

High-Performance Wavelet Compression for Mammography: Localization Response Operating Characteristic Evaluation¹

Maria Kallergi, PhD
 Bradley J. Lucier, PhD
 Claudia G. Berman, MD
 Marla R. Hersh, PharmD, MD
 Jihai J. Kim, MD
 Margaret S. Szabunio, MD
 Robert A. Clark, MD, MBA

Purpose:

To evaluate the accuracy of a visually lossless, image-adaptive, wavelet-based compression method for achievement of high compression rates at mammography.

Materials and Methods:

The study was approved by the institutional review board of the University of South Florida as a research study with existing medical records and was exempt from individual patient consent requirements. Patient identifiers were obliterated from all images. The study was HIPAA compliant. An algorithm based on scale-specific quantization of biorthogonal wavelet coefficients was developed for the compression of digitized mammograms with high spatial and dynamic resolution. The method was applied to 500 normal and abnormal mammograms from 278 patients who were 32–85 years old, 85 of whom had biopsy-proved cancer. Film images were digitized with a charge-coupled device–based digitizer. The original and compressed reconstructed images were evaluated in a localization response operating characteristic experiment involving three radiologists with 2–10 years of experience in reading mammograms.

Results:

Compression rates in the range of 14:1 to 2051:1 were achieved, and the rates were dependent on the degree of parenchymal density and the type of breast structure. Ranges of the area under the receiver operating characteristic curve were 0.70–0.83 and 0.72–0.86 for original and compressed reconstructed mammograms, respectively. Ranges of the area under the localization response operating characteristic curve were 0.39–0.65 and 0.43–0.71 for original and compressed reconstructed mammograms, respectively. The localization accuracy increased an average of 6% (0.04 of 0.67) with the compressed mammograms. Localization performance differences were statistically significant with $P = .05$ and favored interpretation with the wavelet-compressed reconstructed images.

Conclusion:

The tested wavelet-based compression method proved to be an accurate approach for digitized mammography and yielded visually lossless high-rate compression and improved tumor localization.

© RSNA, 2006

¹ From the Department of Radiology, College of Medicine, and H. Lee Moffitt Cancer Center & Research Institute, University of South Florida, 12901 Bruce B. Downs Blvd, Tampa, FL 33612-4799 (M.K., C.G.B., M.R.H., J.J.K., M.S.S., R.A.C.); and Department of Mathematics, Purdue University, West Lafayette, Ind (B.J.L.). Received May 20, 2004; revision requested August 3; final revision received February 15, 2005; final version accepted March 2. Supported by General Telephone and Electronics Award 6121-081, National Cancer Institute award 7-R29-CA71479, and Office of Naval Research contract N00014-91-J-1152.

According to the most recent cancer statistics (1), more than 200 000 new female patients with breast cancer and about 40 000 deaths related to breast cancer were expected in 2004. Although the breast cancer incidence rate has increased almost every year since 1980, the breast cancer death rate has shown a substantial decrease. Both trends may be attributed, in large part, to mammography, which is widely recognized as the most sensitive technique for breast cancer detection (2).

Conventional mammography is a film-based x-ray technique referred to as screen-film mammography (3). Full-field digital mammography is a new technology in which a solid-state detector is used instead of film for the generation of the breast image (4). Modern applications, including computer-aided detection and computer-aided diagnosis, computer display and interpretation (soft-copy mammography), digital image transmission and storage, require a digital format of the mammogram. In screen-film mammography, the digital image is obtained by means of film digitization with dedicated high-resolution imagers. In full-field digital mammography, the output is directly digital. The use of digitized and/or digital mammography is expected to become widespread in the next 5–10 years because of the advent of the new full-field digital mammographic systems; the wider accessibility and effectiveness of picture archiving and communication systems that push the transition from film-based to filmless radiology departments; and the ever-increasing need for tele-mammographic services because of cost and reimbursement issues, the need for consultation by specialists in mammography, or the lack of mammographic clinics (5).

In either digitized or direct digital mammography, however, the resulting electronic image files are very large, and the size poses a major challenge to the transmission, storage, and manipulation processes. For example, a digitized mammographic film can yield a file of up to 90 MB of data, depending on the spatial and dynamic resolution of the digitization process, and a complete study (four films) yields files that amount to 360 MB. Similarly, digital mammograms range from

9–50 MB per view. One mammographic unit (Senographe 2000D; GE Medical Systems, Waukesha, Wis), for example, generates images at 100- μ m and 16 bit/pixel resolution (and 14 of the latter are actually used), which leads to a total of about 9 MB per image. The unit of another manufacturer (Senoscan; Fischer Imaging, Denver, Colo) generates images at 50- μ m resolution and 12 bit/pixel, which yields a total of about 50 MB per image. Even at 9 MB, the times to transmit, retrieve, or process digital images are unacceptably long, and the size of the digitized images places overwhelming requirements on storage. Advances in transmission and storage technologies alone cannot solve the problems associated with digital mammography. Image compression is a desirable process and often is essential for attaining cost and time efficiency in transmission of data (teleradiology or tele-mammography), internal communication and storage (hospital networking and archiving), and display (electronic or soft-copy interpretation).

Researchers in several studies (6,7) have demonstrated the need for high-rate compression algorithms for medical imaging applications. High compression rates, however, can only be achieved through lossy compression techniques that have not been considered acceptable for clinical applications, particularly mammography. In the past, we experimented with various wavelet bases for the compression of digitized mammograms (8,9) at high rates and demonstrated that such techniques could be acceptable for mammography because they cause losses that are not visually perceived and, hence, are not clinically important. This effect often is referred to as “visually lossless” compression. The purpose of our study was to evaluate the accuracy of a visually lossless, image-adaptive, wavelet-based compression method for achievement of high compression rates at mammography.

Materials and Methods

Database

This study was approved by the institutional review board of the University of South Florida as a research study with

existing medical records and was exempt from individual patient consent requirements. Patient identifiers were obliterated from all images. The study was compliant with the Health Insurance Portability and Accountability Act.

Five hundred single-view screening mammograms were used for the evaluation of the proposed compression method. Mammograms were selected by authors (M.K., C.G.B., and M.S.S.) from the files of 278 patients imaged at the Lifetime Cancer Screening Center of the H. Lee Moffitt Cancer Center & Research Institute, Tampa, Fla, from 1987 to 1997. The age range of patients was 32–85 years, with a mean age of 55 years. Mammograms were selected to meet the requirements of the study in terms of lesion size, histologic properties of the lesions, and breast parenchymal density. Half of the mammograms in each of the three subgroups were obtained in the craniocaudal view, and the other half were obtained in the mediolateral oblique view.

Of 500 single-view mammograms, 250 were obtained in patients with negative results and were selected from 125 patients; 131, in patients with benign results and were selected from 68 patients; and 119, in patients with cancer-

Published online

10.1148/radiol.2381040896

Radiology 2006; 238:62–73

Abbreviations:

A_z = area under the ROC curve
 CL = correctly localized
 IL = incorrectly localized
 LROC = localization response operating characteristic
 ROC = receiver operating characteristic

Author contributions:

Guarantors of integrity of entire study, M.K., B.J.L.; study concepts, M.K., B.J.L., C.G.B., R.A.C.; study design, M.K., B.J.L.; literature research, M.K., B.J.L.; clinical studies, M.K., B.J.L., C.G.B., M.R.H., J.J.K., M.S.S., R.A.C.; data acquisition and analysis/interpretation, M.K., B.J.L.; statistical analysis, M.K., B.J.L.; manuscript preparation, M.K., B.J.L.; manuscript definition of intellectual content, editing, revision/review, and final version approval, all authors

Address correspondence to M.K.

