AD_____

GRANT NO: DAMD17-93-J-3021

TITLE: DEVELOPMENT OF METHODS FOR COMPUTER-ASSISTED INTERPRETATION OF DIGITAL MAMMOGRAMS FOR EARLY BREAST CANCER DETECTION

PRINCIPAL INVESTIGATOR: Maryellen L. Giger, Ph.D.

CONTRACTING ORGANIZATION: University of Chicago
5801 South Ellis Avenue
Chicago, Illinois 60637

REPORT DATE: March 27, 1995

TYPE OF REPORT: Annual Report

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick
Frederick, 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
	March 27, 1995	Annual Report (1 Mar 94-28 Feb 95)

4. TITLE AND SUBTITLE
Development of Methods for Computer-Assisted Interpretation of Digital Mammograms for Early Breast Cancer Detection

5. FUNDING NUMBERS
DAMD17-93-J-3021

6. AUTHOR(S)
Maryellen L. Giger, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Chicago
5801 South Ellis Avenue
Chicago, Illinois 60637

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick
Frederick, 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 goal of the research is to develop a computer-vision module as an aid to radiologists The specific aims: (1) Further development of advanced computerized schemes for the detection and classification of masses and microcalcifications in digital mammograms, including quantitative analysis of the radiographic characteristics and the decision-maki processes used by radiologists in making a decision with respect to the likelihood of malignancy and choosing the appropriate course of action. (2) Development of a dedicated "intelligent" modular system with man-machine interfaces and fast computation times appropriate for the effective use of the computer-vision schemes. (3) Evaluation of the efficacy and efficiency of the module using a large clinical database. The significance of this research is that if the detectability of cancers can be increased by employing a computer to aid the radiologist's diagnosis, then the treatment of patients with cancer can be initiated earlier and their chance of survival improved. Systematic introduction of computer-vision tools to radiologists that is presented in this proposal requires minimal modification to the current reading habits of radiologists. When digital mammographic imaging units become commonplace, the computer-vision module can be interfaced to electronic, filmless medical imaging reading areas.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Computer-aided diagnosis, computer vision, mammography	43
	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

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.

____ 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.

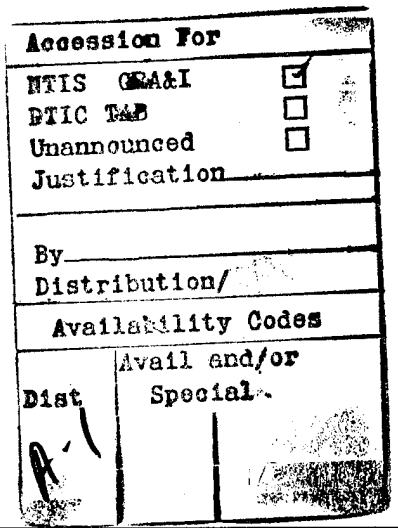
____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

____ 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.



Maryellen Segin 3-26-95

PI - Signature Date

Table of Contents

INTRODUCTION

Nature of the problem

Breast cancer is a leading cause of death in women, causing an estimated 44,000 deaths per year (1). Mammography is the most effective method for the early detection of breast cancer (2-5) and it has been shown that periodic screening of asymptomatic women does reduce mortality (6-11). Various medical organizations have recommended the use of mammographic screening for the early detection of breast cancer (3). Thus, mammography is becoming one of the largest volume x-ray procedures routinely interpreted by radiologists.

It has been reported that between 30 to 50% of breast carcinomas detected mammographically demonstrate clusters of microcalcifications (12-14), although about 80% of breast carcinomas reveal microcalcifications upon microscopic examination (15-18). In addition, studies indicate that 26% of nonpalpable cancers present mammographically as a mass while 18% present both with a mass and microcalcifications (19). Although mammography is currently the best method for the detection of breast cancer, between 10-30% of women who have breast cancer and undergo mammography have negative mammograms (20-24). In approximately two-thirds of these false-negative mammograms, the radiologist failed to detect the cancer that was evident retrospectively (23-26). Low conspicuity of the lesion, eye fatigue and inattentiveness are possible causes for these misses. We believe that the effectiveness (early detection) and efficiency (rapid diagnosis) of screening procedures could be increased substantially by use of a computer system that successfully aids the radiologist by indicating locations of suspicious abnormalities in mammograms.

