

JPEG 2000 compression of medical imagery

David H. Foos^{*a}, Edward Muka^b, Richard M. Slone^b, Bradley J. Erickson^c, Michael J. Flynn^d,
David A. Clunie^e, Lloyd Hildebrand^f, Kevin Kohm^a, and Susan Young^a

^aHealth Imaging Research Laboratory, Eastman Kodak Company, Rochester, New York 14650-2033
USA

^bMallinckrodt Institute of Radiology, Washington University School of Medicine, St. Louis, Missouri
63110 USA

^cMayo Clinic and Foundation, Rochester, Minnesota 55905 USA

^dHenry Ford Health System, Detroit, Michigan 48202 USA

^eQuintiles Intelligent Imaging, Plymouth Meeting, Pennsylvania 19462 USA

^fInoveon, Oklahoma City, Oklahoma 73104-3698 USA

ABSTRACT

A multi-institution effort was conducted to assess the visual quality performance of various JPEG 2000 (Joint Photographic Experts Group) lossy compression options for medical imagery. The purpose of this effort was to provide clinical data to DICOM (Digital Imaging and Communications in Medicine) WG IV to support recommendations to the JPEG 2000 committee regarding the definition of the base standard. A variety of projection radiographic, cross sectional, and visible light images were compressed-reconstructed using various JPEG 2000 options and with the current JPEG standard. The options that were assessed included integer and floating point transforms, scalar and vector quantization, and the use of visual weighting. Experts from various institutions used a sensitive rank order methodology to evaluate the images. The proposed JPEG 2000 scheme appears to offer similar or improved image quality performance relative to the current JPEG standard for compression of medical images, yet has additional features useful for medical applications, indicating that it should be included as an additional standard transfer syntax in DICOM.

Keywords: JPEG, JPEG 2000, image compression, visual quality, wavelets, medical images, DICOM

1. INTRODUCTION

1.1. Background

JPEG^{1,2,3} is an international standard for lossy, lossless, and nearly lossless compression of continuous-tone still images and is the primary compression technology supported in DICOM.⁴ In December 1999, a draft version (committee draft) of a new international standard for image compression based on wavelet technology was completed. The new standard is officially known as the JPEG 2000 Image Coding System.⁵ The scope of the JPEG 2000 standard includes specification of the decoding processes for reconstructing compressed image data, definition of the codestream syntax required for interpreting the compressed image data, definition of a file format, and guidance for implementing the encoding processes. The draft standard specifies the base set of requirements for JPEG 2000 conformance (Part 1) and is structured to provide considerable capability while minimizing complexity. Part 2 of the standard is still under development and will include more advanced technology, but will likely involve greater computational complexity. Part 1 provides for significant improvements over current JPEG including lossy to lossless compression within a single bit stream, multiple resolution image representation, progressive decoding by image quality, region of interest coding, and improved compression efficiency.

The DICOM subcommittee, working group IV (WG IV), is responsible for compression technology evaluation and specification definition for DICOM. WG IV has been actively following the progress of JPEG 2000 development to ensure that the specialized needs of the medical imaging community are represented. In anticipation of the release of the committee draft, a formal visual quality study of the various algorithms under consideration for inclusion in JPEG 2000 was coordinated

*Correspondence: 1999 Lake Avenue, Eastman Kodak Company, Rochester, NY 14650-2033; Email: david.foos@kodak.com, Telephone: 716-722-4565; Fax: 716-722-4771

through DICOM WG IV. The timing of this study was driven by the desire to have results in time for WG IV to influence the definition of the base standard, should any issues be identified. This study was also desired to support decisions regarding the suitability of JPEG 2000 for incorporation into DICOM.

There were two specific objectives for the study. The first was to determine if important image quality performance differences existed between current JPEG and various candidate JPEG 2000 options at bit-rates, i.e., compressed bits-per-pixel, that are relevant for medical imaging applications. The second objective was to determine if important differences in medical image quality performance existed among various JPEG 2000 options under consideration for inclusion in Part 1. This study was based on a subjective rank-order protocol. The results are appropriate for making image quality comparisons about the performance of JPEG 2000 algorithms relative to the JPEG baseline and for highlighting image quality issues. However, because the study evaluated changes in appearance, and not disease interpretation, the results do not provide objective evidence about the diagnostic accuracy of a particular bit-rate for a particular pathology or capture modality. The visual study included participation from multiple university hospitals. The range of bit-rates that were used for compressing the images was determined via a pilot study with the goal of selecting bit-rates that included conservative, moderate and aggressive levels of compression. By covering a full range of bit rates corresponding to usable image quality for different applications, e.g., conservative bit rates for primary diagnostic interpretation in PACS (picture archive and communications systems) and aggressive bit rates for teleradiology for remote sites, a thorough comparison of the different algorithms could be done. An effort was made to include bit rates in the range where the rate of image quality change was greatest to maximize the likelihood of uncovering important differences in algorithm performance.

1.2. Image Compression

The image compression algorithms specified by both the JPEG and JPEG 2000 compression standards belong to the general class of transform coding.⁶ The encoding process, which converts source image data file into a compressed form, can be organized into three basic steps: transformation, quantization, and entropy coding. The decoding process, which converts compressed image data into a reconstructed image, is the inverse of the encoding process. Figure 1 shows the functional blocks for the encoding and decoding processes.

