AD_____

GRANT NO: DAMD17-94-J-4354

TITLE: Compression and Classification of Digital Mammograms for Storage, Transmission, and Computer Aided Screening

PRINCIPAL INVESTIGATOR(S): Robert M. Gray, Ph.D.

CONTRACTING ORGANIZATION: Stanford University
Stanford, California 94305-4125

REPORT DATE: August 21, 1995

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	8/21/95	Annual 1 Aug 94 - 31 Jul 95

4. TITLE AND SUBTITLE
Compression and Classification of Digital Mammograms for Storage, Transmission, and Computer Aided Screening

5. FUNDING NUMBERS
DAMD17-94-J-4354

6. AUTHOR(S)
Robert M. Gray, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Stanford University
Stanford, California 94305-4125

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The general goal of this project is to extend and validate techniques for image compression that we have developed and applied to computerized tomography and magnetic resonance images to the compression of digitized mammograms for efficient transmission, storage, and digital signal processing. This processing will include automatic classification of abnormal tissue so as to permit highlighting of suspicious areas on decompression. A specific goal is to provide compression from 12 bits per pixel (bpp) digital originals to less than 1 bpp while maintaining diagnostic accuracy and clinical utility at least equal to analog film. The performance of compression and classification and the clinical utility of the resulting images will be validated using recently developed clinical simulation and statistical analysis methods. In particular, the diagnostic accuracy of the digitized films at various levels of compression will be compared to that of the analog originals using clinical simulation and statistical analysis methods developed by us.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Compression, classification, digital mammography Breast Cancer	30 16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to ***stay within the lines*** to meet ***optical scanning requirements***.

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	- Contract	**PR**	- Project
G	- Grant	**TA**	- Task
PE	- Program Element	**WU**	- Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

- **DOD** - See DoDD 5230.24, "Distribution Statements on Technical Documents."
- **DOE** - See authorities.
- **NASA** - See Handbook NHB 2200.2.
- **NTIS** - Leave blank.

Block 12b. Distribution Code.

- **DOD** - Leave blank.
- **DOE** - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
- **NASA** - Leave blank.
- **NTIS** - Leave blank.

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the US Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

RJG Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

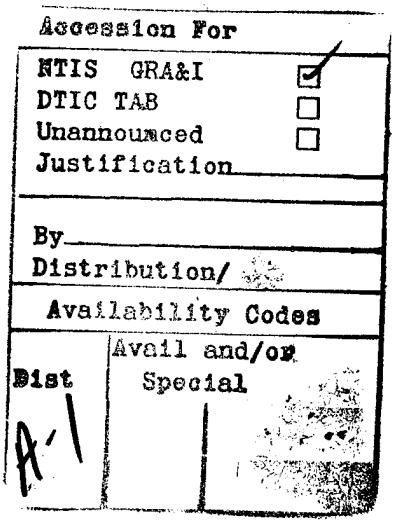
____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

RJG For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

____ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.



Robert M. Gray 8/21/95

PI - Signature Date

Contents

1 Introduction

X-ray-mammography is the most sensitive technique for detecting breast cancer [1] with a reported sensitivity of 85–95% in detecting small lesions. Most non-invasive ductal carcinomas, or DCIS, are characterized by tiny non-palpable calcifications detected at screening mammography [2, 3, 4]. Traditional mammography is essentially analog photography using X-ray in place of light and analog film for display. For a variety of reasons, digital technologies are likely to change and eventually replace most of the existing analog methods. The digital format is required for access to modern digital storage, transmission, and digital computer processing techniques. Hardcopy films use up valuable hospital space and are prone to loss and damage, which undermines the ability of radiologists to carry out future comparison studies. Images in analog format are not easily distributed to multiple sites either in-hospital or off-site, and there is the cost of personnel salary and benefits to store, archive, and retrieve the films. Currently only 30% of women get regular mammograms, and the storage problems will be compounded when this number increases with the advent of a National Health Care program. Digital image processing provides the possibilities for easy image retrieval, efficient storage, rapid image transmission for off-site diagnoses, and the maintenance of large image banks for purposes of teaching and research.

Digital signal processing allows filtering, enhancement, classification, and combining images obtained from different modalities, all of which can assist screening, diagnosis, research, and treatment. Retrospective studies of interval cancers (carcinomas detected in the time intervals between mammographic screenings which were interpreted as normal) show that observer error can comprise up to 10% of such cancers. That is to say, carcinomas present on the screening mammograms were missed by the radiologist because of fatigue, misinterpretation, distraction, obscuration by a dense breast, or other reasons [5, 6, 7]. To this end, computer-aided diagnosis (CAD) schemes may assist the radiologist in the detection of clustered microcalcifications and masses [8, 9, 10, 11]. Current CAD schemes require images in digital format.

To take advantage of digital technologies, analog signals such as X-rays must either be converted into a digital format or directly acquired in digital form. Digitization of an analog signal causes a loss of information and hence a possible deterioration of the signal. In addition, with the increasing accuracy and resolution of analog-to-digital converters, the quantities of digital information produced can overwhelm available resources. A typical digitized mammogram with 4096×4096 picture elements (pixels) with 50 micron spot size and 12 bit per pixel depth can require over 25 megabytes of data. Complete studies can easily require unacceptably long transmission times through crowded digital networks and can cause serious data management problems in local disk storage. Advances in transmission and storage technology do not solve the problem. In recent years these improvements on the internet have been swamped by the growing volume of data. Even with an ISDN line, a single X-ray can take several minutes for transmission. Therefore compression techniques are desirable and often essential for cost and time efficiency of storage and communication. The overall goal is to represent an image with the smallest possible number of bits, or to achieve the best possible fidelity for an available communication or storage bit rate capacity.

Industry alone is not likely to generate solutions to these problems because the specific needs, constraints, and performance of medical imaging are distinct from those of consumer products, which economically dwarf the medical image processing industry. For example, image quality in HDTV is evaluated typically by subjective opinions of untrained viewers, while the quality of medical images can only be determined by experts (e.g., radiologists) simulating actual clinical tasks. The tasks of medical image processing are more closely akin to those of scientific imaging (e.g., from remote sensors) because of the critical importance of subtle detail.

A compression system typically consists of one or more of the following operations, which may be combined with each other or with additional signal processing: **Sampling**: the intensity of an analog image is measured on a regular grid of points called *picture elements* or *pixels*. **Signal decomposition**: the image is decomposed into a collection of images or "bands" for separate processing, typically by linear transformation by a Fourier or discrete cosine transform or by subband filtering, possibly using wavelet filters. **Quantization**: analog or high rate digital pixels are converted into a relatively small number of bits. This operation is "lossy" as it is noninvertible, so information is lost. This loss is unavoidable if the original image is analog, as is ordinary film X-ray. The conversion can operate on individual pixels (scalar quantization) or groups of pixels (vector quantization). Quantization can arise in high resolution analog-to-digital conversion, in the zeroing of signal decomposition coefficients, or in the lossy digital compression of preserved decomposition coefficients. **Lossless compression**: further compression is achieved by a lossless code such as run-length, Huffman, Lempel-Ziv, or arithmetic code.

Decompression reverses the above process, although if the quantization is operative, the system will be lossy because the quantization is only approximately reversible. Theory and experience argue that good compression can be designed by focusing separately on each individual operation, although simpler implementations may be obtained by combining some operations. Lossless coding is well understood, readily available [12], and typically yields compression ratios of 2:1 to 3:1 on still frame greyscale medical images. This modest compression is often inadequate. Lossy coding does not permit perfect reconstruction of the original image, but can provide excellent quality at a fraction of the bit rate [13, 14, 15, 16, 17]. The *bit rate* of a compression system is the average number of bits produced by the encoder for each image pixel. If the original image has 12 bits per pixel (bpp) and the compression algorithm has rate R bpp, then the *compression ratio* is $12 : R$. Compression ratios must be interpreted with care as they depend crucially on the image type, original bit rate, sampling density, and how much coding of background goes into the calculation.

Early studies of lossy compressed medical images performed compression using variations on the standard discrete cosine transform (DCT) coding algorithm combined with scalar quantization and lossless (typically Huffman and run-length) coding. These are variations of the international standard ISO/CCITT Joint Photographic Experts Group (JPEG) compression algorithm [18, 19]. The American College of Radiology–National Electrical Manufacturers Association (ACR-NEMA) standard [20] has not yet recommended a specific compression scheme, but transform coding methods are suggested. These algorithms are well understood and have been tuned to provide good performance in many applications. More recent studies have used subband or wavelet decompositions combined with scalar or vector quantization [21, 22, 23, 24, 25] These signal decompositions provide several potential advantages over traditional Fourier-type decompositions, including better concentration of energy, better decorrelation for a wider class of signals, better basis functions for images than the smoothly oscillating sinusoids of Fourier analysis because of diminished Gibbs and edge effects, and better localization in both time and frequency. Because of their sliding-block operation using 2-dimensional linear filters, they do not produce blocking artifacts (although other artifacts arise at low rates). Vector quantization can provide advantages in some applications in terms of simplicity, speed, performance, natural progressive reconstruction, and amenability to combination with additional signal processing such as enhancement and classification for computer assisted diagnosis.

Since lossy coding can degrade image quality, making precise the notion of "excellent quality" of a compressed or processed image is a serious issue that is at the heart of this proposal. Analog mammography remains the gold standard against which all other imaging modalities can be judged, including both direct digital mammography and digitized analog mammograms. In a medical

application it does not suffice for an image to simply "look good" or to have a high signal-to-noise ratio (SNR), nor should one necessarily require that original and processed images be visually indistinguishable. Rather it must be convincingly demonstrated that essential information has not been lost and that the processed image is at least of equal utility for diagnosis or screening as the original. Image quality is typically quantified objectively by average distortion or SNR, and subjectively by statistical analyses of viewers' scores on quality (e.g., analysis of variance (ANOVA) and receiver operating characteristic (ROC) curves). Examples of such approaches may be found in [26, 15, 27, 28, 14, 13, 29].