Many breast cancers are detected and referred for surgical biopsy on the basis of a radiographically detected mass lesion or cluster of microcalcifications. Although general rules for the differentiation between benign and malignant breast lesions exist (20,27), considerable misclassification of lesions occurs with the current methods. On average, only 10-30% of masses referred for surgical breast biopsy are actually malignant (20,28). Surgical biopsy is an invasive technique that is an expensive and traumatic experience for the patient and leaves physical scars that may hinder later diagnoses (to the

extent of requiring repeat biopsies for a radiographic tumor-simulating scar). A computerized method capable of detecting and analyzing the characteristics of benign and malignant masses, in an objective manner, should aid radiologists by reducing the numbers of false-positive diagnoses of malignancies, thereby decreasing patient morbidity as well as the number of surgical biopsies performed and their associated complications.

The development of computer methods to assist radiologists is a timely project in the sense that digital radiography is on the threshold of widespread clinical use. The arrival of digital radiographic systems allows for the acquisition of image data in a format accessible to computerized schemes. The potential significance of this research project lies in the fact that if the detectability of cancers can be increased by employing a computer to aid the radiologist's diagnosis, then the treatment of patients with cancer can be initiated earlier and their chance of survival improved.

The systematic and gradual introduction of computer-assisted interpretation to radiologists that is presented in this proposal is very important in that it allows for a mode of presentation with minimum modification to the current reading habits of radiologists and does not require a "digital" department in which reading must be done from a CRT screen. These two issues are of concern since (1) some radiologists are not comfortable with computer-based methods and (2) primary diagnosis from a CRT display is still controversial. However, the introduction of computer vision to radiologists presented in this proposal is not affected by either concern. In addition, when filmless image acquisition and/or digital (PACS) radiology departments are commonplace in the future, the computer-vision module can be immediately interfaced to electronic, filmless imaging and reading areas.

Background of previous work

In the 1960's and 70's, several investigators attempted to analyze mammographic abnormalities with computers. Winsberg et al. (29), in an early study, examined areas of increased density in contralateral breasts. They felt that their results demonstrated the feasibility for future computer interpretation of mammograms. Spiesberger (30) developed various feature-extraction techniques and a two-view verification method involving medio-lateral oblique and cranio-caudal views to detect

microcalcifications. Kimme et al. (31) developed a computerized method for the detection of suspicious abnormalities in mammograms based on the statistical measures of textural features. They tested their algorithm on 7 patient cases. A similar approach using texture analysis and bilateral comparison was also employed by Hand et al (32) and Semmlow (33) in the computerized localization of suspicious abnormal areas of breasts. Their results yielded a 66% true-positive rate with approximately 26 false suspicious areas per image. With regard to classification methods, Ackerman et al. (34), using digital xeroradiographs, devised four measures of malignancy: calcification, spiculation, roughness and shape, to perform classification on specific areas selected by human observers. The authors viewed their research as only a small step toward the automated reading of xeroradiographs and appeared to discontinue prematurely their computer vision work. The same group (35) did, however, attempt to improve diagnosis by using 36 radiographic properties which were evaluated semi-quantitatively by a radiologist for input to a computer decision tree. Wee et al. (36) and Fox et al. (37) performed preliminary studies on the classification of microcalcifications. These previous studies demonstrated the potential capability of using a computer in the detection of mammographic abnormalities. Their results, however, yielded a large number of false-positives and were based on small data sets.

Computer-aided diagnosis, in general, has attracted little attention during the last decade, perhaps due to the inconvenience involved in obtaining a radiograph in digital format. Recent work, though, shows a promising future. Magnin et al. (38) and Caldwell (39) used texture analysis to evaluate the breast's parenchymal pattern as an indicator of cancer risk. These preliminary studies raised many unanswered questions regarding topics ranging from the digital recording process to the type of numerical risk coefficient employed. Thus, further studies using texture analysis are indicated. The work by Fam and Olson (40,41) on the computer analysis of mammograms is encouraging; however, their method has only been tested on 20 mammographic regions of interest (each roughly half a mammogram). Davies and Dance (42) have reported on their automatic method for the detection of clustered calcifications using local gray-level thresholding and also a clustering rule. Their results yielded a true-positive rate of 96%; however, no indications of the subtlety and size of the calcifications were given. Astley et al. (43), Grimaud et al. (44) and Jin et al. (45) recently reported on their methods

for the detection of breast lesions. Karssemeijer (46) has described a stochastic method based on Bayesian decision theory that appears promising. Lai et al. (47) and Brzakovic et al. (48) are also developing techniques for the detection of mass lesions. The actual performance level and difficulty of the databases, however, are unknown. Gale et al. (49) and Getty et al. (50) are both developing computer-based classifiers, which take as input diagnostically-relevant features obtained from radiologists' readings of breast images. Getty et al. found that with the aid of the classifier, community radiologists performed as well as unaided expert mammographers in making benign-malignant decisions. Swett et al. (51,52) are developing an expert system to provide visual and cognitive feedback to the radiologist using a critiquing approach combined with an expert system. The system has been demonstrated, though not tested.