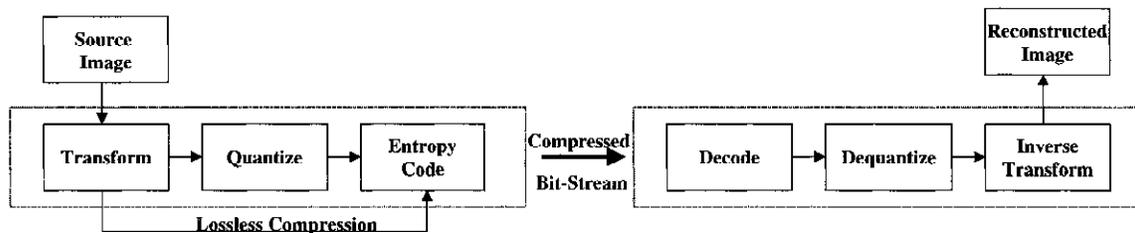


Figure 1. Image chain for transform-based image compression (encoding) and decompression (decoding).

The purpose of the transform is to decorrelate the source image, i.e., to concentrate the energy of the source image into relatively few coefficients. This energy concentration step puts the image into a form that makes it feasible to eliminate many of the coefficients while retaining the important scene information. Images with high correlation, i.e., slowly varying scene content, such as projection radiographs, are relatively easy to compress because much of the image information can be concentrated into a small number of coefficients when transformed into the spatial frequency domain. The less significant coefficients of the transformed image can then be quantized and dramatic file size reductions achieved with minimal loss of image information. Conversely, MR and CT images with high contrast edges and US images that have embedded text or pronounced speckle noise patterns, require that more information in the spatial frequency domain be preserved in order to effectively reconstruct the important scene information. Distortion generally becomes visible in these types of images at higher bit rates. Quantization reduces the amount of data used to represent the image information by limiting the number of possible values that the transformed coefficients can have.⁶ Quantization methods that treat each transform coefficient separately are known as scalar quantizers versus vector quantizers that treat sequences of coefficients. Once the transform coefficients have been quantized, the original image can no longer be identically reconstructed. The final step in the compression process is entropy coding.⁶ Entropy coding is a reversible compression process that maps the quantized coefficients, or sequences of quantized coefficients, into a new set of symbols based on probability of occurrence. Quantized coefficients that occur more frequently are assigned symbols that require fewer bits to represent. Infrequently occurring coefficient values are assigned symbols that require more bits to represent. The most commonly used methods are Huffman and arithmetic entropy coding.⁶

In Fig. 1, notice the distinction between lossless (reversible) and lossy (irreversible) compression. Lossless compression precludes the quantization portion of the encoding process. Fully reversible compression, in general, also requires that the transform algorithm perform calculations using integer arithmetic with no round-off errors. While ensuring the integrity of the source image upon decompression, lossless compression typically limits the amount of file size reduction that can be achieved to about 50% (2:1 compression ratio). Compression ratios for lossy methods, while theoretically unlimited, in practice are constrained by image quality requirements.

1.3. JPEG Standard

The JPEG standard for image compression is comprised of a toolkit that has three distinct components: baseline lossy, extended lossy, and lossless. Baseline lossy JPEG, the most widely implemented of the three, utilizes the discrete cosine transform (DCT) to decompose an image into sets of spatial frequency coefficients. The DCT is done on an 8×8 pixel block-adaptive basis. Baseline lossy JPEG supports 8 bits-per-pixel (per color) source imagery, offers a simple quantization scheme that enables users of the algorithm to trade off the degree of file size reduction, i.e., compression ratio, with image quality, and utilizes sequential Huffman entropy coding. Extended lossy JPEG is also based on the 8×8 pixel block-adaptive DCT. A feature available in extended JPEG that is critical for medical-imaging applications is the ability to handle source images with 12 bits-per-pixel. Other features available in extended JPEG include progressive and hierarchical encoding, variable quantization, and arithmetic entropy coding, however, these features are not widely supported. Lossless JPEG is an entirely separate system for image compression that is based on simple DPCM (differential pulse code modulation). A prediction value is formed for each pixel based on the values of neighboring pixels. The difference between the pixel value and the prediction value is then entropy coded using Huffman coding. Lossless JPEG can handle source images with up to 16 bits-per-pixel. The JPEG committee introduced an improved standard for lossless compression in 1997 that is known as JPEG-LS.³ JPEG-LS uses a more sophisticated context-based prediction algorithm and run-length encoding and has been shown to yield greater than 25% improvement⁷ in compression efficiency over lossless JPEG. JPEG-LS also provides a near-lossless feature that allows the a priori specification of a reconstruction error tolerance, i.e., $\pm 1, 2, 3$, etc. maximum absolute code value difference between compressed-reconstructed version and the original. The near-lossless mode provides an intuitive option for trading off compression ratio with image quality degradation that may be appealing for some medical imaging applications.⁸

1.4. JPEG 2000

JPEG 2000 is based on the two-dimensional DWT (discrete wavelet transform). The wavelet transform⁹ decomposes image signals based on scale or resolution, rather than the frequency content based decomposition resulting from the discrete cosine transform used in today's JPEG algorithm. The signal is decomposed into a lower resolution signal together with a detail signal and is calculated using the steps shown in Fig. 2.

Low-pass filter both image rows and image columns
Low-pass filter image rows and high pass filter image columns
High-pass filter image rows and low pass filter image columns
High-pass filter both image rows and image columns
Downsample by a factor of two in both image row and image column dimensions

Figure 2. Steps used in calculating a wavelet transform.