(e-mail: kallergi@moffitt.usf.edu).

Authors stated no financial relationship to disclose.

ous results and were selected from 85 patients. A total of 375 findings were present on the mammograms of the patients with benign and cancerous results; 182 of the findings were masses (98 patients with benign results and 84 with cancerous results) and 193 were calcification clusters (100 patients with benign results and 93 with cancerous results). All cancers were minimal, namely in situ noninvasive carcinoma or invasive carcinoma with a mass that had a diameter no greater than 1 cm, as established by the Breast Cancer Detection Demonstration Projects (10). Mammograms in patients with negative results were selected from those patients with at least 2 years of follow-up with negative results. Mammograms in patients with negative results matched the mammograms with positive (benign and malignant results) ones in terms of breast size, parenchymal density, and density pattern.

Films were digitized at 30- μm and 16 bit/pixel resolution with a charge-coupled device-based digitizer system (ImageClear R3000; DBA, Melbourne, Fla); hereafter, this system will be referred to as unit A. Fourteen of the 16 bits were actually used by this digitization system, as in the case of one of the imaging units (Senographe 2000D; GE Medical Systems). The pixel size was changed by a factor of two to 60- μm spatial resolution by using neighborhood averaging; pixel depth was not changed. This spatial resolution was selected on the basis of results from prior studies (11,12) that suggested that 50–90- μm and 12–16 bit/pixel resolution are optimum digitization conditions for the majority of applications, including soft-copy interpretation and processing, computer-aided detection, and computer-aided diagnosis development. We also considered the fact that all new and currently used direct digital mammographic systems generate data of similar spatial and dynamic resolution (ie, 50–100 μm and 12–16 bit/pixel, respectively).

To avoid having large areas of background (exposed area of film outside the breast area) changing the calculations of the compression rate calculations and

to focus the process more on the area of interest (breast area), images were automatically cropped to remove as much of the background area as possible. First, the breast boundary was automatically outlined to determine the breast area (13). Then, a tight rectangle that enclosed the entire breast area was drawn. Finally, the exposed part of the film outside the tight rectangle was removed to form an image smaller in size than the original. Images obtained in the same patient were cropped to the same size but images in different patients usually were of different sizes. Compression rates were determined by comparing the sizes of the original and the compressed cropped images at 60- μm spatial resolution.

Image Compression Method

The compression algorithm, designed and implemented by one author (B.J.L.), is a further development of methods used in previous investigations (8,9,14,15). Details of the new method are given in the Appendix.

The compression algorithm was applied to all mammograms in the database. Compressed images were reconstructed for soft-copy display and visual evaluation (Figs 1, 2). The compressed reconstructed images had the same dynamic range, namely the same range of pixel gray-scale values or bit depth, as the original images. There was a difference between the two images, namely between the original and compressed reconstructed data, because some information is lost with wavelet compression. The difference image was a residual noisy image and appeared as though the compression process caused removal of the signal from the original image that was corrupted by Poisson noise (ie, the signal-dependent quantum mottle) (inserts in Figs 1b and 2b) (20,21). Because of this behavior, the same or similar algorithms can be used for both wavelet compression and wavelet denoising (14,16–19,22).

On a practical note, the time for compression of a 1410×3375 -pixel image was 2.24 seconds with a 500-MHz system (UltraSparc II; Sun Microsystems, Santa Clara, Calif). Decompres-

sion was slightly faster, at 2.10 seconds, with the same system. Personal computers available in 2004 are estimated to be about four to six times as fast as the workstation used to compress the images in this study.

A mathematic evaluation of the compression output was performed by one author (B.J.L.) by using standard statistical metrics that included the calculation and comparison of the median, mean, minimum and maximum compression rates, median root mean square error, and median maximum error in gray-scale values (23).

Localization Response Operating Characteristic Experiment

Visual signal detection experiments are based on two main psychophysical techniques. The most popular in medical imaging is the receiver operating characteristic (ROC) method; in this method, a specified signal or finding of a disease may or may not be present on the image, and the observers are asked to rank their confidence about the presence or absence of the signal or the finding of a disease with a rating scale (five-point, 10-point, continuous, or other) (24,25). ROC experiments deal with signal likelihood but not with signal location. In contrast, the localization response operating characteristic (LROC) approach involves both signal likelihood and signal location tasks and, hence, offers a more complete analysis of observer performance (26,27). Thus, the LROC approach was used in our study; the design was determined and the implementation was performed by one author (M.K.).

Three board-certified mammographers (J.J.K., M.R.H., and R.A.C.) participated in the LROC experiment. At the time of the study, their experience in reading mammograms was 1, 4, and 10 years, respectively. All three reviewed 500 images on a computer monitor in two different formats (original digitized images and compressed reconstructed images) in 10 different bi-weekly sessions. One hundred images were presented per session (randomly mixed mammograms from patients with negative findings and mammograms

from patients with positive findings in the original or the compressed format); care was taken to maintain at least an 8-week interval between readings of the same mammogram in different formats. Each reading session lasted approximately 30 minutes and was conducted in a light-controlled environment where the ambient light level was maintained below 50 lux. Session volume and duration were selected to avoid potential memory bias and fatigue associated with soft-copy reading.

The observers were asked to identify and report every finding that was regarded as a suspicious breast lesion on a mammographic image. For each reported finding, the observer specified its location by a click of the computer mouse at the center of the finding and rated the likelihood that the finding was breast cancer. A five-point scale was used for rating the findings; a score of 1 corresponded to a low likelihood for cancer (or a high likelihood that the finding was normal tissue or a benign lesion) and a score of 5 corresponded to a high likelihood that the finding was cancer (or a low likelihood that it was a benign lesion). This rating scale was used for both the mammographic images with cancerous findings and those with benign findings. For the images on which no abnormalities (eg, the mammographic images with negative findings) or for mammographic images with benign findings or those with cancer on which no abnormalities could be detected, the observer was instructed to identify a single "most suspicious" area on the image and to assign a low rating (forced localization choice) according to LROC design requirements (26).

Workstation and User Interface

Display and interpretation hardware consisted of one high-resolution monitor (DR 110; Data-Ray, Westminster, Colo) with a video board (Md5/SBX; Dome, Waltham, Mass) and the same model of workstation as was used for image compression that provided a 2048×2560 -pixel display with an 8-bit digital-to-analog converter. Images were reduced in resolution by a factor of two in each dimension to fit the dis-

play. The resolution reduction was performed in real time while the image was loaded for display. A zoom option to the original spatial resolution was available for evaluation of regions of interest (256×256 pixels) on the original $60\text{-}\mu\text{m}$ -resolution image. Note that this display approach is similar to that used with most direct digital mammographic systems for the soft-copy display of digital mammograms. Monitors were calibrated before each reading session according to the Digital Imaging and Communications in Medicine part 14 gray-scale display standard with a calibration system (28).