ROC analysis is the dominant technique for evaluating the suitability of radiologic techniques for real applications [30, 31, 32, 33]. Its origins are in the theory of signal detection: a filtered version of signal plus Gaussian noise is sampled and compared to a threshold. If the threshold is exceeded, then the signal is said to be there. As the threshold varies, the probability of erroneously declaring a signal absent and the probability of erroneously declaring a signal there when it is not vary too, and in opposite directions. The plotted curve is a summary of the tradeoff in these two quantities; more precisely, it is a plot of *true positive rate* or *sensitivity* against *false positive rate*, the complement of *specificity*. Summary statistics, such as the area under the curve, can be used to summarize overall quality. In typical implementations, radiologists or other users are asked to assign integer confidence ratings to their diagnoses, and thresholds in these ratings are used in computing the curves. This approach generally differs from clinical practice and requires special training. Further, image data are not well modeled as known signals in Gaussian noise and hence methods that rely on Gaussian assumptions are suspect. Modern computer-intensive statistical sample reuse techniques can help get around the failures of Gaussian assumptions, but in fact difficulties with ROC in this specific context are more fundamental. For clinical studies that involve other than binary tasks, specificity does not make sense because it has no natural or sensible denominator, as it is not possible to say how many abnormalities are absent. This can be done for a truly binary diagnostic task for if the image is normal then exactly one abnormality is absent. Previous studies were able to use ROC analysis by focusing on detection tasks which were either truly binary or could be rendered binary. Extensions of ROC to permit consideration of multiple abnormalities have been developed [34], but these still require the use of confidence ratings as well as Gaussian or Poisson assumptions on the data, and we believe that alternative methods are preferable.

During the past seven years our group at Stanford University has developed an alternative approach to evaluating the diagnostic accuracy of lossy compressed medical images (or any digitally processed medical images) that mimics ordinary clinical practice and does not involve special training or artificial subjective evaluations, applies naturally to the detection of multiple abnormalities and to measurement tasks, and requires no assumptions of Gaussian behavior of crucial data. The methods are developed in detail for CT and MR images [35, 36, 37, 38, 39] and are sketched later.

Our general goal is the development and validation in clinical situations of lossy image compression algorithms that permit efficient and fast storage, communication, display, and analysis of digital mammograms. The proposed algorithms incorporate recent advances from signal decomposition, vector quantization, and classification tree design and combine aspects of compression with low-level classification so as to permit the best (or fastest) reproduction in areas of an image of most interest to the user. Stated formally:

Hypothesis: Digitized mammograms and lossy compressed digitized mammograms are at least as good as traditional film/screen mammography for the indication of screening asymptomatic women provided that the bit rate is sufficient. (The particular value will be estimated conservatively as a result of the experiment, but we believe it will be below 0.5 bits per pixel.)

By incorporating classification and associated highlighting into the compression, the compressed

images will be able to provide improved screening and diagnosis capabilities.

2 Technical Objectives

The specific aims of the research are:
1) To evaluate clinically the quality of of digital and lossy compressed images. The emphasis will be on digital mammograms, although the same ideas can be applied to any competing modality. Experiments have been and will be designed in conjunction with biostatisticians and radiologists to simulate as closely as possible ordinary screening and diagnostic reading of mammograms by radiologists. Emphasis is placed on experimental and statistical methods that do not involve the implicit assumptions of traditional ROC methods, but the data will also be amenable to suitable extensions of ROC-style analysis, especially when judging accuracy of patient management decisions. The goal is to judge the coded images quantitatively and qualitatively both for the detection of important features and for the preservation of selected measurements.
2) The long range goal is to compress original 12 bit images to less than 1 bpp with no loss of diagnostic accuracy using compression and decompression that are implementable in real time using currently available technology. We will consider both fully optimized algorithms, which in general will be computationally complex if implemented in software, and fast, software-based approximations. It is also desirable that the algorithms be progressive, so that image quality is improved as additional bits arrive, and scalable, so that users with a wide diversity of decompression and display platforms can extract from the bit stream the best possible reproduction for their particular platforms.
3) To combine compression with enhancement, local classification, and highlighting of features deemed important by radiologists. The goal is to incorporate such diagnostic aids into the compression/decompression algorithm with little or no increase in on-line computer processing. Clinical simulations will be conducted to quantify any gain or loss in diagnostic accuracy due to such image processing using the same basic methods as in the compression studies. Our goal is compression, but ensuring the best possible compression requires optimizing the compression for the specific application of mammography.

This third goal was primarily the topic of the third year of our original proposal, but the proposal was only funded for two years. Work continues on the basic algorithms but there will not be sufficient time to perform formal clinical experiments.

3 Methods

3.1 Study Design

The general methods to be used are extensions to digital mammography and elaborations of techniques developed for CT and MR images by our group and reported in [35, 36, 38, 40, 39, 41, 42], where all details regarding the data, compression code design, clinical simulation protocols, and statistical analyses may be found. We here describe extensions developed during the first year of this project of these methods to digital mammography. The design of the proposed mammogram evaluation study incorporates elements from both the CT and MR studies, as well as many new aspects. We propose to compare the detection of microcalcifications, masses, and other findings on analog and digital mammograms on film and compressed digital mammograms with digital originals on high resolution monitors.

The following general principles for protocol design have evolved from our earlier work on quality and utility evaluation for CT and MR images and our preliminary work with digital mammograms.
- The protocol should simulate ordinary clinical practice as closely as possible. Participating radiologists (judges, observers) should perform in a manner that mimics their ordinary practice. The studies should require little or no special training of their clinical participants.
- The clinical studies should include examples of images containing the full range of possible findings, all but extremely rare conditions.
- The findings should be reportable using the American College of Radiology (ACR) Standardized Lexicon.
- Statistical analyses of the trial outcomes should be based on assumptions as to the outcomes and sources of error that are faithful to the clinical scenario and tasks.
- "Gold standards" for evaluation of equivalence or superiority of algorithms must be clearly defined and consistent with experimental hypotheses.
- Careful experimental design should eliminate or minimize any sources of bias in the data that are due to differences between the experimental situation and ordinary clinical practice, e.g., learning effects that might accrue if a similar image is seen using separate imaging modalities.
- The number of patients should be sufficient to ensure satisfactory size and power for the principal statistical tests of interest.

We have already argued that traditional ROC analysis violates the first goal because of the requirement for confidence levels and the statistical assumptions of Gaussian or Poisson behavior. In addition, it is not well suited to the study of detection and location accuracy when a variety of abnormalities are possible. Traditional ROC analysis also does not come equipped to distinguish among the various possible notions of "ground truth" or "gold standard" in clinical experiments. We focus on three definitions of diagnostic truth as a basis of comparison for the diagnoses on all lossy reproductions of that image. These are:

Personal: Each judge's readings on an original analog image are used as the gold standard for the readings of that same judge on the digitized version of that same image,

Independent: formed by the agreement of the members of an independent expert panel, and

Separate: produced by the results of further imaging studies (including ultrasound, spot and magnification mammogram studies), surgical biopsy, and autopsy.

The first two gold standards are usually established using the analog original films. As a result, they are extremely biased in favor of the established modality, i.e., the original analog film. Thus statistical analysis arguing that a new modality is equal to or better than the established modality will be conservative since the original modality is used to establish "ground truth." The personal gold standard is in fact "hopelessly" biased in favor of the analog films. It is impossible for the personal gold standard to be used to show that digital images are *better* than analog ones. If there is any component of noise in the diagnostic decision, the digital images cannot even be found equal to analog. The personal gold standard is often useful, however, for giving some indication of the diagnostic consistency of an individual judge. The independent gold standard is also biased in favor of the analog images, but not "hopelessly" so, as it is at least possible for the readings of an individual judge on either the digital or analog images to differ from the analog gold standard provided by the independent panel. If the independent panel cannot agree on a film, the film could be removed from the study, but this wouls forfeit potentially valuable information regarding difficult images. By suitable gathering of data, one can instead define several possible independent gold standards and report the statistics with respect to each. In particular, a cautious gold standard declares a finding if any of the panel do so. An alternative is that the panel designates a chair to make a final decision when there is disagreement.

Whenever a separate gold standard is available, it provides a more fair gold standard against which both old (analog) and new (digital, compressed digital) images can be compared. When histologic data are available, they can be used to establish a separate gold standard against which results based on both analog and digital images can be compared.

As part of the first Task in our current USAMRMC project, we have acquired a database of training and test images from the Radiology Department at the University of Virginia. We have also acquired from the University of Virginia a Training Set (learning sample) for use in the vector quantization and combined compression and classification work. This data set consists of 40 images described in Table 2. We have corroborative biopsy information on at least 31 of the test and 24 of the training subjects, which can be used for a separate gold standard.

This initial data set has two shortcomings: It is too small to have good size and power for the tests proposed and the prevalence of abnormalities in this data set does not accurately reflect that of a normal screening population and hence violates the literal goals of accurate simulation and representative statistics for a screening application. The first shortcoming can be resolved by a larger study, although it is a serious and controversial issue as to how large the study must be. We shall return to this issue. The second problem, however, is unavoidable with any study of reasonable size. We will argue, however, that relevant conclusions can be drawn for the true prevalence based on a carefully constructed study using different proportions. In order to well simulate the proportion of normal images to ones containing pathology that actually would be found in a screening situation, we would require thousands of studies as there are only 6–8 cancers/1000 asymptomatic women screened. In our approach we do not *directly* estimate overall statistics for detection (sensitivity, PVP) and management (sensitivity, specificity). This would result in poor size and power for some of the statistics without unreasonably large patient numbers. It would also involve incorporating somewhat arbitrarily abnormality prevelance values reflecting the "general population." Such prevalence can vary widely depending on specific sectors of the population and a purely prospective screening study using commonly assumed prevelance values can result in requirements for more than 10,000 patients, as reported by NCI statistician Dr. L.G. Kessler at a March 6 meeting of the Radiological Devices Panel Meeting (chaired by Francine Halberg, M.D., and held at the FDA) to consider protocols for demonstrating substantial equivalence of film/screen mammography and full field digital mammography. Such an enormous study would be prohibitive in terms of cost and time and is, in our view, unnecessary. Our "retrospective/prospective" approach, reported as an alternative protocol at the 6 March Panel meeting [43] and described in a 24 February 1995 presentation to the Center for Devices and Radiological Health at the FDA by the PI, allows us to compute estimates of our statistics conditional on the presence or absense of abnormalities and to separately estimate size and power for both conditional populations. This then yields by straightforward algebra overall statistics by suitably weighting the conditional statistics to reflect estimated prevalence. The specific numbers of patients needed for good size and power will be estimated in a cumulatively improving manner as the data are gathered and the experiments performed, but preliminary analysis based on standard approximations suggests that this will be far fewer than many thousands. Our preliminary analysis based on standard approximations suggests that the following data set will suffice, as we reported in our March 1995 "strawman" proposed protocol to the FDA [43]: 400 patients of which at least 200 are normal, 110 have mammographically detected breast cancers, 75 have benign findings, and 15 have breast edemas. (See the subsection *Statistical Analysis* below.)