The process can be repeated recursively on the low-resolution image to create a series of reduced resolution images, also referred to as subbands. Figure 3a shows the familiar graphical representation for multiple level wavelet decomposition. Notice that by increasing the number of wavelet decomposition levels the low-low subband is further decorrelated. However, the incremental benefits resulting from successive wavelet decomposition are diminished beyond some practical limit that is primarily driven by the number of pixels in the original image. Figure 3b shows an image example of a wavelet decomposition. The detail contrast was enhanced for the HH, LH, and HL subbands for the illustration.

Like the DCT, the DWT effectively concentrates the image energy into relatively few coefficients in the transform domain. The JPEG 2000 base standard offers two choices of wavelet filters, the Daubechies $9/7$ ⁹ floating point filter because

of its superior image quality performance, and the Integer 5/3 for low complexity and lossless applications. JPEG 2000 (Part 1) quantization is performed using a scalar quantization strategy. The coefficients contained within each subband are scaled by a specified constant value, i.e., quantizer step size. The quantizer step sizes can be independently optimized (weighted) for each subband, based on image viewing conditions and contrast sensitivity considerations. The quantized coefficients are progressively coded from the most significant to least significant bit. This coding strategy enables progressive transmission-by-quality, also known as signal-to-noise-ratio (SNR) scalability. A property of SNR scalability is that the bit-stream can be terminated at any point and the image that is produced is identical to the image that would have been produced had the image been compressed to the bit rate corresponding to the truncated bit stream.¹⁰ For an image that has been lossless compressed, the SNR scalability represents a progressive lossy to fully lossless decompression capability from a single codestream. In fact, JPEG 2000 allows for a reconstructed version of the image to be generated from the lossless compressed version at any specified spatial resolution or image quality level.

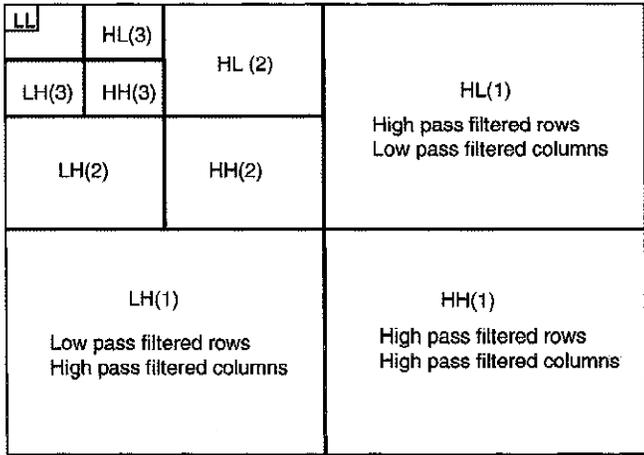


Figure 3a. Graphical representation of a multiple level wavelet decomposition.

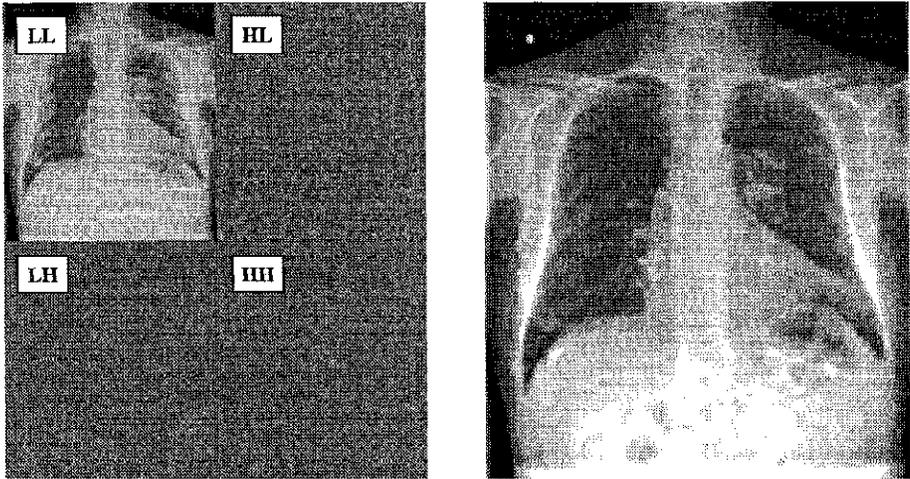


Figure 3b. Image example of a one level wavelet decomposition.

Other capabilities, which are included as part of JPEG 2000, include region-of-interest decoding and handling of source images with bit depth of up to 16 bits-per-pixel. Extended JPEG is limited to source images with 12 bits-per-pixel. Perhaps the key advantage of JPEG 2000 compression over JPEG is the flexibility afforded by the data representation resulting from the wavelet transform and embedded coding.

2. METHODS AND MATERIALS

2.1. Images

A sampling of images representing various imaging modalities including projection radiographic, cross-sectional, and visible light imaging were selected for use in the study. The Thoravision, CR (computed radiography), FD (film digitized), US (ultrasound), MR (magnetic resonance), and CT (computed tomography) images were selected by radiologists that participated in the evaluation (RMS - Washington University School of Medicine, BJE – Mayo Clinic and Foundation). The DR (flat panel digital radiography) image example was provided by the Henry Ford Health System (MJF). The visible light images were selected by an ophthalmologist (LH). Figure 4 shows the matrix of image types that were used in the study.