The user interface software was programmed in the C programming language. The code was used with an image processing library (XIL; Sun Microsystems) and a volume file format (SunVision; Sun Microsystems). A dialog box on a monitor (Sun Microsystems) with a 1280×1024 -pixel display allowed selection of mammograms and reporting. In addition to the high-resolution zoom

option, the interface allowed the following options: the manual adjustment of the gray-scale values of the image by using window level, window width, and γ correction functions; the selection of the locations of the findings on the screen with the computer mouse that automatically recorded the x and y coordinates in a text file for subsequent analysis; the recording of the rating of each finding; and the review and modification of selected locations and ratings.

Statistical Analysis

The x and y coordinates of the detected findings and corresponding likelihood ratings were recorded electronically for each observer. In the first stage of the analysis, the selected x,y coordinates were compared with data in a ground truth file to determine the number of correct and incorrect localizations. The ground truth file was established by one of the authors (C.G.B.), who had 10 years of mammographic imaging experi-

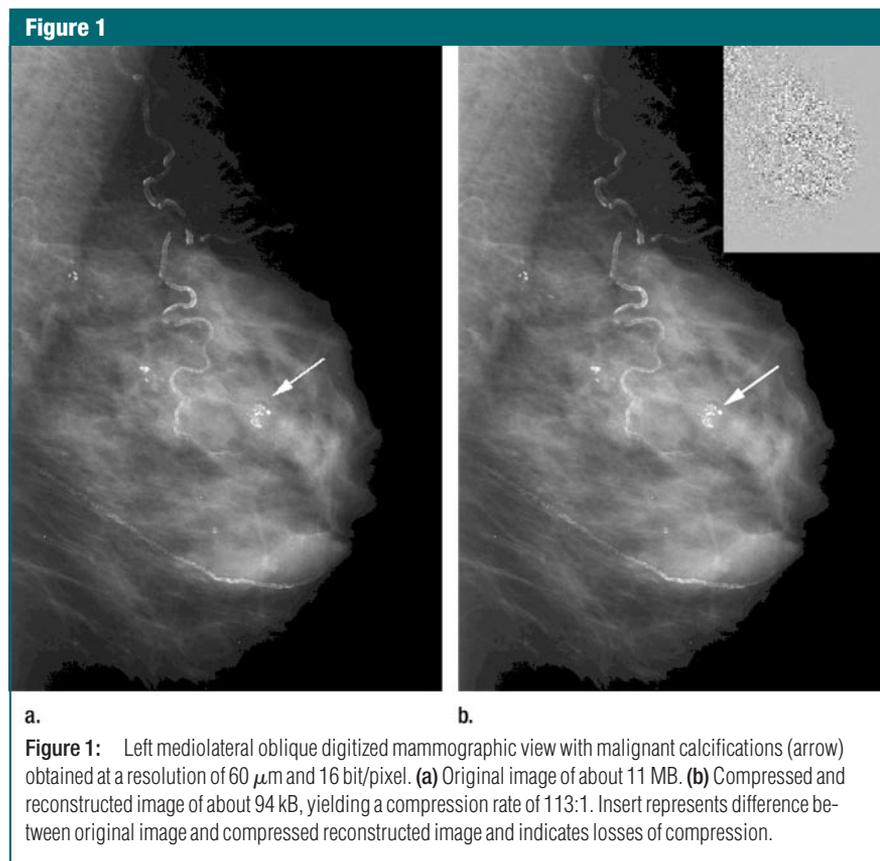


Figure 1: Left mediolateral oblique digitized mammographic view with malignant calcifications (arrow) obtained at a resolution of $60\text{ }\mu\text{m}$ and 16 bit/pixel. **(a)** Original image of about 11 MB. **(b)** Compressed and reconstructed image of about 94 kB, yielding a compression rate of 113:1. Insert represents difference between original image and compressed reconstructed image and indicates losses of compression.

ence at the time of the study. This author reviewed the patients' files, which included radiology and pathology reports, and determined the locations of the findings and their pathologic characteristics. Coordinates of each finding were recorded both in pixels and in millimeters of distance from the lower left corner of the image. A finding reported by the observer was considered correctly localized (CL) if the ordinate point selected by the observer had x,y coordinates that were within ± 200 pixels of those listed in the truth file. If the difference was greater than 200 pixels, then the finding was considered incorrectly localized (IL). The 200-pixel criterion was determined empirically by conducting an experiment in which 10 observers, with mammographic imaging experience that ranged from 3 months to 10 years, were asked to locate the centers of known lesions with ill-defined margins by using the interface described previously. On some images, the centers were different from

each other by as much as 200 pixels. All measurements were performed on the original 60- μm -resolution images.

The 1998 version of the LROC program—developed and provided by Swensson (26), Swensson et al (29), and Swensson (30)—was used in the analysis of the data. This program generates both ROC and LROC fitted curves and estimates of performance indices, which include the area under the ROC curve, A_z ; the area under the LROC curve; and the localization accuracy that corresponds to the ordinate of the LROC curve. The highest rated report of a finding on each image was used as the summary rating to represent the entire image in the analytic process (26,29). Analysis was performed by one author (M.K.) for three mammographic combinations.

1. The first combination included mammograms obtained in patients with negative findings versus those obtained in patients with cancerous findings. In this scenario, the responses in regard to the 250 images in patients with negative

findings and no lesions were compared with the responses in regard to the 119 images in patients with findings of cancerous lesions. This comparison allowed us to evaluate the ability of the readers to detect and diagnose cancer on the digitized images before and after compression. Such a data set (ie, mammograms with negative findings and mammograms with cancerous findings only) is typical in most reported ROC and LROC experiments. Most observer studies do not include mammograms in patients with benign findings in whom biopsies were performed despite the fact that these are the primary source of uncertainty and false interpretations. Here, we studied the effect of benign findings on observer performance in the next two combinations.

2. The second combination included mammograms obtained in patients with benign findings versus those obtained in patients with cancerous findings. In this scenario, the 131 mammograms in patients with benign findings were considered to signify "nontarget" or "nondisease" and were compared with the 119 mammograms in patients with cancerous findings. This comparison allowed us to evaluate the ability of the readers to differentiate between benign and cancerous findings on the soft-copy display before and after compression.

3. The third combination included mammograms obtained in patients with negative and benign findings combined versus those obtained in patients with cancerous findings. In this scenario, the 381 mammograms obtained in patients with negative and benign findings (nontarget) were compared with the 119 mammograms in patients with cancerous findings. This comparison may be closer to a clinical setup, in which the readers are asked to differentiate between patients with negative, benign, and cancerous findings to allow the focus on the latter.

Performance differences were tested for statistical significance with a two-tailed paired t test. In addition, compression rates for CL and IL lesions were calculated and compared as a function of observer, reading modality, and mammographic combination.