Because directly acquired full field digital images are not yet available, the current study uses digitized analog images. The digitized images will be compressed to three bit rates using two compression algorithms. The bit rates are aimed at providing transparent or superior quality to the

original, very good quality, and quality with distinct artifacts present. These are tentatively approximately 1.5 bpp, .45 bpp, and .15 bpp, for compression ratios of 8:1, 27:1, and 80:1, respectively. The goal of this original study is to prove the stated hypothesis for the given compression algorithms and to answer the following questions: 1) Do digital mammograms and lossy compressed digital mammograms provide equal or superior values for important statistical parameters in comparison with film screen mammography? Particular parameters of interest are sensitivity, predicted value positive (PVP or PPV), and, when it makes sense, specificity. 2) Are there any significant statistical differences between the assessment and resulting management recommendations made in clinical studies based on analog, digital, and lossy compressed digital mammograms?

In the current study images will be viewed on hardcopy film on an alternator by four judges in a manner simulating ordinary screening practice as closely as possible. The added diagnostic component is to supplement the screening simulation with additional information on diagnostic accuracy while maintaining the focus on the information in these images alone since patient histories and other image modalities will not be available. It also provides a quantitative and non-artificial rating against which ROC curves can be produced.

The ongoing study is too small to provide definitive results as it does not provide sufficient size and power for the hypotheses being tested. It is intended to demonstrate the protocol (and thereby the potential for compression in screening and diagnostic applications) and to provide data to improve our estimates of the number of patients required for an experiment with good statistical size and power. We are submitting a proposal for future studies based on larger numbers of patients (200 normal, 200 abnormal) and radiologist judges (minimum 6) to compare film screen X-ray to directly acquired digital X-ray on film as in our FDA proposed protocol, to compare digital "original" images to compressed images on high resolution monitors, and to quantify the possible benefits of optional image processing enhancements built into the compression methods.

Two views will be provided of each breast (CC and MLO), so four views will be seen simultaneously for each patient. Each of the four judges will view all the images in an appropriately randomized order over the course of nine sesssions. Two sessions will be held every other week, with a week off in between. A clear overlay will be provided for the judge to mark on the image without leaving visible trace. For each image, the judge either will indicate that the image is normal, or, if something is detected, will fill out the Observer Form in Figure 1 using the American College of Radiology (ACR) Standardized Lexicon by circling the appropriate answers or filling in blanks. The instructions for the form are given in 2. The form is intended to capture the essential information of screening with supporting detail regarding detection and assessment in a form useful for statistical analysis. The form will be filled out by a student assistant querying the radiologist for each item detected, so there may be several filled out for one patient. It attempts to preserve the information noted and considered by radiologists in drawing their conclusions. The judges will be asked to use a grease pencil to circle the detected item. The instructions to the judges specify that ellipses drawn around clusters should include all microcalcifications seen, as if making a recommendation for surgery. The masses should be outlined carefully to include the main tumor as if grading for clinical staging, without including the spicules (if any) that extend outward from the mass. This corresponds to what is done in clinical practice except for the requirement that the markings be made on copies. The judges will be allowed to use a magnifying glass to examine the films.

Although the judging form is not standard, the ACR Lexicon is used to report findings, and hence the judging requires no special training. The reported findings permit subsequent analysis of the quality of an image in the context of its true use, finding and describing anomalies and using them to assess and manage patients.

To confirm that each radiologist identifies and judges a specific finding, the location of each

lesion is confirmed both on the clear overlay and the judging form. Many of these lesions will be judged as 'A' (assessment incomplete), since it is often the practice of radiologists to obtain addtional views in two distinct scenarios: (1) to confirm or exclude the presence of a finding, that is, a finding which may or may not represent a true lesion, or (2) to further characterize a true lesion, that is, to say a lesion clearly exists but is incompletely evaluated. It is important to distinguish these two separate uses of the 'A' code since the first scenario hints at the presence of a lesion, and can be a source of false-positives if identified too often, leading to unnecessary studies or a source of false-negatives if subtle abnormalities which hint at the presence of a true cancer are missed. Similarly, it is important that true lesions identified by the radiologist should be identified in all cases, and that the use of the 'A' code is not mistaken for a possible lesion instead of a real lesion for purposes of the study.

In order to accomplish the task of separating the true meaning of the 'A' code, the judging form separates the two meaningsof the 'A' code into possible lesion or definite lesion. Furthermore, if the lesion is definite, the judges are asked to determine their suspicion of all true findings based on the standard two-view mammogram. In this way we will be able to identify possible false-positives in our data versus true findings.

The initial question requesting a rating of diagnostic utility on a scale of 1-5 is not itself used to quantify actual diagnostic utility. Rather, it is intended for a separate evaluation of the general subjective opinion of the radiologists of the images. The degree of suspicion registered in the Management portion also provides a subjective rating, but this one is geared towards the strength of the opinion of the reader regarding the cause of the management decision. It is desirable that obviously malignant lesions in a gold standard should also be obviously malignant in the alternative method.

3.2 Statistical Analysis

Detection accuracy: Once a gold standard is established, a value can be assigned to the sensitivity, the probability that something is detected given that it is present in the gold standard. Sensitivity makes sense for non-binary detection tasks, and is a crucial statistic that quantifies results. *Predictive value positive* (PVP, also called PPV), the chance an abnormality is actually present given that it is marked, fills the role of specificity in penalizing false positive reporting. Sensitivity and PVP can be measured separately for each specific lesion type. They can also be measured for the collection of all anomalies, i.e., for the identification of any of the listed lesions as opposed to none. For this case specificity also makes sense as a statistic.

Mean values for both quantities for both analog and digital images will be determined together with the two-sided 95% confidence regions. Because such data are neither Gaussian nor binary, some care is required in summarizing them and forming confidence intervals for their "true values." We will adapt computer-intensive schemes such as permutation statistics and bootstrapping [44, 42] as we have in the past to form valid confidence intervals for these two fundamental parameters.

Relative to the independent gold standard, sensitivity and PVP for the findings of the judging radiologists will be determined by whether their outlined sites largely contain the smaller circles of the independent panel (taking into account possible positioning differences on the digital mammograms). Differences in sensitivity or PVP between analog and digitized images will be analyzed using the permutation distribution of the Behrens-Fisher (Welch) statistic. The test is a variation of the two-sample t-test that takes account of differences in sample variances. As we implement the test with its permutation distribution, the test is exact in a certain sense, and does not rely on Gaussian assumptions that would be patently false for this data set. These comparisons will

be conducted for both personal and independent gold standards to demonstrate both consistency and accuracy. Sensitivity and PVP for the masses, calcifications, and other abnormalities can be evaluated both separately and combined.

Management: Management is a key issue in digital mammography. There is concern that artifacts could be introduced leading to an increase in false positives and hence in unnecessary biopsies. Statistical analysis should quantify the degree, if any, to which any such differences exist. One way to analyze the management portion of the task is to record the management decisions of (ordinary followup, further study [spot mammo, magnification mammo, other imaging]) for the two modalities in a two dimensional array of all possible pairs of the two essential decisions as in Figure 3. The counts can be used to estimate sensitivity, PVP, and specificity with respect to the personal and independent gold standards. Standard statistical methods (including simple χ^2 tests) can be used to quantify any significant differences between the management judgements of each type and as a whole. An additional statistic measuring the degree of agreement of two methods (along with confidence intervals) can be developed as follows [45, 46]: If you have several categories into which you can classify, and two ways of acquiring information, as in digital and analog, then the two methods will agree to the extent that the diagonal entries have all the probability. One can look quantitatively for the increase in agreement beyond what it would be by chance if the ratings were independent. That amounts to looking at a diagonal entry (viewed as a probability) and subtracting off the product of its row and column estimated probabilities. Summing differences over the diagonals gives a statistic for the "excess agreement." Then approximate confidence intervals follow from an asymptotic analysis. A McNemar test then can be applied to test for significant differences in management decisions as is done in [38] when there were only two categories.

An ROC-style curve can be produced by plotting the (sensitivity, specificity) pairs for the management decision for the levels of suspicion. Sample reuse methods (rather than common Gaussian assumptions) could be applied to provide confidence regions around the sample points. (Sample reuse ROC methods are considered, e.g., in [47].)

Statistical Power: We have no experimental data upon which to base precise computations of size and power in the present mammographic context. Hence we can provide only coarse approximations without resorting to additional and possibly unwarranted assumptions on the data.

It should be emphasized that "power" alone is not the issue. It makes sense only in the context of a specific size, test statistic, null hypothesis, and alternative. Once some preliminary data are available, the power and size can be computed for each test statistic described above to test the hypothesis that digital mammography of a specified bit rate is equal or superior to film/screen mammography with the given statistic and alternative hypothesis to be suggested by the data. In the absence of data, we can only guess the behavior of the collected data to approximate the power and size. We consider a one-sided test with the "null hypothesis" that, whatever the criterion (sensitivity, specificity, or predictive value positive), the digitally acquired mammograms are worse than analog. The "alternative" is that they are better. In accordance with standard practice, we take our tests to have size .05.