Images	No.	Rows	Cols	Signal Space	Bits	Vendor	Type
Digitized Film	1	2440	2048	millidensity	12	Lumisys	Abdomen
CR	1	2048	2500	p-values	12	Kodak	Portable Chest
CR	1	2500	2048	p-values	12	Kodak	Shoulder
Thoravision	1	2444	2166	log luminance	15	Philips	Chest
DR	1	2048	2048	p-values	14	GE	Chest
CT	4	512	512	crt	16	GE	Brain
MR	4	256	256	crt	16	GE	Brain
Ultrasound	4	465	640	crt	8	Accuson	Kidney, Renal Artery, Uterus, Testicle
Ophthalmology	8	1152	1152	crt	24	Kodak	Diabetic retinopathy

Figure 4. Matrix of image types used in the study.

2.2. Image Processing

The images were compressed in the signal space nominally output by the modality. Log luminance images included Philips Thoravision and FD. Kodak CR images were output in a form consistent with the DICOM Grayscale Standard Display Function (P-Values). CT, MR, US, and visible-light images were assumed to be CRT (cathode ray tube) display-ready. The image processing parameters for the projection radiographic images were based on those used for laser film prints. Appropriate LUTs (look up tables) were applied after decompression, but prior to display to calibrate the CRT to the image signal space. The CT images were displayed using window level (1064) and window width (100) specified by the radiologist that provided the images (BJE). The MR images were displayed using the values for window level and window width specified in the DICOM header for each image.

Figure 5 lists the six image compression algorithms that were tested. Five of the algorithms represent various wavelet options that are either part of the base JPEG 2000 standard, or that were under consideration to be part of the base standard. The JPEG 2000 Verification Model (VM 5.1) was used for performing the compression processing for the JPEG 2000 algorithm options. The sixth algorithm was the current JPEG operated in either the base configuration for 8-bit-per-pixel source image data or using extended JPEG for source imagery with 12-bits-per-pixel. An Eastman Kodak Company (Rochester, NY) implementation of the 12 bits-per-pixel JPEG algorithm was used. A single quantization table shape generated using tools developed by Kodak, and based on the visual optimization techniques developed by Daley,¹¹ was applied to the images for the specific viewing conditions of the study. The factors used for deriving the CSF (contrast sensitivity function) curve shape were image size (number of pixels), displayed pixel pitch, viewing distance, the light adaptation level (in cd/m^2 unit), and for the wavelet algorithms, the number of wavelet levels. Note that the CSF curve shape was derived assuming nominal 1X viewing. The images were then evaluated at 2X as a means of maximizing the perceptible differences among the algorithm variations. Another feature available in JPEG 2000 that was tested is known as visual masking. Visual masking is a non-uniform quantization method that provides an additional level of sophistication for improving image quality performance and is specifically targeted at low-resolution displays. It makes use of a nonlinear mapping of the wavelet coefficients prior to quantization. The JPEG software allowed for the specification of a target bit rate, then iterated and converged (typically 3 iterations or less) to the target bit rate to within $\pm 5\%$. The JPEG target bit rates were achieved by scaling the quantization table; the quantization table shape did not change. The JPEG 2000 target bit rates were achieved for the algorithms that used scalar quantization by truncating the bit-stream. The TCQ (trellis coded quantization) algorithm employed an iterative rate control mechanism. JPEG entropy coding was done using Global Huffman codes. JPEG 2000 utilized arithmetic coding. Because of the inherent 12-bit limitation with extended JPEG, the

Thoravision image was scaled linearly from 15 bits-per-pixel to 12 bits-per-pixel prior to JPEG compression. JPEG 2000 compression is able to handle source images with up to 16 bits-per-pixel so the compression was performed on the full 15-bit data.

The source images were compressed with each algorithm to the bits-per-pixel values shown in Fig. 6, then reconstructed. The number of levels and range of bit-rates for each modality type were determined via a pilot assessment.

Algorithms	Filter Taps	Filter Bank	Quantization
F97	Floating Point	Daubechies (9,7)	Uniform Scalar
F97_CSF	Floating Point	Daubechies (9,7)	CSF Weighted Scalar
F97_CSF_MSK	Floating Point	Daubechies (9,7)	CSF Weighted & Transducer Masked Scalar
TCQ/LRA_CSF	Floating Point	Daubechies (9,7)	CSF Weighted Trellis Coded Quantization
I53	Integer	Legall I(5,3)	CSF Weighted Scalar
JPEG_CSF	Floating Point	DCT	CSF Weighted Scalar

Figure 5. Image compression algorithms that were tested.

2.75	2.50	2.25	2.00	1.75	1.50	1.25	1.00	0.75	0.50	0.25	0.125
MR											
			CT								
			X		X	X	X	X	X	X	X
					US	US	US	US	US		
	Oph					Oph			Oph	Oph	Oph

Figure 6. Compressed image bits-per-pixel (column labels) for each modality type. An X indicates projection radiography and includes Thoravision, CR, FD, and DR; the bits-per-pixel levels were selected for each modality via a pilot assessment.

2.3. Observers

The combination of the large experiment matrix and the short three month timeframe for the study, resulted in an unbalanced set of acquired data. Different observers rated different combinations of images, in part driven by particular areas of expertise, e.g., ophthalmology, as well as capacity to participate. Figure 7 shows a breakdown of the acquired ratings data. The observers included radiologists from multiple institutions (RMS, Mallinckrodt Institute of Radiology, Washington University School of Medicine; BJE, Mayo Clinic and Foundation; WE and DW, Henry Ford Health System). One ophthalmologist (LH) participated. In addition, three image scientists experienced with viewing medical images and image compression participated as observers (KK and DHF, Eastman Kodak Company; BW, Mallinckrodt Institute of Radiology). Note that every image was evaluated by at least one radiologist and for the case of the ophthalmology images rated by the ophthalmologist. Ratings were obtained for the Thoravision image at both 1X and 2X magnification.