Figure 2

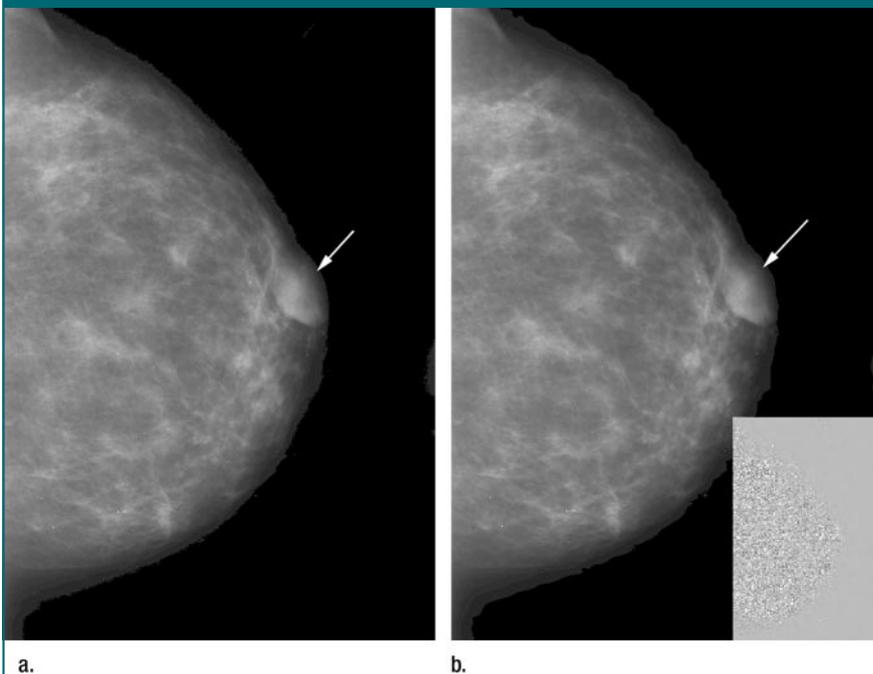


Figure 2: Left craniocaudal digitized mammographic view with a malignant mass (arrow) obtained at a resolution of 60 μm and 16 bit/pixel. **(a)** Original image of about 16 MB. **(b)** Compressed reconstructed image of about 61 kB, yielding a compression rate of 268:1. Insert represents difference between original image and compressed reconstructed image and indicates losses of compression.

Results

Compression rates ranged from 14:1 to 2051:1 (Table 1). Minimum rates were similar for the various types of lesions, whereas maximum rates differed substantially (Table 1). Of all mammographic images used in this study, 60% (302 of 500) of the images were compressed at a rate less than 100:1 and greater than 20:1. Only 6% (29 of 500) of the images were compressed at rates of 20:1 or less. Similar distributions of compression rate were observed for the various mammographic types (Fig 3). The mean compression rate was 59:1 for the mammograms in patients with negative findings, 56:1 for those in patients with benign findings, and 53:1 for those in patients with cancerous findings.

ROC and LROC curves were generated for each observer for original images and compressed reconstructed images (Fig 4). The A_z , the area under the LROC curve, the localization accuracy, and the corresponding standard error of the mean were determined for all three mammographic data combinations that were discussed earlier, before and after wavelet compression (Tables 2–4). The results of the analysis with the 95% confidence interval and the two-tailed paired t test showed that all but one of the differences were statistically significant (Table 5). The exception was for the difference in A_z values obtained for mammographic combination 2 in which observer performances for interpretation of the images from patients with benign findings and those with cancerous findings were compared. All statistically significant differences favored the compressed reconstructed version of the digitized mammograms (Table 5).

ROC and LROC parameters were also estimated for a subset of the original database that contained only independent mammograms (ie, mammograms from different patients). This analysis was performed by one of the authors (M.K.) to determine the effect of data clustering on the ROC and LROC parameters. This subset consisted of 227 images, and 106 of the

Table 1

Compression Rates, Median Root Mean Square Error, and Median Maximum Error Obtained from Mammograms

Findings on Mammograms	Compression Rate			Median Root Mean Square Error*	Median Maximum Error*
	Median	Minimum	Maximum		
Negative	65:1	14:1	2051:1	33.5	407
Benign					
Calcification clusters	67:1	14:1	1231:1	33.7	403
Masses	94:1	19:1	1045:1	29.8	422
Cancerous					
Calcification clusters	72:1	14:1	409:1	30.9	411
Masses	74:1	18:1	832:1	32.4	426

* All errors were measured in gray-scale values.

Figure 3

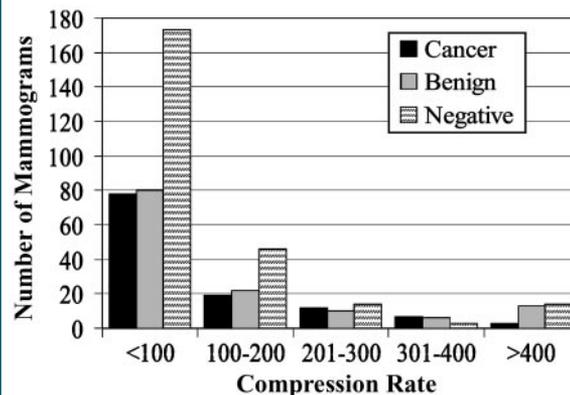


Figure 3: Bar diagram shows distribution of compression rates for mammograms from patients with negative, benign, and cancerous findings in the tested set.

images were from patients who had negative findings; 66, from patients who had benign findings; and 55, from patients who had cancerous findings. The A_z , the area under the LROC curve, and localization accuracy values of the subset differed by $\pm 2\%$ from the values obtained from the full set (Tables 2–4). The standard errors for the subset were consistently higher by as much as 10%–25% than those obtained for the full set (Tables 2–4). Statistically significant outcomes remained the same as those obtained for the full set (Table 5).

The numbers of CL and IL lesions, determined for each observer with original images and compressed reconstructed images, showed an increase in the number of CL lesions and a corresponding decrease in the number of IL lesions when interpretation was per-

formed with the compressed reconstructed images rather than with the original images. The CL and IL indices were also calculated as a function of lesion type and disease detected on images obtained before and after compression. These results showed an increase in CL lesions ranging from 7% (three of 46) to 29% (11 of 38), depending on the finding and the disease (Fig 5). The smallest increase in CL lesions was observed for the mammograms in patients who had cancerous findings with calcification clusters, and the largest increase in CL lesions was observed for the mammograms in patients who had benign findings with masses. The decrease in IL lesions ranged from 17% (three of 18) to 42% (11 of 26). The smallest decrease in IL lesions was observed for the mammograms in patients

who had cancerous findings with calcification clusters, and the largest decrease in IL lesions was observed for the mammograms in patients with benign findings with calcification clusters. Overall, readers localized findings more accurately on the compressed mammograms with benign findings than they did on the compressed mammograms with malignant findings. This result implies that the compression algorithm had an effect on image quality. Random noise seems

to have been removed from the original image (inserts in Figs 1b and 2b), and that removal affected image smoothness and appearance. At this time, it is not clear what exactly caused improved visual interpretation. Denoising effects of the wavelet algorithm seem to increase the smoothness of the image without necessarily increasing contrast. The changes in image smoothness may have contributed to the improvement in CL and IL findings seen in this study.