Approximate computations of power devolve from Figure 3. Similar methods can be applied to a table listing the detection possibilities. The key idea is twofold. In the absence of data, a guess as to power can be computed using standard approximations. Once preliminary data are obtained, however, more accurate estimates can be obtained by sample reuse techniques taking advantage of the estimates inherent in the data. One approach is to modify Figure 3 to reflect the gold standard (of whatever kind) and whatever the nonstandard decisions produce. This can test against personal and independent gold standards and, where available, against the separate standard. We abbreviate the gold standard to "Right" and the alternative to "Wrong." Figure 4

shows the possibilities and their corresponding probabilities. If the parameter γ can be estimated, then the size and power of the test for various values of the other parameters and critical values can be estimated. We do our computations of critical value for tests with size .05 in the case $h = 0$; in that case the conditional distribution of the number observed in the lower left square of the 2×2 table, given that the two off-diagonal squares have, say, N observations in total, has a binomial distribution with parameters N and .5 . If the method summarized by the columns is better ($h > 0$), then we expect the lower left square to have more observations than the upper right one. The probabilities are in terms of parameters that have not yet been estimated; they can only be guessed at conservatively. The results that follow immediately are of most interest in considering sensitivity; we will turn subsequently to computations that bear upon specificity. In this analysis we consider the experiment being proposed for a future larger study, an analysis that forms part of the current study. Suppose that ψ were .8, h .05, and γ .05 (so that the method summarized by the columns is 5% "better" than that summarized by the rows). Then for a test of size .05 (5%), the power is approximately .76 for detecting the difference by our test based on the binomial computation for our 400 overall subjects, of which, 200 are normal. Changing the parameters a bit does not alter the basic conclusion that we have reasonable power for detecting differences in sensitivity. The current study will provide data to improve the patient number estimates.

The results for size and power as they concern sensitivity are conservative in that they hold individually for each judge. If we can defend the assumption that two judges are equal in behavior, then the power increases to .95 for the combined data of those two judges. And if the data from four judges can be combined, then power increases to .999+ (for our size .05 test). If the six judges could be combined, then we could lower size to nearly 0 and have power nearly 1.

We turn now to the more delicate issue of comparing specificities. And here our approach is rather different from the approach that we have taken regarding sensitivity. Sensitivity is a "breast by breast" issue in that one commits an egregious mistake by missing disease in a single breast. Each woman was assumed in the computations thus far to contribute two breasts to the computation of sensitivity except regarding diagnoses in which asymmetry is the defining parameter. With specificity, the egregious mistake is to take a woman to biopsy of either breast when she does not require it. Here, the units for computation are individuals, and the effective sample sizes therefore are much smaller than before. The values of the parameters are quite different as well. Thus, supppose that ψ is .5, γ .25, and h .05. Then for an individual judge, the power of a test of our null hypothesis for which the size is .05 is only .27. If, however, we can combine the results of four judges, then the power of the size .05 test rises to .71, while if we can combine the results of all six judges, then the power increases to .83. The parameters we have chosen present a stern challenge to the digital technology; we could change them somewhat and change the power for various numbers of judges for our size .05 tests. We can draw a clear conclusion without presenting tables of results. That is, for a careful, powerful study of specificity, it will not be possible to make suitable conclusions without being able to combine the results of several judges – at least three and better six.

It should be emphasized that these are approximate computations in the absence of data, but that we believe the totals to be reasonable. Based on the data, size and power can be recomputed using resampling methods as a check and, if found inadequate, additional patients acquired to improve the size and power.

3.3 Compression Algorithms

This project focuses on a family of compression algorithms based on combining signal decompositions, especially subband and wavelet, with vector quantization (VQ), the conversion of vectors (typically a block of pixel intensity values in the original image such as a 2×2 square) into binary vectors which tell the decompressor which reproduction template (or codeword) from a limited set called a *codebook* should be used to best approximate the original vector. The general approach of subband/wavelet vector quantization is surveyed in Cosman, Gray, and Vetterli [48] and the basics of subband coding and vector quantization are developed, for example, in Gersho and Gray [49]. Our research specifically focuses on the quantization aspect, although we will continue to look at various choices of wavelet and of wavelet packets [50, 51]. We defer to other groups to look at the relative merits of differing decompositions [52, 53].

Basic VQ decompression is simply table lookup, yielding extremely fast image reconstruction. Recent developments provides a means of doing combined transform or subband/wavelet decoding entirely by table lookup [54, 55]. The VQ codebooks are usually either constrained to a lattice structure or designed using statistical clustering techniques that attempt to find a small number of representatives for a large data set that do a good job of representing the entire set in the sense of minimizing the average distortion between the original and the representative. A common example is the generalized Lloyd (or k-means) algorithm which has a variety of forms and successful applications [49, 56]. To lower the codebook search complexity, techniques from the design of statistical classification trees can be extended to design codebooks with a tree structure, that is, codebooks that can be searched by a sequence of simple comparisons (hyperplane or correlation) instead of a large number of distortion computations. The complexity of tree-structured VQs (TSVQs) grows linearly in bit rate instead of exponentially, as is the case with unstructured codes. This approach combines clustering with ideas from the classification and regression tree (CART) design technique of Breiman, Friedman, Olshen, and Stone [57]. TSVQ yields lower distortion than fixed rate full search VQ for a given *average* bit rate, has a simple encoder, and is well matched to variable-rate environments. TSVQ has a natural successive approximation (progressive) property, which means that instead of waiting for all the bits describing an image to arrive before displaying it, a TSVQ decoder can construct increasingly better quality images as bits arrive. A tree can be tailored by using weighted distortion measures, an attribute that plays a key role in one of the aims of this project: the optional incorporation of enhancement or highlighting into compression by using distortion measures that assign increased importance to specified features, where the features can be automatically classified or marked by a human expert in a learning data set.

After experimenting with a variety of compression algorithms, our current USAMRMC project chose two of the best current compression algorithms for evaluation: A variation of Shapiro's embedded zero tree algorithm [58] and a perceptually optimized JPEG. These schemes both use scalar quantization following the signal decomposition, but they provide good quality with reasonable complexity, demonstrate distinct low bit rate artifacts, and permit us to emphasize the validation protocol by using popular algorithms. Shapiro's embedded zerotree wavelet (EZW) algorithm [59] uses the discrete wavelet transform to generate wavelet coefficients. The algorithm then uses the idea of "zerotree" coding, in which certain coefficients are deemed "insignificant" and not coded. The insignificance of coefficients across scales are predicted by exploiting the self-similarity inherent in images. Adaptive arithmetic coding is performed on the output bit stream. An embedded code is produced since the bits in the bit stream are generated in order of importance.

Perceptually-optimized algorithms are intended to minimize distortion in a manner matched to the human psycho–visual system. The JPEG compression algorithm allows the user to customize

performance on a per image basis by modification of various parameters. One such set of parameters is the DCT quantization coefficients. Watson [60] developed an algorithm for modifying these coefficients in a perceptually optimal fashion. The work expands on the idea of threshold amplitudes for DCT basis functions presented by Peterson et al. [61]. Watson's algorithm addresses problems of luminance masking, contrast masking, error pooling and selectable quality. The algorithm produces a set of 64 numbers which can then be used by the JPEG algorithm to achieve perceptually–optimized compression.

Future compression algorithms of primary interest will be tree-structured VQ, including memoryless, predictive, and finite-state, all used in conjunction with subband/wavelet signal decompositions. We will also consider extensions and improvements of wavelet coding methods based on low complexity scalar quantization, e.g., [58]. We are actively pursuing research on these algorithms as part of our NSF and USAMRMC projects.

The general procedure with both novel and benchmark systems continus to be to simulate and test the various systems for varying parameters including different predictors, classifiers, block sizes, bit allocations, and other choices. Initial comparisons will be made on the basis of SNR vs. bit rate tradeoffs, computational complexity, and informal evaluations by radiologists. Only the most promising code structures will be selected for careful validation by clinical simulation.

In work with Professor Pamela Cosman of the University of California at San Diego (formerly a Post Doc with this project) we have developed an algorithm that combines tree-structured vector quantization with wavelet image coding [62]. This technique also uses the idea of zerotree coding. In this method, however, we explicitly use distortion/rate tradeoffs to determine the significance of the coefficients in the higher subbands. Preliminary results have shown improvements over the more common technique of using constant pre-determined thresholds to determine significance. We plan to extend the algorithm by using different VQ structures, such as lattice VQs. In addition, we plan to combine the zerotree algorithm with ideas from weighted universal VQ and classified VQ to allow the code to better match distinct local behavior [63, 64]. We have also explored a low complexity multiresolution approach that uses pruned nested TSVQs. This technique has produced several dB improvement over basic VQ schemes. Furthermore, the algorithm produces images that can be transmitted progressively in both a spatial and frequency multiresolution manner. The low complexity nature of the algorithm makes it useful for applications that require fast decoding with low complexity in software [62, 65].

Two additional methods are of particular interest because of their intimate connection with combined compression and classification for computer assisted diagnosis of mammograms to be discussed later and because of their potential for improving compression alone: classified VQ and finite-state VQ [49]. Both of these methods have a collection of small codebooks (which can be thought of as custom compression algorithms) available to the encoder for the current pixel block, where each codebook corresponds to a distinct mode or type of behavior. For example, images will have different local dynamic ranges or different textures such as fatty vs. dense tissue. If one is able to distinguish a small collection of classes or types for the local behavior, then a smart compression system might have a separate code available for each. The encoder picks the best codebook for the class to which the current block belongs and then code the block using that codebook. Classified VQ and finite-state VQ differ in how the class is chosen and communicated to the decoder, but the design techniques for the two systems are quite similar [49]. Both schemes provide useful byproducts in the identification of classes of possible use to the physician. The design is complicated, but the actual compression/decompression once the codebook is fixed is simple. These codes have not yet been applied to wavelet coding systems, but they appear to be naturally suited for the application in that the classification for quantizing the output of each level can be computed from the higher

resolution previous level. In a way Shapiro's embedded zero tree algorithm does this, effectively coding all descendent coefficients from a low energy high resolution coefficient as a zero rate "zero tree." More generally, each high resolution pixel or pixel block could be classified to determine which codes would be used on descendent pixel blocks in the decomposition, with bit rate being traded off for average distortion. Lastly, finite-state VQ appears a good match to the sliding-block nature of wavelet coders, which are also a form of finite-state machine when run on discrete data. We propose to investigate a variety of finite-state VQ design algorithms, including bit allocation techniques, variable rate state codebooks, and differing classifiers such as CART, VQs, and the Bayes classifiers for abnormalities considered next.

