; however at compression ratios from 1:20 to 1:40, JPEG demonstrates improved quality over JPEG 2000. Both of these images contain large areas of fine detail. This detailed information is of one predominant color in both scenes. Blocking, ringing and smoothing artifacts are present in the images produced by both algorithms; however the level of ringing is similar and may be discounted. Because there is so much fine detail within the scenes, the smoothing artifact is highly visible, however the blocking artifact is somewhat masked (Figure 11) and this may be the reason for the improved results from JPEG.

The images in group 5 (Figure 12) produce extremely similar curves for both algorithms, although in "*Chinatown*" it appears that there is a slight preference for JPEG 2000, whereas JPEG seems to be preferred for the "*Boats*" image. The "*ISO table*" scene produces results that vary in preference for one algorithm or another across the scene.

The most unusual results are produced from the "*African tree*" image (Figure 13). In this scene, the JPEG images at compression ratios 1:20 and 1:40 have higher quality than both the uncompressed original and JPEG 2000 versions. For the rest of the range, JPEG is preferred to JPEG 2000 and there is little quality loss from JPEG. This is the only image that has a positive quality scale value at JPEG compression of 80:1. Examining the scene characteristics, it is clear that this image is quite different to the other scenes, having low chroma, low contrast and virtually no fine detail. Significantly, the scene is an image of a tree in mist, and therefore contains soft edges. The blurring artifacts produced by JPEG 2000 therefore represent a loss in image quality, whereas the slight sharpening produced by JPEG might be viewed as an improvement.

Figure 14 shows the average across most scenes. The values for "*African tree*" have not been included, as they are so unusual compared to the rest of the images and cause a large increase in the standard deviation of the distribution. From these curves it is quite clear that JPEG 2000 outperforms JPEG across most of the range, with much more significant differences at high compression ratios. At lower compression ratios, the large standard deviations indicate a large spread of results and there seems much less of a performance advantage using JPEG 2000. JPEG was originally developed to be visually lossless at low compression rates and this perhaps indicates that both algorithms perform well at these lower levels of compression.

PSNR provides a measure of the absolute error within an image compared to the original. Because the images are all of a standard size, the values between different scenes are comparable. Interval scales provide a measure of image quality loss, across a range of compressed images compared to the original, but do not provide information about the relative perceived quality of different scenes, as their zero point is not fixed and absolute. For this reason, error measures can be a useful method for quantifying the effects of an algorithm across different scenes and may predict the types of scenes that will produce fewer artifacts when compressed. The problem with such a simple approach, however, is that the values give no prediction of the perceptibility of such errors.

The results for the two algorithms, shown in Table II, indicate higher PSNR, which is often associated with better image quality, for JPEG 2000 compared to baseline JPEG across all scenes at all compression rates, apart from the two highest compression rates for the "*African tree*" image. This confirms the results from previous similar