Discussion

In this study, we conducted an LROC experiment to measure the ability of mammographers to recognize and localize breast cancer associated with a mass or calcification clusters on original and compressed digitized mammograms displayed on high-resolution computer monitors. This allowed us to evaluate the accuracy of a visually lossless, image-adaptive, wavelet-based compression

Figure 4

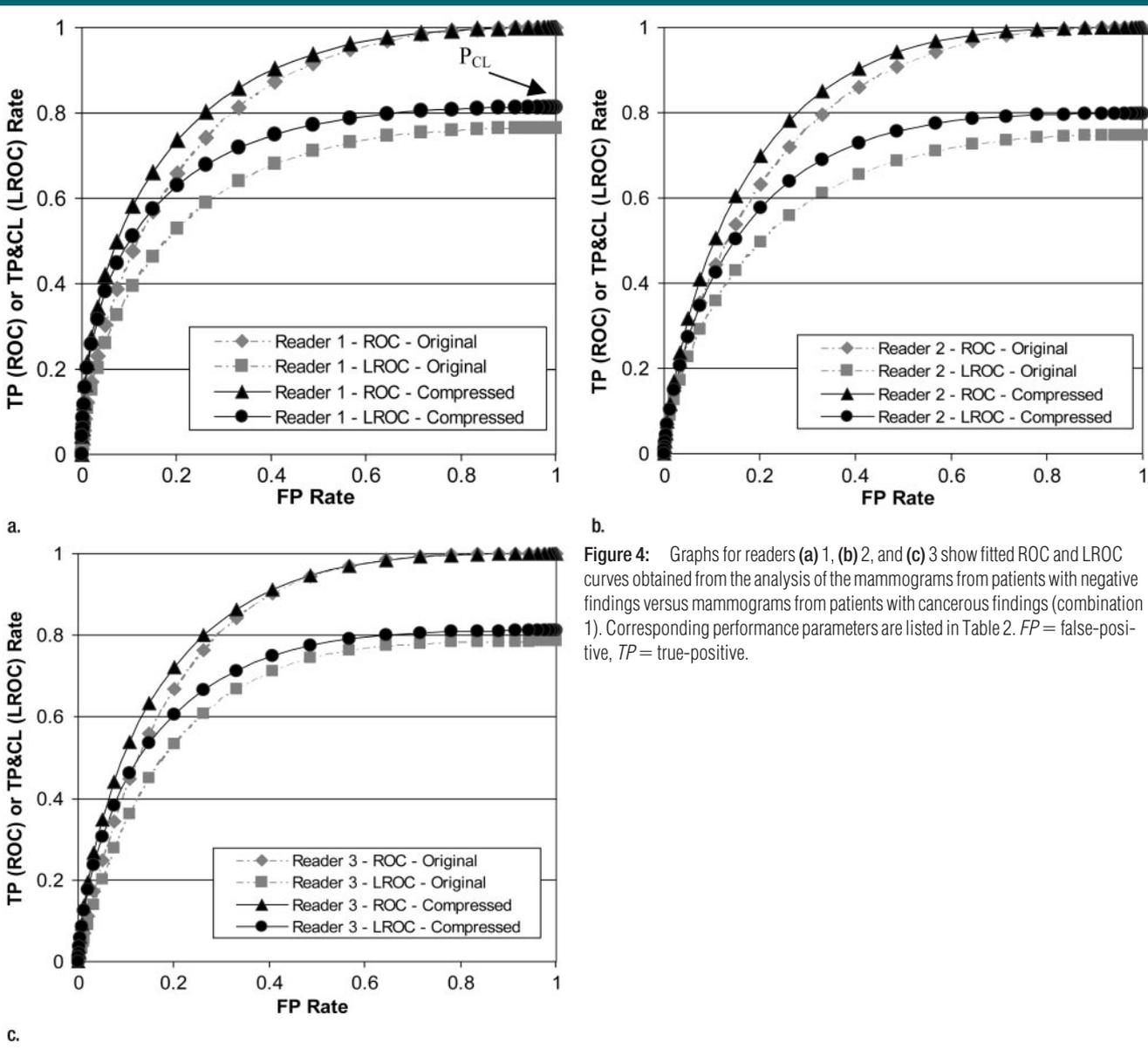


Figure 4: Graphs for readers (a) 1, (b) 2, and (c) 3 show fitted ROC and LROC curves obtained from the analysis of the mammograms from patients with negative findings versus mammograms from patients with cancerous findings (combination 1). Corresponding performance parameters are listed in Table 2. *FP* = false-positive, *TP* = true-positive.

sion method with which high compression rates were achieved. Soft-copy reading of original digitized mammograms was shown to be no more accurate than soft-copy reading of compressed reconstructed digitized mammograms, and, in many instances, soft-copy reading of original digitized mammograms underperformed soft-copy reading of compressed reconstructed digitized mammograms. This may be explained by the observation that the compression algorithm removed signal-dependent quantum noise from the images. In mammography, quantum noise is relatively low, but its removal is expected to result in a smoother image appearance that, in this study, seemed to have a favorable effect on visual perception with soft-copy display. Interestingly, the performance obtained with the images that were processed with wavelet-based compression and were reconstructed was better than that obtained with wavelet-enhanced mammograms from a separate LROC experiment in which the same database, the same workstation-user interface, and the same observers were used (31). The different results could be attributed to the different wavelet methods, which had different effects on image noise and characteristics, used in these studies.

The compression algorithm outlined in this article was specifically designed to keep diagnostically important features on mammograms unchanged. This compression algorithm was used with a local image quality measure, and it allowed both compression rates and differences between the original and compressed images to vary, depending mainly on the degree of parenchymal density. The results showed that compression rates were generally smaller for mammograms in patients with negative findings and for mammograms in patients with benign or malignant calcification clusters and larger for mammograms in patients with benign or malignant masses. Root mean square error and maximum gray-scale errors were comparable for all mammograms and types of abnormalities. These results suggested the potential for an adaptive

Table 2 **A_z , Area under LROC Curve, Localization Accuracy, and Standard Errors for Original and Compressed Images: Combination 1**

Data Set	A_z		Area under LROC Curve	Localization Accuracy	
	Value	Standard Error		Value	Standard Error
Original images					
Observer 1	0.8201	0.0180	0.6401	0.7651	0.0285
Observer 2	0.8080	0.0179	0.6160	0.7478	0.0289
Observer 3	0.8268	0.0177	0.6535	0.7853	0.0268
Compressed images					
Observer 1	0.8551	0.0173	0.7102	0.8124	0.0268
Observer 2	0.8389	0.0169	0.6778	0.7975	0.0252
Observer 3	0.8491	0.0170	0.6981	0.8112	0.0256

Table 3 **A_z , Area under LROC Curve, Localization Accuracy, and Standard Errors for Original and Compressed Images: Combination 2**

Data Set	A_z		Area under LROC Curve	Localization Accuracy	
	Value	Standard Error		Value	Standard Error
Original images					
Observer 1	0.7135	0.0201	0.4270	0.5686	0.0395
Observer 2	0.7047	0.0199	0.4095	0.5501	0.0400
Observer 3	0.6955	0.0190	0.3909	0.5730	0.0365
Compressed images					
Observer 1	0.7225	0.0201	0.4449	0.5927	0.0383
Observer 2	0.7164	0.0195	0.4327	0.5975	0.0366
Observer 3	0.7208	0.0194	0.4416	0.6138	0.0355

Table 4 **A_z , Area under LROC Curve, Localization Accuracy, and Standard Errors for Original and Compressed Images: Combination 3**

Data Set	A_z		Area under LROC Curve	Localization Accuracy	
	Value	Standard Error		Value	Standard Error
Original images					
Observer 1	0.7720	0.0182	0.5441	0.6886	0.0314
Observer 2	0.7625	0.0180	0.5249	0.6723	0.0318
Observer 3	0.7635	0.0173	0.5269	0.6928	0.0290
Compressed images					
Observer 1	0.7927	0.0177	0.5853	0.7272	0.0289
Observer 2	0.7864	0.0169	0.5728	0.7244	0.0271
Observer 3	0.7871	0.0171	0.5742	0.7247	0.0279

performance of the algorithm, depending on the clinical relevance and content of the mammogram.