2.4. Evaluation Procedure

Flicker style image presentation on softcopy was used as the mechanism for visually comparing the compressed-reconstructed versions against the uncompressed originals. Flicker style comparison consists of the rapid sequential display of two versions of the same image and allows observers to readily visualize differences between the two renderings. The method has been shown to be sensitive for detecting small differences between two renderings¹² of the same image and is ideally suited for evaluating lossy compression, because the original image represents truth. Using a flicker comparison-

enabled-workstation, the CT, MR, US, CR, FD, and DR images were viewed for evaluation on 21" monochrome monitors with 2.0K × 2.5K addressable pixels. The images were displayed at 2X magnification to maximize the ability of observers to perceive subtle differences. The ophthalmology images were viewed for evaluation on a Mitsubishi Diamond Plus 100e 21" color monitor configured to display at 1024 × 768 resolution. The ophthalmology images were displayed for evaluation at 1X magnification because of the relatively large displayed pixel size of the color display.

Modality	Images per Modality	Radiologists or Ophthalmologists	Image Scientists	No. of Bit-Rates	Algorithms	Total Ratings
CR Bone	1	4	3	8	6	336
CR Chest	1	4	2	8	6	288
Thoravision	1	3	3	8	6	288
Thor (1x)	1	3	0	8	6	144
US	4	1	2	5	6	360
CT	4	1	2	7	6	504
MR	4	1	2	8	6	576
DR	1	1	1	8	6	96
FD	1	1	2	8	6	144
Oph	8	1	1	5	6	480

Figure 7. Breakdown of acquired ratings data by image type.

The flicker comparison software was installed on three separate systems, one at Kodak, one at the Mallinckrodt Institute, and one at the Mayo Clinic. Similar hardware platforms were used at all three sites. Because the experiment protocol involved image comparisons on a single display, the residual differences, after calibration, between the three systems had minimal impact on the results. The primary difference among the systems was the choice of monitor, however, all three systems used 21" monochrome monitors with 2.0K × 2.5K addressable pixels. The Kodak system used a DataRay (Westminster, CO) DR 110, 2.0K × 2.5K, 21 inch, P45 phosphor, 71 Hz monochrome monitor, the Mayo Clinic system used a Barco (Kortrijk, Belgium) MGD 521, 2.0K × 2.5K, 21 inch, P45 phosphor, 71 Hz monochrome monitor, and the Mallinckrodt system used a Siemens (Karslsruhe, Germany) SMM 21200 P, 2.0K × 2.5K, 21 inch, P45 phosphor, 71 Hz monochrome monitor.

The monitor setup and calibration procedure was performed the same way at all three sites. The maximum luminance output was calibrated to 220 Cd/m² with a dynamic range of 650:1. The natural response of the monitors to input drive level was measured and compensated to the desired aim. The desired aim resulted in a displayed image exhibiting the same tonal characteristics as the image when laser printed to film. The underlying hardware platforms were comprised of dual 300 MHz (or greater) Pentium II processors, a Microsoft (Redmond, WA) Windows NT operating system, and sufficient disk space and RAM to support the large number of evaluation images. The Kodak and Mallinckrodt systems used a Metheus (Beaverton, OR) P1540 display driver card, and the Mayo Clinic system used a Dome (Waltham, MA) MD5, 2.0K × 2.5K video card.

An introductory training session was conducted for each test participant, as necessary, prior to the start of the evaluation. The training session included a demonstration of the distortion introduced by compression, a description of the protocol, and a walk through of the procedure.

The evaluation procedure began with an observer selecting the first test image from the image list. The initial display included the control, or uncompressed original image, and the test image. The presented images were labeled as either "control" or "test". The images were displayed at 2X magnification, spatially registered, and flickering (toggling) with a period of approximately 1-second. The evaluation participant was allowed to adjust the flicker rate via the user interface or could disable the automatic flicker process in order to manually toggle (via mouse click) between control and test images. The evaluation participants were able to change magnification between 2X and 1X and, by moving the translation window, look at different regions-of-interest. Viewing distance was unrestricted. The observers were not allowed to adjust window width and window level; these controls were disabled. Using the subjective rank-order scale, shown in Fig. 8, the observers were asked to assign a numerical score to the visual degradation of the test image relative to the original. The observers

recorded their scores electronically via the user interface. Once the score was recorded, the next test image was automatically displayed. For each unique input image, the test image presentation sequence was randomized with respect to bit rate and compression algorithm. As an additional measure of control, a small number of uncompressed original images were included as test images in the evaluation.

9	None	No Detectable Difference
8		
7	Just Noticeable	No Loss of Diagnostic Information
6		
5	Conspicuous	Bordering on Loss of Information; Subtle Abnormalities Could be Overlooked
4		
3	Significant	Important Diagnostic Information Could be Lost; Degradation Impairs Interpretation
2		
1	Unacceptable	Inadequate for Diagnosis; Definite Loss of Diagnostic Information

Figure 8. Rating scale.