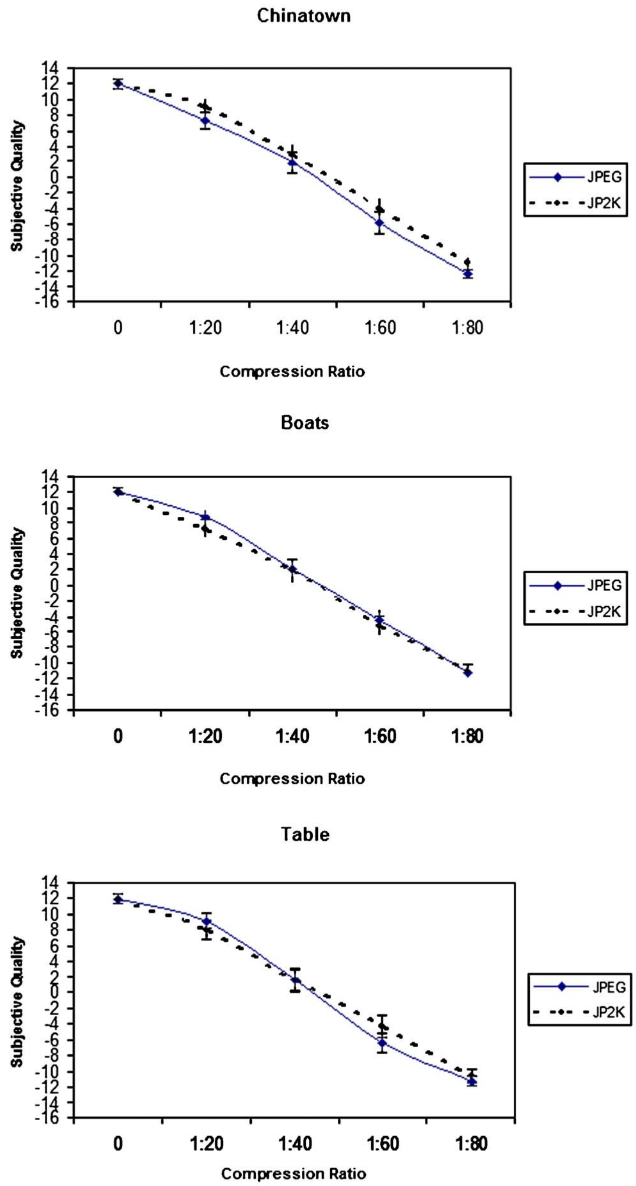

Figure 12. Group 5.

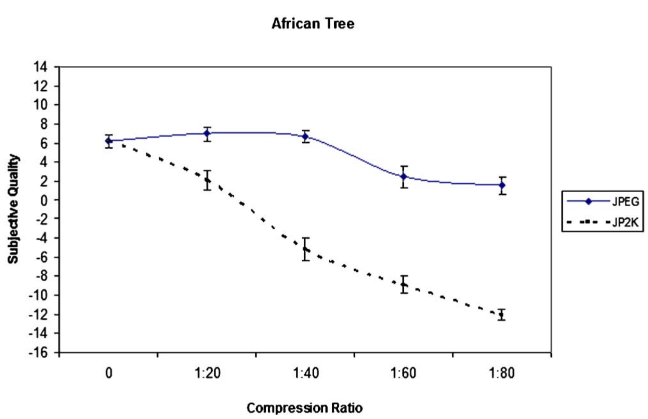

Figure 13. The "*African tree*" scene produces the most anomalous results, with JPEG being preferred to both the original and JPEG 2000 at low compression rates.

investigations[4,11,12,16–18] indicating that JPEG 2000 has better error resilience than JPEG. Figure 15 shows the average PSNR results across all scenes except "*African tree*," which was again removed due to results anomalous with the remaining images. The average results for both subjective and objective evaluations confirm that JPEG 2000 outperforms JPEG; however, PSNR does not predict the scene dependency influencing the perceptual results. This confirms the assertion that PSNR and associated error measures are limited in their value as a tool in image quality studies.

Finally, PSNR results do not correlate with the conclusion from the subjective investigation that in some scenes, JPEG results are preferred to either JPEG 2000 or to the original. Error measures may be considered to in some way quantify the scene dependency of the algorithm, but as there are a number of other influencing factors, they cannot predict the perceived image quality results.[15,26]

CONCLUSIONS

The aim of this work was an evaluation of the subjective image quality of JPEG 2000 compressed images compared to their JPEG compressed equivalents. The derived quality interval scales were obtained using paired comparison of images displayed under calibrated viewing conditions.

In summary:

1. JPEG 2000 is capable of achieving much higher compression ratios than JPEG across most images.
2. The results from both JPEG and JPEG 2000 are highly scene dependent, due to the nature of their characteristic artifacts.
3. Such scene dependencies are mainly due to the architecture of the algorithms and their operation on specific scene content, as well as the visibility of the artifacts in particular scenes.
4. For most scenes, there are small gains in quality for JPEG 2000 compared to baseline JPEG across most of the compression range (up to 80:1).
5. JPEG 2000 outperforms JPEG in terms of subjective quality for the majority of images at high compression rates (>60:1). This is likely to be due to the localization of errors within JPEG, and the visibility

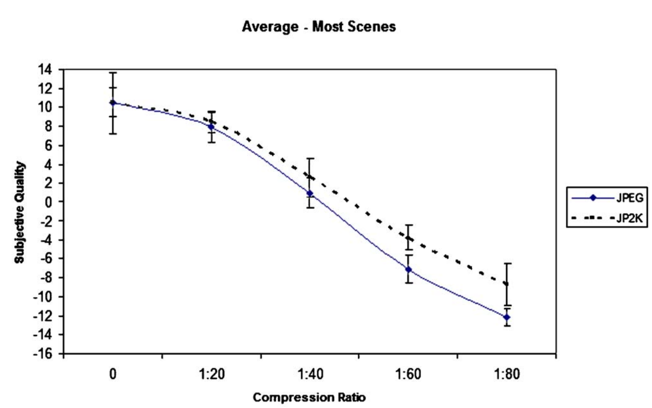

Figure 14. Average results of all scenes (except "*African tree*").