Mammograms with negative, be-

nign, and cancerous findings were combined in three ways in an effort to evaluate different perspectives in observer performance. In all individual compari-

sons and for all three scenarios, the ROC and LROC curves of data from compressed images were above the curves of data from the original images. Performance parameters (eg, A_z , area under the LROC curve, and localization accuracy) also were higher for the compressed data than they were for the original soft-copy data. All but two of the differences were statistically significant with $P = .05$, and all favored the compressed reconstructed display. It should be noted that the paired t test used here to determine the statistical significance of the differences is a limited approach because the variability of the individual area measurements was not considered, and with so few degrees of freedom (two degrees of freedom), the paired t test is not very powerful. A possibly more powerful approach for multireader multicase studies might have been to apply the jackknife technique to the values of the LROC indices and apply analysis of variance, as recently suggested by Charkraborty and Berbaum (32). Overall, use of the LROC is not entirely satisfactory for experiments that involve search and localization. An alternative is the free-response ROC paradigm, which has its own

weaknesses (32). We selected the LROC method for this work to allow relative comparisons with the results of our previous studies in which the same database and analysis were used (31,33).

Localization accuracy increased for mammograms in patients with both benign and cancerous findings, but it was greater for those in patients with benign findings than it was for those in patients with cancerous findings and greater for patients with benign masses than for those with benign calcification clusters. A review of the size and contrast characteristics of the benign and cancerous lesions showed that benign lesions tended to be larger and had more contrast than did malignant lesions. Hence, it is possible that compression losses introduced by our technique had a smaller effect on the benign lesions than they did on the cancerous lesions because of the larger size of the benign lesions and increased contrast.

To put the results of our study in perspective and to help in the understanding of their potential effect on images generated with other units or various digital systems, we shall briefly discuss relevant properties of unit A. A detailed evaluation of this imaging unit

is presented elsewhere (34). This imager was based on a charge-coupled device and was developed for digital mammographic applications, including tele-mammography. It was one of the least noisy film digitization systems on the market, although its cost prevented a full commercialization of the product. Its average signal-to-noise ratio for 60- μm and 16 bit/pixel resolution data was determined to be 7.6. This is higher than the signal-to-noise ratio of a system with that of 2.8 for 50- μm resolution and 12-bit data (LS85; Kodak, Rochester, NY), hereafter referred to as unit B, and an imager with a signal-to-noise ratio of 3.3 for 43.5- μm resolution and 12-bit data (MultiRAD 850; Howtek, Nashua, NH), hereafter referred to as unit C.

Average background image grayscale values (ie, in the area outside the breast) were 12 for unit A, 257 for unit B, and 160 for unit C. Note that unit B is a laser-based system, and unit C is a charge-coupled device-based system. The low noise properties of unit A may suggest that the results obtained in this study are on the low end, as far as digitized mammography is concerned, and are possibly at the high end, as far as digital mammography is concerned. The latter conclusion is based on our initial observation that the background signal

Table 5

Statistical Values according to Mammographic Combination and Parameter

Mammographic Combination and Parameter*	Parameter Value	P Value	95% Confidence Interval
Combination 1			
Mean difference in A_z value	0.029	.016	0.013, 0.046
Mean difference in area under LROC curve	0.059	.016	0.027, 0.091
Mean difference in localization accuracy	0.041	.032	0.008, 0.074
Combination 2			
Mean difference in A_z value	0.015	.081	-0.005, 0.034
Mean difference in area under LROC curve	0.030	.083	-0.010, 0.068
Mean difference in localization accuracy	0.035	.037	0.005, 0.064
Combination 3			
Mean difference in A_z value	0.023	.002	0.018, 0.027
Mean difference in area under LROC curve	0.046	.002	0.036, 0.055
Mean difference in localization accuracy	0.041	.020	0.015, 0.066

Note.—Statistical values were determined with a two-tailed paired t test ($t = 4.303$, 2 df , $P = .05$).

* Each parameter represents the mean difference between the value for the compressed images and that for the original images for the three mammographic combinations. Combination 1 represents mammograms in patients with negative findings versus those in patients with cancerous findings; combination 2, mammograms in patients with benign findings versus those in patients with cancerous findings; and combination 3, mammograms in patients with negative and benign findings combined versus mammograms in patients with cancerous findings.

Figure 5

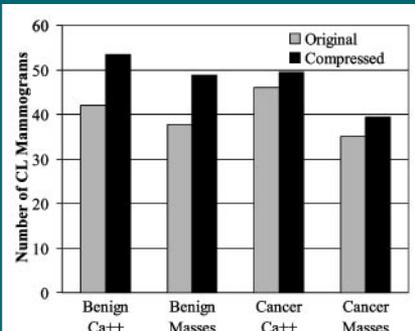


Figure 5: Bar diagram of the three-observer average values of the number of CL mammograms that contained benign and malignant calcification clusters and masses for both original and compressed reconstructed images. In all subgroups of mammograms, findings were more accurately localized on the compressed reconstructed images than on the original images.

in digital mammograms is zero for at least two of the current commercial digital systems (35). One might expect that similarly high compression rates would be achieved with our wavelet method with digital, as well as with digitized, mammography. The improved performance seen with the compressed digitized data, however, may not necessarily translate to the digital domain because of noise differences between direct digital and digitized information.

It may be argued that the compressed reconstructed soft-copy mammograms should have been compared with the original film images instead of with the digitized unprocessed images. In an earlier independent ROC study (33,36), however, in which the same mammographic database, same observers, and a similar workstation-user interface were used, we demonstrated equivalency between the interpretation with the standard film mammographic image and the interpretation with the digitized soft copy. Hence, the comparison in this study is indirectly a comparison between the standard film mammographic image and the soft-copy compressed reconstructed mammographic image. An additional incentive for the present study design was that image compression is mostly applicable to digital applications, and it is highly likely that the reference image will be the digital image and not the film image.

One may also be skeptical about the applicability and value of our compression method to film mammograms digitized differently and to direct digital mammograms. Certainly, whether these observations can be generalized to other digitized and direct digital mammograms remains to be seen. Furthermore, the implementation of the method in clinical practice will have to be carefully evaluated.

In conclusion, the tested wavelet-based compression technique proved to be an accurate approach for digitized mammography and yielded visually lossless high-rate compression and improved tumor localization. The technique could offer a clinically acceptable and effective solution to problems associated with display, transfer, and archiving

of high-resolution digitized, and possibly digital, mammograms.