3.4 Combined Compression and Classification for Highlighting

A variety of techniques for automatically locating abnormalities such as lung nodules, microcalcifications, and masses have been reported in the literature (e.g., [10, 11, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76]). The techniques typically involve full frame sophisticated signal decompositions and segmentation for enhancing the image, extraction of important features, and application of pattern recognition algorithms to classify regions based on the observed features. Many of the algorithms apply morphological methods to thresholded images, effectively eliminating relative pixel intensity information that clustering tree-structured methods can use to advantage. Most published algorithms are computationally complex, often requiring long times to perform the image analysis. A notable exception is the approach of Kegelmeyer [8] who uses the CART algorithm, developed in part by R.A. Olshen and intimately connected with the techniques proposed here. Since such algorithms are performed digitally, quantization is necessary if the original image is an analog X-ray. Quantization may be desirable even for digital images, however, as reducing the bit rate can speed the subsequent processing. With the exception of our recent work to be described, all published techniques of which we are aware make no attempt to match the quantizer design to the subsequent classification step, but rather separately and independently design the compressor and classifier or design the two in simple cascade using separate criteria. For example, a VQ could be designed to minimize average squared error and then a Bayes classifier could be designed for the VQ output. This approach is common and intuitive—if the quantization has enough bits, the digitized signal should well approximate the original and hence a classifier designed for the original should still work well. The intuition is not necessarily appropriate, however, when high compression is required and there is no guarantee that high SNR will translate into preservation of essential classification information. A potential solution is to incorporate the classifier's goal into the quantizer design. This can provide a simple, fast, and useful quantizer that provides some classification and preserves essential information for a subsequent more sophisticated classifier. The idea of combining VQ with classification is based on the simple observation that both techniques are optimized by best balancing a tradeoff between distortion or cost and complexity and hence one can incorporate both notions of distortion–error energy and Bayes risk–to a single general distortion measure which can be used to design the code. By combining these ideas with pyramid or other multiresolution coding schemes, increasingly larger features can be included in the optimization algorithm used to design the codes. For example, small pixel blocks can attempt to identify individual microcalcifications; larger blocks can look for clusters and masses.

The basic idea is to consider coding not just a pixel intensity block X, but the pair (X, Y), where Y is a "class label" which takes values in a finite set $\mathcal{H} = \{0, \cdots, M-1\}$; we wish to accurately guess the class Y when only the observable X or, in our approach, a quantized pixel block, is known. In otherwords, the information necessary to segment the image into classes is contained in

bits describing the image. For mammography there could be two class labels corresponding to "microcalcification present" and its complement, or three labels corresponding to "microcalcification," "mass," and "neither microcalcification nor mass." Composite classes could assign elementary classes to each pixel within a vector, e.g., separately identifying microcalcifications, mass, and other classes.

Typically classifier performance is measured by Bayes risk, a weighted combination of error probabilities where different error types can be assigned different costs. For example, in the two class problem (microcalcification present or not), the Bayes risk is a weighted combination of the probabilities of missing a microcalcification that is there and declaring a microcalcification that is not there; our weighting makes the first error type far more important. Here the decision is made on the same size pixel group as is used for the VQ and hence this decision problem can be considered to be binary: either the small pixel square is part of a microcalcification or it is not. Much of the theory and practice of classification is aimed at finding a classifier that minimizes Bayes risk. Our approach is a variation on empirical Bayes detection where the necessary probabilities are learned from a labeled set of training data, e.g., radiographs marked to indicate important features such as calcifications.

Our method [77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87] uses a modified distortion measure in the design and application of the code and allows simultaneous optimization for both compression (using squared error or other objective distortion measure for general appearance) and Bayes risk (for classification accuracy) by combining the two terms with a Lagrangian importance weighting. The Lagrange multiplier determines the relative importance of squared error and classification, but preliminary results show that the classification accuracy can be weighted quite heavily while still producing excellent compression. A simple variation of the Lloyd algorithm can then be used to design the code. The intensive computation occurs during code design, not during compression or decompression. For our current set of training images, the masses and clusters of calcifications were marked on the mammograms with a grease pencil by a radiologist, and the transference of those class assignments to the digitized data has been done using a mouse to perform an extremely time-consuming labeling of those abnormalities on the monitor. The labeling on the monitor is then reviewed and verified by the radiologist. Work will begin this fall to perform a similar labeling of our dataset from the University of Virginia. Labeling can also incorporate other information such as biopsy results, as we propose to do. It is conceivable that as the biopsy data base grows, the design algorithms could succeed in producing codes that can distinguish between features such as microcalcification clusters that are visually identical, but which might be benign or malignant.

Our studies using pixel intensities as features (no signal decomposition) have shown the approach to to provide superior performance in terms of classification and compression to Kohonen's LVQ in the detection of lung nodules in CT scans, where a Bayes tree-searched VQ with posterior estimation produced a pixel block sensitivity and specificity of .856 and .970, respectively [85]. Preliminary results for digitized mammograms were reported by us in [82, 83], where the sensitivity and specificity were 41.2 and 92.6, respectively. The results are depicted in Figure 5. Although this is not good performance considered only as a classifier, it is promising for several reasons: 1) the performance is much better than that of the independent cascade design of quantizer and classifier, as seen by comparing (B) and (D) in Figure 5 with the gold standard (C); 2) the decision is based only on 2×2 pixel blocks and performance will improve with context or suitable signal decomposition; and (3) the point of the algorithm is only to highlight suspicious regions as an aid to radiologist viewing and screening for more sophisticated algorithms. Local probability of error or sensitivity and specificity or PVP can be improved by combining the implicit classification with hierarchical algorithms that take more context into account. An attractive facet of this approach

is that it automatically incorporates the true purpose of the image, detecting microcalcifications and masses, into the optimization algorithms used to design the codes. This overall optimization of compression including the application is capable of better performance than is separately cascading compression and classification algorithms for detecting pathology.

The algorithmic and theoretical development of the algorithm is proceeding with the support of the NSF Grant MIP-9311190.

4 Conclusions

We have modified the our basic validation protocol to the comparison of analog with digital and lossy compressed digital mammograms [43]. The protocol was described in the Methods section and was presented by PI Gray and Co-PI Olshen to the Digital Mammography Panel meeting at the FDA on 6 March 1995 for consideration for use in demonstrating substantial equivalence of film/screen mammography and full field digital mammography. We have acquired the image data base for the current experiment, we have made small pilot experiments with the protocol, and we are currently coding the image data base for the full clinical experiment. The clinical experiments will begin in early September 1995, as originally planned in the Statement of Work. Looking toward a future studies comparing analog and digital, both compressed and uncompressed, we have used traditional approximations to estimate the number of patient studies that will be required for definitive size and power, and the current experiments will provide initial data which will allow us to refine these estimates.

During the past year we have continued to explore alternative compression algorithms of possible use in digital radiography. These include multiresolution, combined wavelet and vector quantization, and finite state codes [62, 81, 88, 89, 90, 65].

During the second and final year of this grant we will complete the clinical experiment described in this report and the accompanying statistical analyses. We will perform an additional experiment in the spring and summer using high resolution monitors instead of film on the same database. Work will continue on compression and classification algorithm development and on refining our estimates of size and power and the number of patients required for future, definitive, studies.

5 Figures and Tables

6 benign mass
6 benign calcifications
6 malignant mass
6 malignant calcifications
3 malignant combination of mass & calcifications
3 benign combination of mass & calcifications
4 breast edema
4 malignant architectural distortion
3 malignant focal asymmetry
3 benign asymmetric density
15 normals

59 studies, with 4 views per study.
The data were scanned by a a Lumisys Lumiscan 150 with
12 bits per pixel and 50 micron spot size.

Films printed using a Kodak2180 X-ray film printer,
a 79 micron 12 bit greyscale printer which
writes with a laser diode of 680 nm bandwidth. (Film and technician
time donated by Kodak.

Table 1: Test Data Set: Current Experiment

4 benign mass
4 benign calcifications
4 malignant mass
4 malignant calcifications
2 malignant combination of mass & calcifications
2 benign combination of mass & calcifications
4 breast edema
4 malignant architectural distortion
4 malignant focal asymmetry
4 benign asymmetric density
4 normals

Table 2: Training Data Set: Current Experiments

ID number _____ Session number _____ Case number _____

Reader initials: _____
Mammograms were of (**Left, Right, Both**) breast(s).

Subjective rating for diagnostic quality (sharpness, contrast)?

(bad) 1 – 5 (good):	Left CC	Left MLO	Right CC	Right MLO

Breast Density : Left 1 2 3 4 Right 1 2 3 4

1) almost entirely fat 2) scattered fibroglandular densities 3) heterogeneously dense 4) extremely dense

Finding side: **Neither, Left, Right, Both**

Findings (detection):

Individual finding side: **Left, Right** Finding # _____ of _____

Projection in which finding is seen: **CC** **MLO** **CC and MLO**

Location:

1) UOQ	**5) 12:00**	**9) retroareolar**	**13) inner**
2) UIQ	**6) 3:00**	**10) central**	**14) upper**
3) LOQ	**7) 6:00**	**11) axillary tail**	**15) lower**
4) LIQ	**8) 9:00**	**12) outer**	**16) whole breast**

Finding type: (possible, definite)

1) mass	**5) architectural distortion**	**9) breast edema**
2) calcifications	**6) solitary dilated duct**	**10) other**
3) mass containing calcifications	**7) asymmetric breast tissue**	___
4) mass with surrounding calcs	**8) focal asymmetric density**	___

CC View	**MLO View**
Size: _____ cm long axis by _____ cm short axis	Size: _____ cm long axis by _____ cm short axis
Distance from the nipple: _____ cm	Distance from the nipple: _____ cm

Associated findings include: (p= possible, d= definite)

1) breast edema	**(p , d)**	**5) lymphadenopathy**	**(p , d)**	**9) multiple similar masses**	**(p , d)**
2) skin retraction	**(p , d)**	**6) trabecular thickening**	**(p , d)**	**10) dilated veins**	**(p , d)**
3) nipple retraction	**(p , d)**	**7) architectual distortion**	**(p , d)**	**11) asymmetric density**	**(p , d)**
4) skin thickening	**(p , d)**	**8) calcs associated with mass**	**(p , d)**		

Assessment: The finding is

(A) indeterminate, additional assessment needed
What? 1) spot mag 2) extra views 3) U/S 4) old films
What is your *best guess* as to the finding's 1–5 assessment? _____ or are you uncertain if the finding exists? *Y*

(1) (N) negative – return to screening

(2) (B) benign (also negative but with benign findings) – return to screening

(3) (P) probably benign finding requiring 6-month followup

(4) (S) suspicion of malignancy (low), biopsy

(4) (S) suspicion of malignancy (moderate), biopsy

(4) (S) suspicion of malignancy (high), biopsy

(5) radiographic malignancy, biopsy

Comments: _____

Figure 1: Observer Form

Instructions to mammogram readers

You have been invited to participate in a reading of mammograms to detect breast abnormalities as seen on analog and digital studies. The study has been designed to simulate the clinical scenario as closely as possible. The films have been hung so that you will not be able to identify the patient names, and separate study numbers have been assigned to each patient for purposes of the study. A clear overlay has been taped to each film, but this should not interfere with your reading of the image. You may use a magnifying glass and you may use a bright light as you would ordinarily in clinical practice. The reading of the films is not timed.