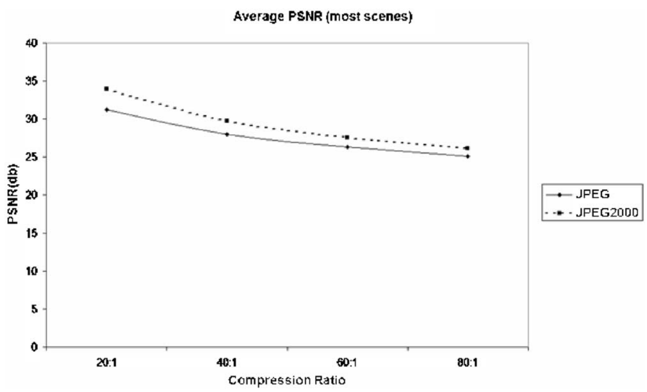

Figure 15. Average PSNR results for most scenes.

of the blocking artifacts produced by JPEG compared to the smoothing produced by JPEG 2000.

6. The differences in performance between the two algorithms is much less noticeable at lower compression ratios, (<40:1) and indeed, the slight sharpening effect of the increased ringing artifacts in JPEG is judged as a quality improvement in some images.
7. At high compression ratios blocking artifacts are generally more bothersome than smoothing artifacts in images containing large areas of flat tone or low frequencies.
8. JPEG 2000 produces less distortion of text and numerical data than JPEG.
9. Large areas of fine detail within images may mask blocking artifacts, and in such images the smoothing artifacts produced by JPEG 2000 may reduce perceived image quality.
10. JPEG 2000 outperforms JPEG in terms of error resilience across most images and most of this compression range.
11. PSNR is an inadequate predictor of subjective image quality and in particular the scene dependency affecting image quality studies.

ACKNOWLEDGMENTS

The authors are grateful to Dr. Efithimia Bilissi for allowing the use of her Paired-Comparison software programmed using Visual Basic™ 6.