Appendix

The image compression method applied and evaluated in this study is a further development of methods reported previously (8,14,16). An image is considered to be a function $f(x)$ from a rectangular region of the plane to the interval $[0,65535]$, which contains all possible pixel values reportable with the film digitizer we used. The image is decomposed by using a biorthogonal wavelet decomposition, delineated with an equation thus:

$$f(x) = \sum_j d_{j,k} \phi_{j,k}(x) + \sum_{j,k \geq K, \psi} c_{j,k,\psi} \psi_{j,k}(x), \quad (\text{A1})$$

where we have used the normalization $\phi_{j,k}(x) = \phi(2^k x - j)$, $\psi_{j,k}(x) = \psi(2^k x - j)$ found in DeVore et al (15); $x = (x_1, x_2)$, $j = (j_1, j_2)$ is a multiindex denoting the approximate location $(j_1/2^k, j_2/2^k)$, k denotes the scale, K corresponds to the coarsest scale, ϕ is the scaling function, and each ψ is a "feature" wavelet. There are three such feature wavelets, one that models oscillations vertically, one that models them horizontally, and one for "checkerboard" patterns. General information on wavelet decompositions can be found elsewhere (17). The first sum in Equation (A1) models only the broadest parts of the image at the coarsest scale; the second sum models the features that exist on the image at each dyadic scale 2^{-k} . We used the biorthogonal fifth-order accurate wavelets with piecewise constant duals of Cohen, Daubechies, and Feauveau, found in the text by Daubechies (17, p 272). We chose this family of wavelets because they are symmetric and have few oscillations for a given approximation order. The first-order member of this family is the Haar wavelet, which is discontinuous and presents too many artifacts to be used for serious compression. The third-order wavelet in this family gives reasonable visual and compression results; preliminary experi-

ments with 20 mammograms showed, however, that the fifth-order wavelet gave measurably smaller root mean square errors at the same compression rates.

An accurate description of the process requires more practical details. Specifically, the images were decomposed until there were no more than eight rows or columns (this determined the coarsest scale K noted before), at which point the wavelet rewrite rules for the associated one-dimensional synthesis and analysis scaling functions ϕ , $\tilde{\phi}$, and one-dimensional synthesis and analysis wavelets ψ , $\tilde{\psi}$, namely

$$\phi(x) = \sum_j a_j \phi(2x - j), \quad (\text{A2})$$

$$\tilde{\phi}(x) = \sum_j \tilde{a}_j \tilde{\phi}(2x - j), \quad (\text{A3})$$

$$\psi(x) = \sum_j (-1)^j \tilde{a}_{1-j} \phi(2x - j), \quad (\text{A4})$$

and

$$\tilde{\psi}(x) = \sum_j (-1)^j a_{1-j} \tilde{\phi}(2x - j), \quad (\text{A5})$$

did not have enough data to be applied. Note that in these one-dimensional expressions x is a number, not a vector, and j is an integer, not a multiindex.

We dealt with the image boundaries by reflecting the image across each edge. This approximation is only first-order accurate, but one can show mathematically that this degree of accuracy has a small effect on compression (14). The underlying wavelet rewrite rules required an even number of rows and columns at each scale of the image decomposition; whenever an odd number of rows or columns arose at a certain dyadic scale, we modeled the last unpaired value as the difference between that value and the average of the previous two values.

Given the wavelet decomposition of the image $f(x)$, the compressed image, expressed with

$$\tilde{f}(x) = \sum_j d_{j,k} \phi_{j,k}(x) + \sum_{j,k \geq K, \psi} \tilde{c}_{j,k,\psi} \psi_{j,k}(x), \quad (\text{A6})$$

was obtained by replacing the original wavelet coefficients $c_{j,k,\psi}$ with scalar quantized coefficients $\bar{c}_{j,k,\psi}$ given by

$$\bar{c}_{j,k,\psi} = q_k \text{round}(c_{j,k,\psi}/q_k), \quad (\text{A7})$$

where “round” is a function that rounds its argument to the nearest integer and q_k is a scale-dependent quantization level. It was pointed out previously that one can choose the scale-based quantization level to match the purpose of the images (15). For example, in the investigation of DeVore et al (15), it was argued that for natural images (eg, outdoor scenes, portraits) viewed at a fixed distance, choosing $q_{k+1} = 4q_k$ (which requires that a feature with half the diameter of another feature must have four times the contrast to be similarly visible) gives the best match for the human visual system. This quantization rule attempts to minimize the mean absolute error (15), expressed as follows:

$$\int |f(x) - \bar{f}(x)| dx. \quad (\text{A8})$$

Mammographic images, however, are not natural images, in that they contain specific diagnostically important features (microcalcification clusters, lesion boundaries, spiculations, architectural distortions, etc) that occur at various scales and must be preserved. On the basis of a pilot examination of 40 compressed images with various diagnostic features, the scale-dependent quantization levels of $q_{k+1} = 2q_k$ preserved the required diagnostic information better than other quantization rules. This quantization rule attempts to minimize the root mean square error and is expressed as follows:

$$[\int |f(x) - \bar{f}(x)|^2 dx]^{1/2}. \quad (\text{A9})$$

The pilot examination was performed by two senior mammographers (C.G.B. and R.A.C.) who compared the compressed images with the original images in terms of visibility and appearance of lesions, density of parenchymal tissues, demarcation of skin lines, and presence of artifacts. Compressed images were classified in one of three groups: (a) identical to the original, (b) similar to the original with minor differences, or (c) different from the original with major differences.

When we compress an image, we can choose exactly one of the following three quantities, after which the other two are fixed: (a) the number of non-zero compressed coefficients $\bar{c}_{j,k,\psi}$, which has been shown experimentally to be directly related to compression rates (15); (b) the global root mean square error, expressed as

$$[\int |f(x) - \bar{f}(x)|^2 dx]^{1/2}; \quad (\text{A10})$$

or (c) the quantization q_k at the finest dyadic level. We chose the quantization level at the finest scale (ie, 2×2 -pixel blocks) to be 128; that is, features that oscillated on a scale of 2 pixels must have had a contrast change of at least 64 gray-scale levels to be kept, otherwise they were quantized to zero. Similarly, features that oscillated on a scale of 4 pixels must have had a contrast change of at least 32 gray-scale levels to be kept, and features on a scale of 8 pixels must have had a contrast change of 16 gray-scale levels, etc. For the largest features, the quantization level q_k was not allowed to be less than one-eighth. Although these local (in each region of the image and at each scale) image quality measures q_k are fixed, the amount of compression and the measures of various global error vary, depending on the image “complexity,” and for mammograms, this variation depends mainly on the degree of parenchymal density.

Once the quantized coefficients $\bar{c}_{j,k,\psi}$ have been chosen, the compressed image is completely determined; the rest of the compression method affects only the compression rate (15). Briefly, we set up a binary decision tree (Is the quantized coefficient zero? If not, is it negative? etc). We also used the so-called Q-coder (18) as an adaptive binary arithmetic coder to encode the results. The most important factor in achieving high compression rates is encoding in a small fraction of a bit whether or not each quantized coefficient is zero (19).

References

1. Jemal A, Tiwari RC, Murray T, et al. Cancer statistics, 2004. *CA Cancer J Clin* 2004;54: 8–29.
2. Clark R. Breast cancer screening: is it worthwhile? *Cancer Control* 1995;2:189–194.