2.5. Analysis

A linear model was fit using Proc GLM in the SAS programming language. GLM is capable of handling class factors (which are not numerical), such as algorithm type. The model included selected two-factor interactions that were of interest, such as algorithm and image type, or that were important in explaining variability, such as observer. Other two-factor interactions that did not prove to be statistically significant, such as observer \times algorithm, were omitted. Because the data was not balanced, omitting such two-factor interactions increased the precision with which the remaining regression coefficients could be estimated. The response variable, rating, was assumed to be linear, e.g., the difference between a rating of 8 and 9 was the same as that between a rating of 1 and 2. Further, for purposes of calculating the significance of factors, the response variable was treated as if it was continuous and with normally distributed errors, although in reality it could only take on the integer values of 1 through 9. This seemed a reasonable approximation, because there were nine levels and a majority of the observers used the full range. Statistical significance was judged using a cutoff p value of 0.05, i.e., that such a result or a more extreme one, would occur by chance only 0.05 of the time if the factor in truth had no effect at all. Finally, for judging whether algorithms were different from each other for a fixed image type, Tukey's method¹³ of multiple comparison was used.

3. RESULTS

The statistically significant factors and two-factor interactions from the regression analysis are shown in Fig. 9. Several important, but obvious, factors fell out from the regression model as statistically significant. As expected, decreasing bits-per-pixel, i.e., increasing compression ratio, was found to have the greatest overall effect on visual degradation of image quality. A second-order (curvature) bit rate effect was also found to be statistically significant. This effect is consistent with the general behavior of rate distortion curves where image quality degrades gradually (almost undetectably) at first as bit rate decreases followed by a sudden quality fall-off below some threshold bit rate. This effect can be observed from the ratings versus bits-per-pixel plots shown in Figs. 10-14. Image type, or modality, was also found to be a statistically significant effect. The regression analysis separated the image types into two statistical classes, projection radiographs (Thoravision, CR, DR, FD) and everything else (US, CT, MR, Ophthalmology), with the projection radiograph image class retaining higher image quality ratings at lower bit rates than the other image class. This effect is not surprising, because the high spatial frequency content of projection radiographic images is generally less than that of the cross-sectional modalities. The differences among individual observers were also found to be statistically significant. This result is expected for many reasons, including varying contrast sensitivity thresholds for different observers and individual differences in the interpretation and application of the scoring criteria. However, the general shape of the ratings versus bit rate curves for each observer followed a similar trend.

The remaining effects shown in Fig. 9 were found to be smaller, hence less important, but also less obvious, and therefore more interesting. A significant difference was detected between two classes of observers,

radiologists/ophthalmologist and image scientists. The on-average ratings for the image scientists were lower suggesting that the radiologist/ophthalmologist observer class was more forgiving of the compression distortion effects. This effect is likely due to image scientists rating the degree of visibility of the distortion without a clear understanding of the impact on the underlying diagnostic content. Although only assessed for the Thoravision image, a strong interaction was detected between the magnification factor (1X versus 2X) and bit rate. Figure 15 shows that the images viewed at 1X magnification were rated higher than the images viewed at 2X magnification at low bit rates. This effect supports the assertion that the results from the images rated at 2X magnification are conservative estimates of the visual sensitivity to compression distortion. There was insufficient statistical power to detect if differences existed among algorithms at 1X magnification. The overall ratings performance among the six compression algorithms that were tested resulted in two statistical groupings, the floating point wavelet algorithms comprising one class and the integer wavelet algorithm (I53) and JPEG comprising the other. Although the mean differences across bit rate and image type among the algorithms was small, the differences among the algorithms at the lower bit rates are easily observed in Figs. 10-14. The wavelet/TCQ (the most computationally intensive of the algorithms that were tested), was found to have the highest overall score, but could not be statistically separated from the rest of the floating point wavelet class. One final small, but interesting effect was detected, namely a difference in the rate of degradation between the class of wavelet algorithms and JPEG. This effect suggests that although wavelet compression algorithms may degrade image quality more gradually than current JPEG as bit rate is decreased, the distortion introduced by wavelet compression may be detected sooner.

Figures 10-14 show the ratings, averaged across observers and images, plotted as a function of bit rate (compressed bits-per-pixel) for each of five image types: projection radiography, CT, MR, US, and ophthalmology. The statistical groupings for the algorithms for each image type are shown as insets adjacent to the plots. The algorithm groupings were determined using Tukey's method of multiple image comparison.

Bits-per pixel
Bits-per-pixel curvature
Image type (x-ray, everything else)
Observer
Observer type (radiologist/ophthamologist or image scientist)
Magnification x bits-per-pixel
Algorithm class (f97 wavelets vs JPEG & i53)
Algorithm x bits-per-pixel
Bits-per-pixel x image type

Figure 9. Statistically significant factors and two-factor interactions determined from regression analysis.

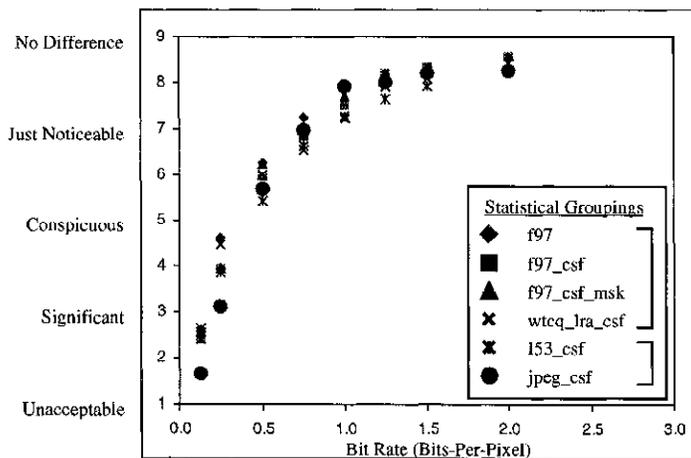


Figure 10. Average subjective image quality rating versus bit rate (bits-per-pixel) for projection radiographic image type. Ratings averaged across 7 observers for Thoravision, CR Chest, CR Bone, Digitized Film, and flat panel digital radiograph. Subjective quality rating represents perceived degradation of compressed-reconstructed image relative to the original. At low bit rate (below 0.5 bits-per-pixel) the JPEG compressed images are clearly rated as having greater degradation than images compressed using the various wavelet algorithms.