A student will be assigned to you to prompt you for specific answers to questions on breast density, location, and suspicion of breast findings as stated on a questionnaire. You will also be asked to circle the abnormalities on the clear overlays with a grease or wax pencil and number them. You will also be asked to mark the location of the nipple on each film. Please be as specific as possible and follow these guidelines:

1. Please rate each mammogram for its sharpness and contrast as based on the technique of the year it was obtained. Rate each individual view for quality, e.g., "The right CC is good (5), and all the others are pretty good (4)." Note motion unsharpness in the comments.
2. Rate the right and left breast densities separately, for example the left breast could be rated as 1 and the right breast could be rated as 2.

Abnormalities:

1. Tell the student how many abnormalities are present in each breast, then describe each abnormality individually, e.g., "There are two lesions in the left breast. Lesion 1 of 2 is ...." The student will fill out extra forms when there are lesions in both breasts, or multiple lesions in one breast. The student will not re-fill out the ratings for diagnostic quality or breast density for each abnormality.
2. Circle all abnormalities, whether benign or malignant (i.e. circle fibroadenomas, fat necrosis, benign appearing clustered calcifications as well as malignant appearing calcifications). Please also note the location of the nipple by a grease or wax pencil mark on the clear overlay.
3. For each abnormality, rate it as a definite or possible abnormality. **Possible** abnormalities are those in which you are not sure that a lesion exists, for example, possible architectural distortion for which you would get additional views to confirm or exclude a lesion. **Definite** abnormalities are ones that are conclusively present, such as a mass or focal asymmetric density.
4. If you can only see an abnormality on one view, please circle it only on that view.
5. Circle spiculated masses such that you include the body of the mass but not its tiny extensions. For architectural distortion that may not have a central mass, include the spiculations.
6. Note and encircle architectural distortion, even when you think it is due to post-biopsy change and include the spiculations in your outline.
7. If you are unsure whether an apparent lesion exists, encircle it and judge the assessment as 'A' (assessment incomplete), and note your uncertainty by circling the *Y*. Here extra views are needed to confirm or exclude the presence of the abnormality.
8. If you are sure an apparent lesion exists and is a true mass, calcification, calcification cluster, or other finding, but the assmenent is 'A' because ultrasound or extra views are needed to evaluate mass borders or calcifications shapes, or to determine if the finding is a cyst, please mark down your BEST GUESS as to whether the lesion is benign or malignant using the ACR lexicon codes.
9. If the lesion has a differential, such as post-biopsy change vs. cancer, or cyst, fibroadenoma or well-circumscribed cancer, and you would like to note it, please do so in the comments section.

Thank you for your participation in this study. If you have questions or comments, please direct them to Debra M. Ikeda, M.D. at (415) 723-7672.

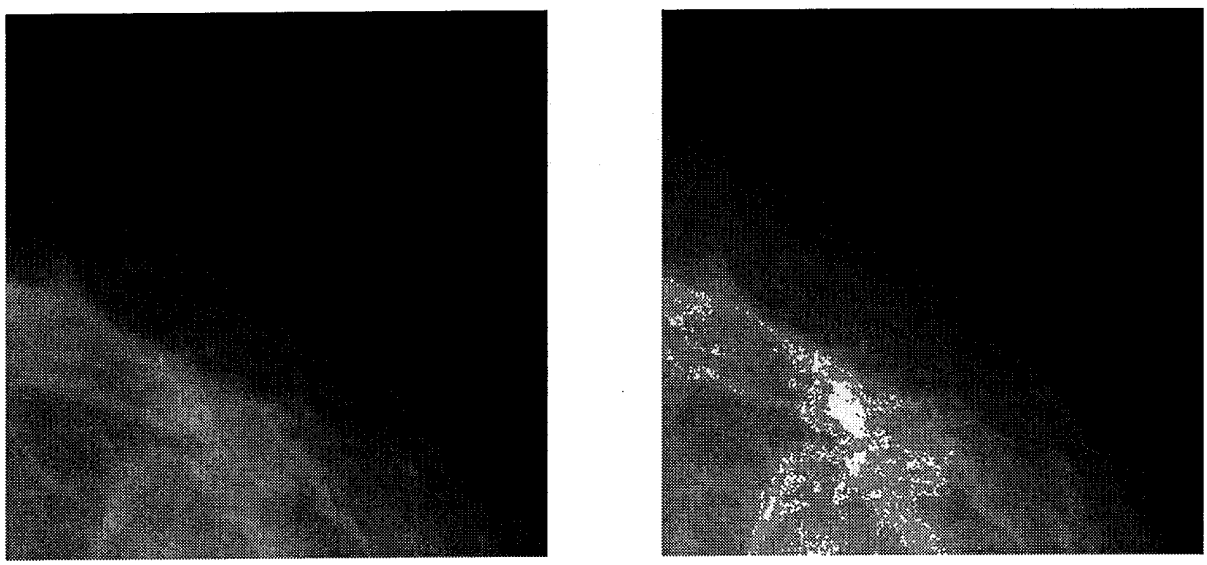
Figure 2: Observer Form Instructions

	routine f/u	further study
routine f/u		
further study		

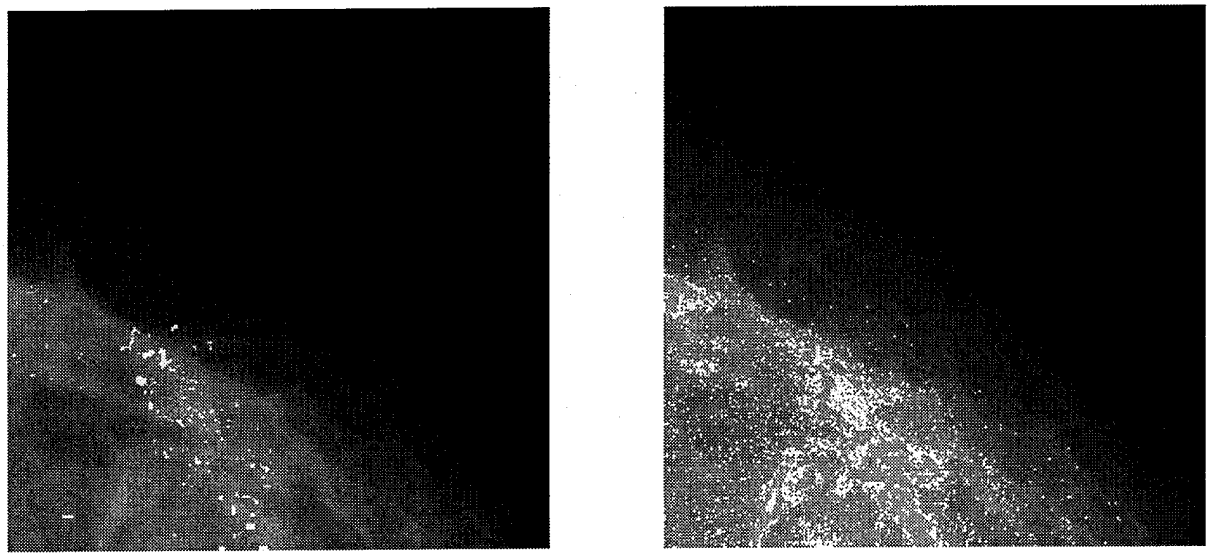
Figure 3: Management Outcomes

	Right	Wrong	
Right	$2\psi + h - 1 + \gamma$	$1 - \psi - h - \gamma$	ψ
Wrong	$1 - \psi - \gamma$	γ	$1 - \psi$
	$\psi + h$	$1 - \psi - h$	

Figure 4: Management Outcome Probabilities



(A) (B)



(C) (D)

Figure 5: Compression and classification of digitized mammograms at 2 bpp for calcifications: (A) Portion of Compressed Mammogram using BTSVQ with posterior estimation (B) Compressed/Classified image using BTSVQ with posterior estimation (white highlighted areas denote pixel blocks classified as microcalcifications) (C) Original 12 bit image with microcalcifications highlighted in white (D) Compressed/Classified image using independent TSVQ design (white highlighted pixel areas denote pixel blocks classified as microcalcification)

6 Bibliography

References

[1] I. Andersson, "Mammography in clinical practice," *Med Radiography and Photography*, vol. 62, no. 2, p. 2, 1986.

[2] P. Stomper, J. Connolly, J. Meyer, and J. Harris, "Clinically occult ductal carcinoma in situ detected with mammography: analysis of 100 cases with radiographic-pathologic correlation," *Radiology*, vol. 172, pp. 235–241, 1989.

[3] D. Dershaw, A. Abramson, and D. Kinne, "Ductal carcinoma in situ: mammographic findings and clinical implications," *Radiology*, vol. 170, pp. 411–415, 1989.

[4] D. Ikeda, I. Andersson, C. W. rd L. Janzon, and F. Linell, "Radiographic appearance and prognostic consideration of interval carcinoma in the Malmö mammographic screening trial," *Amer. J. Roentgenology*, vol. 159, pp. 287–294, 1992.