3. Haus AG. Technologic improvements in screen-film mammography. *Radiology* 1990; 174(3):628–637.
4. Kallergi M. Digital mammography: from theory to practice. *Cancer Control* 1998;5(1): 72–79.
5. Joe JE, Penhoet EE, Petitti DB, eds. Saving women's lives: strategies for improving breast cancer detection and diagnosis. Report of Committee on New Approaches to Early Detection and Diagnosis of Breast Cancer, Institute of Medicine and National Research Council. Washington, DC: National Academies Press, 2004.
6. Maitz GS, Chang TS, Sumkin JH, et al. Preliminary clinical evaluation of a high-resolution tele mammography system. *Invest Radiol* 1997;32(4):236–240.
7. Good WF, Maitz GS, Gur D. Joint photographic experts group (JPEG) compatible data compression of mammograms. *J Digit Imaging* 1994;7(3):123–132.
8. Lucier BJ, Kallergi M, Qian W, et al. Wavelet compression and segmentation of mammographic images. *J Digit Imaging* 1994; 7(1):27–38.
9. Yang Z, Kallergi M, DeVore R, et al. The effect of wavelet bases on the compression of digital mammograms. *IEEE Eng Med Biol* 1995;14(5):570–577.
10. Beahrs OH, Shapiro S, Smart C, eds. Report of the Working Group to Review the National Cancer Institute-American Cancer Society Breast Cancer Detection Demonstration Projects. *J Natl Cancer Inst* 1979;62: 639–709.
11. Kallergi M, He L, Gavrielides M, Heine J, Clarke LP. Resolution effects on the morphology of calcifications in digital mammograms. In: *Medicon '98, Proceedings of VIII Mediterranean Conference on Medical and Biological Engineering and Computing*, Lemesos, Cyprus, June 14–17, 1998 [CD-ROM]. Nicosia, Cyprus: Department of Computer Science, University of Cyprus, 1998.
12. Chan HP, Niklason LT, Ikeda DM, Lam KL, Adler DD. Digitization requirements in mammography: effects on computer-aided detection of microcalcifications. *Med Phys* 1994;21(7):1203–1211.
13. Heine JJ, Kallergi M, Chetelat SM, Clarke LP. Multiresolution wavelet approach for separating the breast region from the background in high resolution digital mammography. In: *Karssemeijer N, Thijssen M, Hendriks J, van Erning L, eds. Digital Mammography*, Nijmegen, 1998. Proceedings of the Fourth International Workshop on Digital Mammography, June 2–6, Nijmegen, the

- Netherlands. Dordrecht, the Netherlands: Kluwer Academic Publishers, 1998; 295–298.
14. Chambolle A, DeVore RA, Lee NY, Lucier BJ. Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Trans Image Process* 1998;7:319–335.
 15. DeVore RA, Jawerth B, Lucier BJ. Image compression through wavelet transform coding. *IEEE Trans Inform Theory* 1992;38:719–746.
 16. DeVore RA, Lucier BJ. Fast wavelet techniques for near-optimal image processing. In: *IEEE Military Communications Conference Record*, San Diego, October 11–14, 1992. Piscataway, NJ: Institute of Electrical and Electronics Engineers, 1992; 1129–1135.
 17. Daubechies I. Ten lectures on wavelets. Philadelphia, Pa: Society for Industrial and Applied Mathematics, 1992.
 18. Pennebaker WB, Mitchell JL, Langdon GG Jr, Arps RB. An overview of the basic principles of the Q-Coder adaptive binary arithmetic coder. *IBM J Res Dev* 1988;32:717–727.
 19. Langdon GG. An introduction to arithmetic coding. *IBM J Res Dev* 1984;28:135–149.
 20. Yaffe MJ, Nishikawa RM. X-ray imaging concepts: noise SNR, and DQE. In: Seibert JA, Barnes GT, Gould RB, eds. *Specification, acceptance testing, and quality control of diagnostic x-ray imaging equipment*. *Med Phys Monograph no. 20*. College Park, Md: American Association of Physicists in Medicine, 1994; 109–144.
 21. Kasturi R, Krile TF, Walkup JF. Image recovery from signal-dependent noise. *Opt Lett* 1983;8(7):401–403.
 22. Chang SG, Yu B, Vetterli M. Adaptive wavelet thresholding for image denoising and compression. *IEEE Trans Image Process* 2000;9:1532–1546.
 23. Eskicioglu AM, Fisher PS. Image quality measures and their performance. *IEEE Trans Commun* 1995;43(12):2959–2965.
 24. Metz CE. ROC methodology in radiologic imaging. *Invest Radiol* 1986;21:720–733.
 25. Dorfman DD, Berbaum KS, Metz CE. Receiver operating characteristic rating analysis: generalization to the population of readers and patients with the jackknife method. *Invest Radiol* 1992;27:723–731.
 26. Swensson RG. Unified measurement of observer performance in detecting and localizing target objects on images. *Med Phys* 1996;23(10):1709–1725.
 27. Seltzer SE, Swensson RG, Nawfel RD, Lentini JF, Kazda I, Judy PF. Visualization and detection-localization on computed tomographic images. *Invest Radiol* 1991;26:285–294.
 28. Digital imaging and communications in medicine (DICOM) Part 14: grayscale standard display function. PS 3.14-2004. Rosslyn, Va: National Electronic Manufacturers Association, 2004.
 29. Swensson RG, Maitz G, King JL, Gur D. Using incomplete and imprecise localization data on images to improve estimates of detection accuracy. In: Krupinski EA, ed. *Proceedings of SPIE: Medical imaging 1999—image perception and performance*. Vol 3663. Bellingham, Wash: International Society for Optical Engineering, 1999; 74–81.
 30. Swensson RG. LROC program. <http://www.radiology.arizona.edu/krupinski/mips/rocprog.html>. Accessed August 26, 2004.
 31. Kallergi M, Heine JJ, Berman CG, Hersh MR, Romilly AP, Clark RA. Improved interpretation of digitized mammography with wavelet processing: a localization response operating characteristic study. *AJR Am J Roentgenol* 2004;182:697–703.
 32. Chakraborty DP, Berbaum KS. Observer studies involving detection and localization: modeling, analysis, and validation. *Med Phys* 2004;31(8):2313–2330.
 33. Kallergi M, Hersh MR, Thomas JA. Using BIRADS categories in ROC experiments. In: Chakraborty DP, Krupinski EA, eds. *Proceedings of SPIE: medical imaging 2002—image perception, observer performance, and technology assessment*. Vol 4686. Bellingham, Wash: International Society for Optical Engineering, 2002; 60–67.
 34. Kallergi M, Gavrielides MA, Gross WW, Clarke LP. Evaluation of a CCD-based film digitizer for digital mammography. In: Van Metter RL, Beutel J, eds. *Proceedings of SPIE: medical imaging 1997—physics of medical imaging*. Vol 3032. Bellingham, Wash: International Society for Optical Engineering, 1997; 282–291.
 35. Heine JJ, Velthuisen RP. Spectral analysis of full field digital mammography data. *Med Phys* 2002;29:647–661.
 36. Kallergi M, Berman CG, Heine JJ, et al. Effect of default display and presentation protocol on softcopy mammography. In: Peitgen HO, ed. *Digital mammography IWDM 2002*. New York, NY: Springer-Verlag, 2003; 446–448.