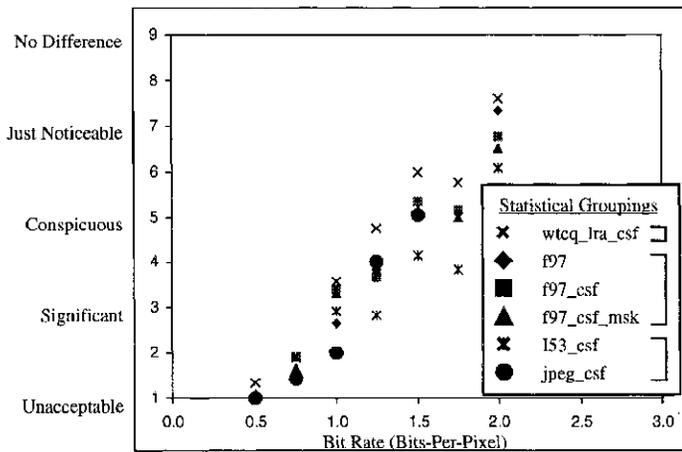


Figure 11. Average subjective image quality rating versus bit rate (bits-per-pixel) for CT image type. Ratings averaged across 3 observers and 4 images. Subjective quality rating represents perceived degradation of compressed-reconstructed image relative to the original. The plots show a trend for JPEG compressed images to be rated comparatively low at 1.0 bits-per-pixel and below and for the I53 compressed images to be rated comparatively lower between 1.5 and 2.0 bits-per-pixel. The wavelet/TCQ compressed images were consistently rated highest at all the bit rates that were tested.

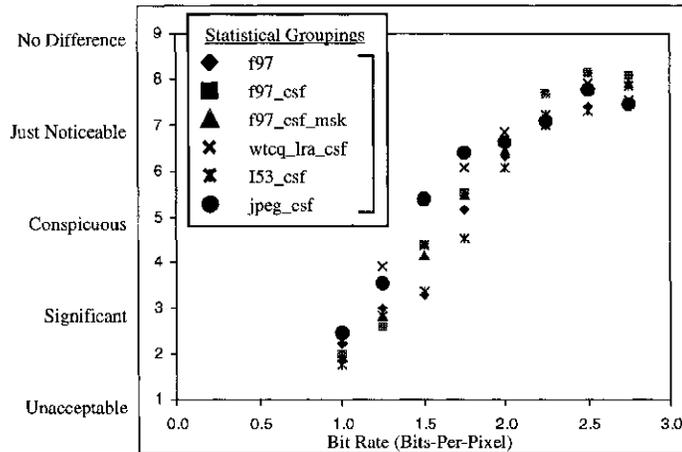


Figure 12. Average subjective image quality rating versus bit rate (bits-per-pixel) for MR image type. Ratings averaged across 3 observers and 4 images. Subjective quality rating represents perceived degradation of compressed-reconstructed image relative to the original. There were no statistically significant differences detected among the algorithms.

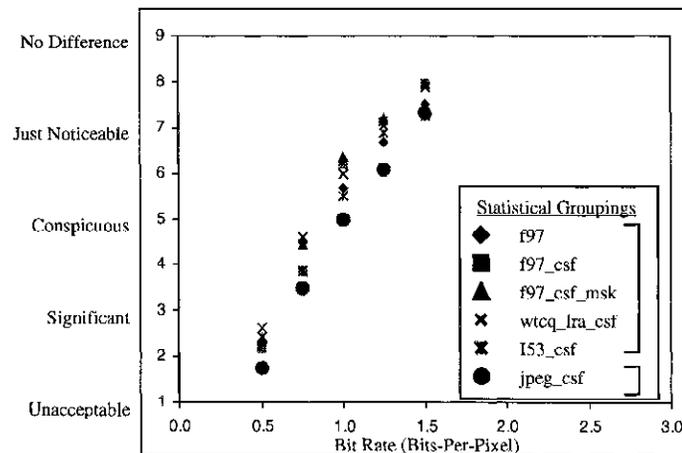
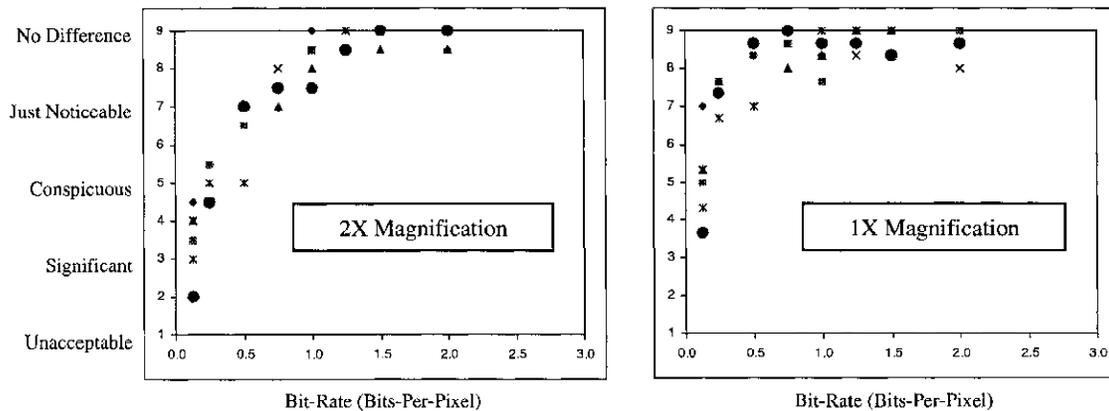
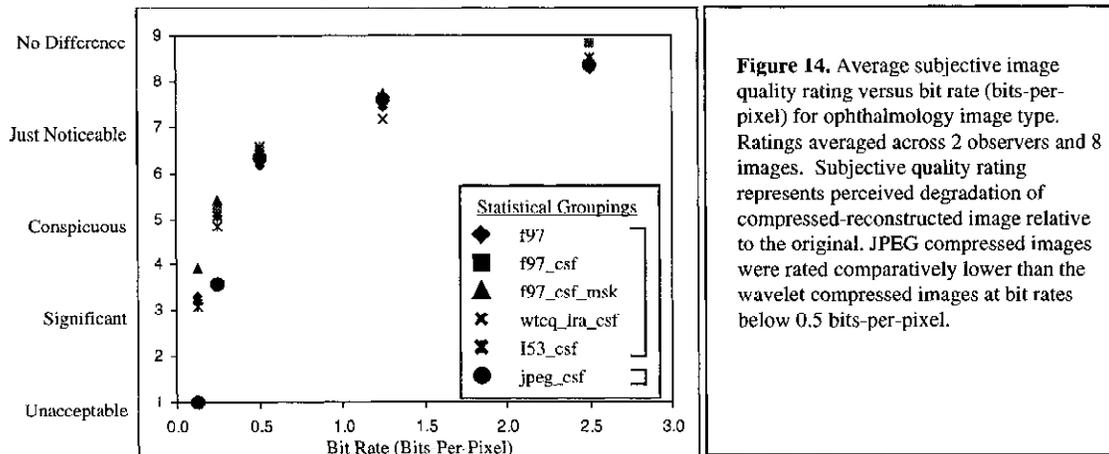