[5] J. Martin, M. Moskowitz, and J. Milbrath, "Breast cancer missed by mammography," *Amer. J. Roentgenology*, vol. 132, pp. 737–739, 1979.

[6] D. Ikeda and I. Andersson, "Atypical mammographic presentations of ductal carcinoma in situ," *Radiology*, vol. 172, pp. 661–666, 1989.

[7] J. Frisell, G. Eklund, L. Hellstrom, and A. Somell, "Analysis of interval breast carcinomas in a randomized screening trial in Stockholm," *Breast Cancer Res Trea*, vol. 9, pp. 219–225, 1987.

[8] W. Kegelmeyer, "Software for image analysis aids in breast cancer detection," *OE Reports*, p. 7, February 1993.

[9] W. P. Kegelmeyer, Jr., "Evaluation of stellate lesion detection in a standard mammogram data set," in *Proceedings IS&T/SPIE Annual Symposium on Electronic Imaging, Science & Technology*, (San Jose, CA), Jan–Feb 1993.

[10] K. Woods, J. Solka, C. Priebe, C. Doss, K. Bowyer, and L. Clarke, "Comparative evaluation of pattern recognition techniques for detection of microcalcifications," in *Proceedings IS& T/SPIE Annual Symposium on Electronic Imaging, Science & Technology*, (San Jose, CA), January–February 1993.

[11] L. Clarke, G. Blaine, K. Doi, M. Yaffe, F. Shtern, and G. Brown, "Digital mammography, cancer screening: Factors important for image compression," in *Proceedings Space and Earth Science Data Compression Workshop*, (Snowbird, Utah), NASA, April 1993.

[12] J. Storer, *Data Compression.* Rockville, MD: Computer Science Press, 1988.

[13] J. Sayre, D. R. Aberle, M. I. Boechat, T. R. Hall, H. K. Huang, B. K. Ho, P. Kashfian, and G. Rahbar, "Effect of data compression on diagnostic accuracy in digital hand and chest radiography," in *Proceedings of Medical Imaging VI: Image Capture, Formatting, and Display*, vol. 1653, pp. 232–240, SPIE, Feb. 1992.

[14] H. MacMahon, K. Doi, S. Sanada, S. Montner, M. Giger, C. Metz, N. Nakamori, F. Yin, X. Xu, H. Yonekawa, and H. Takeuchi, "Data compression: effect on diagnostic accuracy in digital chest radiographs," *Radiology*, vol. 178, pp. 175–179, 1991.

[15] H. Lee, A. H. Rowberg, M. S. Frank, H. S. Choi, and Y. Kim, "Subjective evaluation of compressed image quality," in *Proceedings of Medical Imaging VI: Image Capture, Formatting, and Display*, vol. 1653, pp. 241–251, SPIE, Feb. 1992.

[16] T. Ishigaki, S. Sakuma, M. Ikeda, Y. Itoh, M. Suzuki, and S. Iwai, "Clinical evaluation of irreversible image compression: Analysis of chest imaging with computed radiography," *Radiology*, vol. 175, pp. 739–743, 1990.

[17] K. Chan, S. Lou, and H. Huang, "Full-frame transform compression of CT and MR images," *Radiology*, vol. 171, no. 3, pp. 847–851, 1989.

[18] G. Wallace, "The JPEG still picture compression standard," *Communications of the ACM*, vol. 34, pp. 30–44, April 1991.

[19] M. Rabbani and P. W. Jones, *Digital Image Compression Techniques*, vol. TT7 of *Tutorial Texts in Optical Engineering*. Bellingham, WA: SPIE Optical Engineering Press, 1991.

[20] H. Blume, "ACR-NEMA Digital Imaging and Communication Standard Committee, Working Group # 4, MEDPACS Section, Data Compression Standard # PS2." 1989.

[21] J. W. Woods, ed., *Subband Image Coding*. Boston: Kluwer Academic Publishers, 1991.

[22] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Processing*, vol. 1, pp. 205–220, April 1992.

[23] T. Senoo and B. Girod, "Vector quantization for entropy coding of image subbands," *IEEE Transactions on Image Processing*, vol. 1, pp. 526–532, October 1992.

[24] A. S. Lewis and G. Knowles, "Image compression using the 2-D wavelet transform," *IEEE Trans. Image Processing*, vol. 1, pp. 244–250, April 1992.

[25] M.Vetterli and J.Kovacevic, *Wavelets and Subband Coding*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[26] P. Wilhelm, D. R. Haynor, Y. Kim, and E. A. Riskin, "Lossy image compression for digital medical imaging system," *Optical Engineering*, vol. 30, pp. 1479–1485, Oct. 1991.

[27] H. H. Barrett, T. Gooley, K. Girodias, J. Rolland, T. White, and J. Yao, "Linear discriminants and image quality," in *Proceedings of the 1991 International Conference on Information Processing in Medical Imaging (IPMI '91)*, (Wye, United Kingdom), pp. 458–473, Springer-Verlag, July 1991.

[28] J. Bramble, L. Cook, M. Murphey, N. Martin, W. Anderson, and K. Hensley, "Image data compression in magnification hand radiographs," *Radiology*, vol. 170, pp. 133–136, 1989.

[29] M. Goldberg, M. Pivovarov, W. Mayo-Smith, M. Bhalla, J. Blickman, R. Bramson, G. Boland, H. Llewellyn, and E. Halpem, "Application of wavelet compression to digitized radiographs," *American Journal of Radiology*, vol. 163, pp. 463–468, August 1994.

[30] C. E. Metz, "Basic principles of ROC analysis," *Seminars in Nuclear Medicine*, vol. VIII, pp. 282–298, Oct. 1978.

[31] J. A. Swets, "ROC analysis applied to the evaluation of medical imaging techniques," *Investigative Radiology*, vol. 14, pp. 109–121, March–April 1979.

[32] J. Hanley and B. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC)curve," *Diagnostic Radiology*, vol. 143, pp. 29–36, 1982.

[33] B. McNeil and J. Hanley, "Statistical approaches to the analysis of receiver operating characteristic (ROC) curves," *Medical Decision Making*, vol. 4, pp. 137–150, 1984.

[34] D. Chakraborty and L. Winter, "Free-response methodology: alternate analysis and a new observer-performance experiment," *Radiology*, vol. 174, no. 3, pp. 873–881, 1990.

[35] H. C. Davidson, C. J. Bergin, C. Tseng, P. C. Cosman, L. E. Moses, R. A. Olshen, and R. M. Gray, "The effect of lossy compression on diagnostic accuracy of thoracic CT images." Presented at the 77th Scientific Assembly of the Radiological Society of North America, Chicago, Illinois, Dec. 1991.

[36] R. A. Olshen, P. C. Cosman, C. Tseng, C. Davidson, L. Moses, R. M. Gray, and C. Bergin, "Evaluating compressed medical images," in *Proceedings COMCON III*, (Victoria, B.C., Canada), pp. 830–840, Oct. 1991.

[37] P. C. Cosman, H. C. Davidson, C. J. Bergin, C. Tseng, , L. E. Moses, R. A. Olshen, and R. M. Gray, "The effect of lossy compression on diagnostic accuracy of thoracic CT images," *Radiology*, vol. 190, no. 2, pp. 517–524, 1994.

[38] P. Cosman, C. Tseng, R. Gray, R. Olshen, L. E. Moses, H. C. Davidson, C. Bergin, and E. Riskin, "Tree-structured vector quantization of CT chest scans: image quality and diagnostic accuracy," *IEEE Trans. Medical Imaging*, vol. 12, pp. 727–739, Dec. 1993.

[39] C. Tseng, S. Perlmutter, P. Cosman, K. Li, C. Bergin, R. Olshen, and R.M.Gray, "Effect of tree-structured vector quantization on the accuracy of vessel measurements in MR chest scans." Submitted for possible publication.

[40] P. Cosman, H. C. Davidson, C. Bergin, C. Tseng, L. E. Moses, E. Riskin, R. Olshen, and R. Gray, "Thoracic CT images: effect of lossy image compression on diagnostic accuracy," *Radiology*, vol. 190, pp. 517–524, 1994.

[41] S. Perlmutter, C. Tseng, P. Cosman, K. . Li, R. Olshen, and R. Gray, "Measurement accuracy as a measure of image quality in compressed mr chest scans," in *Proceedings of the IEEE 1994 International Conference on Image Processing*, vol. 1, (Austin, Texas), pp. 861—8, November 1994.

[42] P. Cosman, R. Gray, and R. Olshen, "Evaluating quality of compressed medical images: SNR, subjective rating, and diagnostic accuracy," *Proceedings of the IEEE*, vol. 82, pp. 919–932, June 1994.

[43] R. Gray, R. Olshen, D. Ikeda, P. Cosman, S. Perlmutter, C. Nash, and K. Perlmutter, "Full breast digital imaging (fbdi) as a screening device: Summary of clinical protocol," March 1995. Report submitted to Panel on Digital Mammography, FDA, 6 March 1995.

[44] B. Efron, "Better bootstrap confidence intervals and bootstrap approximations," *J. Amer. Stat. Assoc.*, vol. 82, pp. 171–185, 1987.

[45] J. Cohen, "Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit," *Psychological Bulletin*, vol. 20, pp. 213–220, 1968.

[46] Y. Bishop, S. Feinberg, and P. Holland, *Discrete Multivariate Analysis*. Cambridge, Mass: MIT Press, 1975.

[47] A. Garber, R. Olshen, H. Zhang, and E. Venkatraman, "Predicting high-risk cholesterol levels," *International Statistical Review*, vol. 62, no. 2, pp. 203–228, 1994.

[48] P. Cosman, R. Gray, and M. Vetterli, "Vector quantization of image subbands: A survey," *IEEE Trans. on Image Processing*, 1995. to appear.

[49] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Boston: Kluwer Academic Publishers, 1992.

[50] R. Coifman and M. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 713–718, 1992.

[51] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate-distortion sense," *IEEE Trans. Image Processing*, vol. 2, pp. 160–176, April 1993.