REFERENCES

[1] G. K. Wallace, "The JPEG still picture compression standard", IEEE Trans. Consum. Electron. **38**, 18–34 (1992).
[2] J. L. Mitchell, and W. B. Pennebaker, "Evolving JPEG color data compression standards", Standards for Electronic Imaging systems, Critical reviews **CR37**, 68–95 (YEAR).
[3] B. E. Usevitch, "A tutorial on modern lossy wavelet image compression: Foundations of JPEG 2000", IEEE Signal Process. Mag. **18**, 22–35 (2001).
[4] S. Lawson and J. Zhu, "Image compression using wavelets and JPEG200: A tutorial", Electronics. and Communications. Engineering. Journal **14**, 112–121 (2002).
[5] J. C. Russ, *The Image Processing Handbook*, 2nd ed. (CRC Press, Inc., Boca Raton, FL, 1995), pp. 130–132.
[6] S. G. Mallat, "A yheory for multiresolution signal decomposition: The wavelet representation", IEEE Trans. Pattern Anal. Mach. Intell. **11**, PAGE (1989).
[7] C. Edwards, "Wavelet analysis transforms data processing", Scientific Computing World PAGE (1996).
[8] B. A. Cipra, "Wavelet applications come to the fore", SIAM News PAGE (1993).
[9] G. Strang, "Wavelet transforms versus fourier transforms", Bull. Am. Math. Soc. **28**, 288–305 (1993).
[10] ISO/IEC JTC 1/SC 29/WG 1 (ITU-T SG8) The JPEG 2000 Still Image Compression Standard, M. D., Adams, Dept. of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada V6T 1Z4 (2001).
[11] M. D. Adams, H. Manz, F. Kossentiniy, and T. Ebrahimi, "JPEG 2000: The next generation still image compression standard". Available at the time of writing from www.jpeg.org.
[12] A. Skodras, C. Christopoulos, and T. Ebrahimi, "The JPEG 2000 still image compression standard", IEEE Signal Process. Mag., Vol. #, 36–58 (2001).
[13] P. G. Engeldrum, *Psychometric Scaling: A Toolkit for Imaging Systems Development* (Imcotek Press, Winchester, MA, 2000), p. 1.
[14] K. Sayood, *Introduction to Data Compression* (Morgan-Kauffman Publishers, Inc., San Francisco, 1996), p. 142.
[15] A. Ford, Ph.D. thesis, University of Westminster, London, England, 1997.
[16] D. Santa-Cruz, R. Grosbois, and T. Ebrahimi, "JPEG 2000 performance evaluation and assessment", Signal Process. Image Commun. **17**, 113–130 (2002).
[17] M. Rabbani and R. Joshi, "An overview of the JPEG 2000 still image compression standard", Signal Process. Image Commun. **17**, 3–48 (2002).
[18] D. Santa Cruz and T. Ebrahimi, "A study of JPEG 2000 still image coding versus other standards", *Proc. of the X European Sign. Proc. Conference*, Vol. **2** (EURASIP, Tampere, Finland, 2000), pp. 673–676.
[19] J. C. Russ, *The Image Processing Handbook*, 4th ed. (CRC Press Inc., Boca Raton, FL, 2002), p. 121.
[20] D. R. Fuhrmann, J. A. Baro, and J. R. C. Cox, "Experimental evaluation of psychophysical distortion metrics for JPEG-encoded images", J. Electron. Imaging **4**, 397–406.
[21] R. E. Jacobson and S. Triantaphillidou, *Metric Approaches to Image Quality*, edited by L. W. MacDonald and M. R. Luo (John Wiley and Sons Ltd., New York, 2001), Ch. 18, Color Image Science.
[22] R. E. Jacobson, "An evaluation of image quality metrics", J. Photogr. Sci., **43** PAGE (1995).
[23] S. Triantaphillidou, E. Allen, and R. E. Jacobson, "Image quality comparison between JPEG and JPEG 2000 II-Scene dependency, scene analysis and classification", J. Imaging Sci. Technol. **50**, yyy (2007).
[24] B. W. Keelan, *Handbook of Image Quality: Characterization and Prediction* (Marcel Decker, Inc., New York, 2002).
[25] K. Sayood, *Introduction to Data Compression* (Morgan-Kauffman Publishers Inc, San Francisco, 1996), p. 348.
[26] U. Steingrimmson and K. Simon, "Perceptive quality estimations: JPEG 2000 versus JPEG", J. Imaging Sci. Technol. **47**, 572–585 (2003).
[27] D. Buckley, "Color imaging with JPEG 2000", *Proc. IS&T/SID 9th Color Imaging Conference* (IS&T, Springfield, VA, 2001) pp 113–119.
[28] L. L. Thurstone, "A law of comparative judgment", Psychol. Rev. **34**, 273 (1927).
[29] P. G. Engeldrum, *Psychometric Scaling: A Toolkit for Imaging Systems Development* (Imcotek Press, Winchester, MA, 2000), pp. 93–108.
[30] Software: Advanced JPEG Compressor™ Copyright 1999–2001 WinSoftMagic Development http://www.winsoftmagic.com/
[31] Software: Lurawave SmartCompress 3.0 Copyright 1996–2002 Algo Vision LuraTech GmbH Helmholtzstrasse 2–9 D-10587 Berlin. http://www.algovision-luratech.com
[32] M. Anderso, R. Motta, S. Chandrasekar, and M. Stokes, "Proposal for a standard default color space for the internet—sRGB", *Proc. IS&T/SID 4th Color Imaging Conference* (IS&T, Springfield, VA, 1996) pp. 87–93.
[33] S. Süsstrunk, R. Buckley, and S. Swen, "Standard RGB color spaces", *Proc. IS&T/SID 7th Color Imaging Conference* (IS&T, Springfield, VA, 1999) pp 127–134.
[34] E. Bilissi, R. E. Jacobson, and G. G. Attridge, "Perceptibility and acceptability of gamma differences of displayed sRGB images", *Proc. PICS 2003* (IS&T, Springfield, VA, 2003) pp. 120–125.
[35] E. M. Biederman, Photogr. Korresp. **103**, 25, 41 (1967).