Figure 13. Average subjective image quality rating versus bit rate (bits-per-pixel) for US image type. Ratings averaged across 3 observers and 4 images. Subjective quality rating represents perceived degradation of compressed-reconstructed image relative to the original. JPEG compressed images were consistently rated lower than the wavelet compressed images.



4. CONCLUSIONS

The proposed JPEG 2000 scheme appears to offer similar or improved image quality performance relative to the current JPEG standard for compression of medical images, yet has additional features useful for medical applications, indicating that it should be included as an additional standard transfer syntax in DICOM.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the help of William Eyler, M.D. and David Wang, M.D. of the Henry Ford Health System, and Bruce Whiting, Ph.D. of the Mallinckrodt Institute for providing their time and expertise in participating as observers in the study. The authors also thank Craig Vanostrand, Ph.D. of Eastman Kodak Company for performing the statistical analysis of the observer ratings data.

REFERENCES

- ¹ W. B. Pennebaker and J. L. Mitchell, *JPEG Still Image Data Compression Standard*, Van Nostrand Reinhold, New York (1992).
- ² ITU-T Recommendation T.81 | ISO/IEC 10918-1, "Information technology – Digital compression and coding of continuous-tone still images," JPEG standard, Part 1 – Requirements and guidelines (1994).
- ³ ISO/IEC 14995-1 FDIS (Final Draft International Standard), "Information technology – Lossless and near lossless coding of continuous-tone still images: Baseline", JPEG-LS standard, Part 1 (March 1999).
- ⁴ National Electrical Manufacturers Association (NEMA) Digital Imaging and Communications in Medicine (DICOM) Publication PS 3, Rosslyn, VA (1996). Specifically image compression is discussed in Part 5 (PS 3.5), Data Structures and Encoding.
- ⁵ ISO/IEC CD 15444-1, ITU-T Rec. T.800 CDV1.0 (Committee Draft International Standard), "Information technology – JPEG 2000 Image Coding System", (December 1999).
- ⁶ M. Rabbani and P. W. Jones, *Digital Image Compression Techniques*, Vol. TT7, SPIE Press, Bellingham, WA (1991).
- ⁷ J. Kivijarvi, T. Ojala, and T. Kaukoranta et al., "A comparison of lossless compression methods for medical images," *Comput. Med. Imag. Graphics*, 22, pp. 323-339 (1998).
- ⁸ S. Young, B. Whiting, and D. Foos, "Statistically lossless image compression for CR and DR," *Proc. SPIE* 1999, Vol 3658, pp. 406-419.
- ⁹ M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image Coding Using Wavelet Transform." *IEEE Trans. Image Proc.* Vol 1-2, pp. 205-219, 1992.
- ¹⁰ P. Jones and M. Rabbani, "JPEG compression in medical imaging," *Handbook of Medical Imaging: Volume 2, Progress in Medical Image Display and PACS*, to be published by SPIE Press in Feb. 2000.
- ¹¹ S. Daley, "Applications of a Noise-Adaptive Contrast Sensitivity Function to Image Data Compression." *Opt. Eng.* vol. 29, no. 8, pp. 977-987, 1990.
- ¹² D. H. Foos, R. M. Slone, and B. Whiting et al., "Dynamic viewing protocols for diagnostic image comparison," *Proc. SPIE* 1999; Vol 3663, pp. 108-120.
- ¹³ D. C. Montgomery, *Design and Analysis of Experiments*, John Wiley and Sons, Inc., New York, (1984).