[52] Z. Gao, F. Chen, B. Belzer, and J. Villasenor, "A comparison of the Z, E_8, and Leech lattices for image subband quantization," in *Proceedings of the 1995 IEEE Data Compression Conference* (J. Storer and M. Cohn, eds.), (Snowbird, Utah), pp. 312–321, IEEE Computer Society Press, March 1995.

[53] J. Villasenor, B. Belzer, and J. Liao, "Wavelet filter evaluation for image compression," *IEEE Transactions on Image Processing*, vol. 4, pp. 1053–1060, August 1995.

[54] M. Vishwanath and P. Chou, "An efficient algorithm for hierarchical compression of video," in *Proceedings ICIP-94*, vol. III, (Austin, Texas), pp. 275–279, IEEE, IEEE Computer Society Press, November 1994.

[55] N. Chaddha, M. Vishwanath, and P. A. Chou, "Hierarchical vector quantization of perceptually weighted block transforms," in *Proceedings of the 1995 IEEE Data Compression Conference (DCC)* (J. A. Storer and M. Cohn, eds.), (Snowbird, Utah), pp. 3–12, IEEE Computer Society Press, March 1995.

[56] H. Abut, ed., *Vector Quantization*. IEEE Reprint Collection, Piscataway, NJ: IEEE Press, May 1990.

[57] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Belmont, CA: Wadsworth, 1984.

[58] A. Said and W. Pearlman, "A new fast and efficient image codec based on set partitioning in hierarchical trees," 1995. submitted for possible publication.

[59] J. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Transactions on Signal Processing*, vol. 41, pp. 3445–3462, December 1993.

[60] A. B. Watson, "Visually optimal DCT quantization matrices for individual images," in *DCC*, pp. 178–187, 1993.

[61] H. A. Peterson, H. Peng, J. H. Morgan, and W. B. Pennebaker, "Quantization of color image components in the DCT domain," in *SPIE*, pp. 210–222, 1991.

[62] P. Cosman, S. Perlmutter, and K. Perlmutter, "Tree-structured vector quantization with significance map for wavelet image coding," in *Proceedings of the 1995 IEEE Data Compression Conference (DCC)* (J. Storer and M. Cohn, eds.), (Snowbird, Utah), IEEE Computer Society Press, March 1995.

[63] M. Effros, P. Chou, and R. Gray, "Variable dimension weighted universal vector quantization and noiseless coding," in *Proceedings Data Compression Conference (DCC)* (J. Storer and M. Cohn, eds.), pp. 1–11, IEEE Computer Society Press, March 1994.

[64] M. Effros, P. Chou, and R. Gray, "One-pass adaptive universal vector quantization," in *Proceedings ICASSP*, (Adelaide, Austrailia), IEEE Press, April 1994.

[65] S. M. Perlmutter, K. O. Perlmutter, and P. C. Cosman, "Vector quantization with zerotree significance map for wavelet image coding," in *Proc. Twenty-ninth Asilomar Conference on Signals, Systems and Computers*, (Pacific Grove, Calif.), October 1995. to appear.

[66] S. Astley, I. Hutt, S. Adamson, P. Rose, P. Miller, C. Boggis, C. Taylor, T. Valentine, and J. Davies, "Automation in mammography: computer vision and human perception," in *Proceedings IS&T/SPIE Annual Symposium on Electronic Imaging, Science & Technology*, (San Jose, CA), Jan–Feb 1993.

[67] I. N. Bankman, W. A. Christens-Barry, D. W. Kim, I. N. Weinberg, O. B. Gatewood, and W. R. Brody, "Automated recognition of microcalcification clusters in mammograms," in *Proceedings IS&T/SPIE Annual Symposium on Electronic Imaging, Science & Technology*, (San Jose, CA), Jan–Feb 1993.

[68] H. Barman and G. Granlund, "Hierarchical feature extraction for computer-aided analysis of mammograms," in *Proceedings IS&T/SPIE Annual Symposium on Electronic Imaging, Science & Technology*, (San Jose, CA), Jan–Feb 1993.

[69] D. Brzakovic, P. Brzakovic, and M. Neskovic, "An approach to automated screening of mammograms," in *Proceedings IS&T/SPIE Annual Symposium on Electronic Imaging, Science & Technology*, (San Jose, CA), Jan–Feb 1993.

[70] A. P. Dhawan, Y. Chitre, and M. Moskowitz, "Artificial neural network based classification of mammographic microcalcifications using image structure features," in *Proceedings IS&T/SPIE Annual Symposium on Electronic Imaging, Science & Technology*, (San Jose, CA), Jan–Feb 1993.

[71] N. Karssemeijer, "Recognition of clustered microcalcifications using a random field model," in *Proceedings IS&T/SPIE Annual Symposium on Electronic Imaging, Science & Technology*, (San Jose, CA), Jan–Feb 1993.

[72] R. M. Nishikawa, M. L. Giger, K. Doi, F.-F. Yin, Y. Wu, Y. Jiang, W. Zhang, Z. Huo, U. Bick, C. J. Vyborny, and R. A. Schmidt, "Computer-aided detection and diagnosis of masses and clustered microcalcifications from digital mammograms," in *Proceedings IS&T/SPIE Annual Symposium on Electronic Imaging, Science & Technology*, (San Jose, CA), Jan–Feb 1993.

[73] J. Parker, L. J. Yeoman, D. R. Dance, D. H. Davies, M. J. Michell, and S. Humphreys, "Classification of ductal carcinoma in-situ by image analysis of calcifications from digital mammograms," in *Proceedings IS&T/SPIE Annual Symposium on Electronic Imaging, Science & Technology*, (San Jose, CA), Jan–Feb 1993.

[74] W. Qian, L. P. Clarke, M. Kallergi, H. D. Li, R. Velthuizen, R. A. Clark, M. L. Silbiger, and H. Lee, "Tree-structured nonlinear filter and wavelet transform for microcalcification segmentation in mammography," in *Proceedings IS&T/SPIE Annual Symposium on Electronic Imaging, Science & Technology*, (San Jose, CA), Jan–Feb 1993.

[75] L. Shen, R. M. Rangayyan, and J. E. L. Desautels, "An automatic detection and classification system for calcifications in mammograms," in *Proceedings IS&T/SPIE Annual Symposium on Electronic Imaging, Science & Technology*, (San Jose, CA), Jan–Feb 1993.

[76] D. Zhao, "Morphological size filters for feature extraction of mammographic images," in *Proceedings IS&T/SPIE Annual Symposium on Electronic Imaging, Science & Technology*, (San Jose, CA), Jan–Feb 1993.

[77] K. Oehler and R. Gray, "Combining image classification and image compression using vector quantization," in *Proceedings of the 1993 IEEE Data Compression Conference (DCC)* (J. Storer and M. Cohn, eds.), (Snowbird, Utah), pp. 2–11, IEEE Computer Society Press, March 1993.

[78] K. Oehler, *Image compression and classification using vector quantization.* PhD thesis, Stanford University, Stanford, CA, August 1993.

[79] P. Cosman, K. Oehler, E. Riskin, and R. Gray, "Using vector quantization for image processing," *Proc. IEEE*, vol. 81, pp. 1326–1341, Sep. 1993.

[80] R. Gray, K. Oehler, K. Perlmutter, and R. A. Olshen, "Combining tree-structured vector quantization with classification and regression trees," in *Proceedings of the Twenty-Seventh Annual Asilomar Conference on Signals, Systems and Computers* (R. Madon, ed.), (Pacific Grove, CA), November 1993.

[81] K. O. Perlmutter, M. Effros, S. M. Perlmutter, and R. M. Gray, "An iterative joint codebook and classifier improvement algorithm for finite-state vector quantization," in *Proc. Twenty-eighth Asilomar Conference on Signals, Systems and Computers*, (Pacific Grove, CA), 31 October – 2 November 1994.

[82] K. O. Perlmutter, C. L. Nash, and R. M. Gray, "A comparison of bayes risk weighted vector quantization with posterior estimation with other VQ-based classifiers," in *IEEE International Conference on Image Processing*, vol. 2, (Austin, TX), pp. 217–221, Nov. 1994.

[83] C. Nash, K. Perlmutter, and R. Gray, "Evaluation of Bayes risk weighted vector quantization with posterior estimation in the detection of lesions in digitized mammograms," in *Conference Record of the Twenty-Eighth Asilomar Converence on Signals, Systems, & Computers*, vol. 1, pp. 716–720, IEEE Computer Science Press, October 1995.

[84] K. L. Oehler and R. M. Gray, "Combining image compression and classification using vector quantization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, pp. 461–473, May 1995.

[85] K. O. Perlmutter, S. Perlmutter, R. M. Gray, R. A. Olshen, and K. L. Oehler, "Bayes risk weighted vector quantization with posterior estimation for image compression and classification," *IEEE Trans. Image Processing*, 1995. to appear.

[86] K. O. Perlmutter, R. M. Gray, R. A. Olshen, and S. M. Perlmutter, "Bayes risk weighted vector quantization with cart estimated posteriors," in *Proceedings ICASSP*, (Detroit, Mich.), IEEE, May 1995.

[87] K. O. Perlmutter, N. Chaddha, J. B. Buckheit, R. M. Gray, and R. A. Olshen, "Text segmentation in mixed-mode images using classification trees and transform tree-structured vector quantization." Submitted to ICASSP, 1995.

[88] S. M. Perlmutter and R. M. Gray, "A finite-state multiresolution approach to image compression using pruned nested tree-structured vector quantization," in *Proc. Twenty-eighth Asilomar Conference on Signals, Systems and Computers*, (Pacific Grove, CA), 31 October – 2 November 1994.

[89] S. M. Perlmutter and R. M. Gray, "A low complexity multiresolution approach to image compression using pruned nested tree-structured vector quantization," in *IEEE International Conference on Image Processing*, vol. 1, (Austin, TX), pp. 588–592, Nov. 1994.

[90] K. Perlmutter, W. Tchoi, S. Perlmutter, and P. C. Cosman, "Wavelet/vq image coding with a recursive segmentation structure," in *Proc. Twenty-ninth Asilomar Conference on Signals, Systems and Computers*, (Pacific Grove, Calif.), October 1995. to appear.