We in the Kurt Rossmann Laboratories for Radiologic Image Research at The University of Chicago have vast experience in developing various computer-aided diagnosis (CAD) methods in mammography, chest radiography, and angiography (53-66). We believe that our CAD methods in digital mammography, which include the computerized detection of microcalcifications and masses, have achieved levels of sensitivity and specificity that warrant testing in a clinical environment.

Our detection scheme for clustered microcalcifications includes a preprocessing step referred to as a difference-image approach (53,54). Basically, the original digital mammogram is spatially filtered twice: once to enhance the signal-to-noise ratios of the microcalcifications and a second time to suppress them. The difference between the two resulting processed images yields an image (a difference image) in which the variations in background density are largely removed. Microcalcifications are then segmented from the difference image using global gray-level thresholding and local thresholding techniques. The segmented image is next subjected to feature-extraction techniques in order to remove signals that likely arise from structures other than microcalcifications. An area filter (56), based on mathematical morphology, is used to eliminate small features. Next, each region of interest that contains remaining features is subjected to low-frequency background correction and is characterized by the first moment of its power spectrum, defined as the weighted average of radial spatial frequency over the two-dimensional power spectrum (55). A clustering filter

(57) is next used so that only clusters that contain more than a preselected number of signals within a region of preselected size are retained by the computer. The computerized scheme, using 78 mammograms (39 normal and 39 abnormals) in which most clusters were quite subtle, the scheme yielded a sensitivity of 85% with approximately 2.5 false-positive detections per image (58).

The computerized scheme for detection of clustered microcalcifications (55) developed at The University of Chicago has been tested as an aid to radiologic diagnosis. Using a database of 60 clinical mammograms, half of which contained subtle clusters of microcalcifications, a human observer study was conducted in order to examine the effect of the computer-vision aid on radiologists' performance in a situation that simulated rapid interpretation of screening mammograms. The computer scheme attained an 87% true-positive detection rate with an average of four false-positive clusters per image. The effect of the number of false-positive detections on radiologist performance was also examined by simulating a computer performance level of 87% sensitivity with one false-positive detection per image. Radiologist detection performance was evaluated using ROC (receiver operating characteristic) methodology (68). It was found from the ROC analysis that there was a statistically significant improvement in the radiologists' accuracy when they were given the computer-generated diagnostic information (at either false-positive level), compared with their accuracy obtained without the computer output.

Our scheme for the detection of mammographic masses is based on deviations from the architectural symmetry of normal right and left breasts, with asymmetries indicating potential masses (60,61). The input to the computerized scheme, for a given patient, are the four conventional mammograms obtained in a routine screening examination: the right cranio-caudal (CC) view, the left CC view, the right medio-lateral-oblique (MLO) view, and the left MLO view. After automatic registration of corresponding left and right breast images, a nonlinear subtraction technique is employed in which gray-level thresholding is performed on the individual mammograms prior to subtraction. Ten images thresholded with different cutoff gray levels are obtained from the right breast image, and ten are obtained from the left breast image. Next, subtraction of the corresponding right and left breast images is performed to generate ten bilateral-subtraction images. Run-length

analysis is then used to link the data in the various subtracted images. This linking process accumulates the information from a set of 10 subtraction images into two images that contain locations of suspected masses for the left and right breasts. Next, feature-extraction techniques, which include morphological filtering and analysis of size, shape and distance from border, are used to reduce the number of false-positive detections. Currently, using 150 pairs of clinical mammograms (from 75 cases), the approach achieves a true-positive detection rate of approximately 85% with 3 to 4 false-positive detections per image (62).

We have also investigated the application of artificial neural networks to the detection and classification of mammographic lesions. We used an artificial neural network (ANN) to extract microcalcification image data from digital mammograms (59). The ANN, which was supplied with the power spectra of remaining suspected regions (from the CAD scheme) as input, distinguished actual clustered microcalcifications from false-positive regions and was able to eliminate many of the false positives. Also, we are applying ANNs to the decision-making task in mammography (63). Three-layer, feed-forward neural networks with a back-propagation algorithm were trained for the interpretation of mammograms based on features extracted from mammograms by experienced radiologists. The database for input to the ANN consisted of features extracted from 133 textbook cases and 60 clinical cases. Performance of the ANN was evaluated by ROC analysis. In tests, using 43 initial image features (related to masses, microcalcifications and secondary abnormalities) that were later reduced to 14 features, the performance of the neural network was found to be higher than the average performance of attending and resident radiologists in classifying benign and malignant lesions. At an optimal threshold for the ANN output value, the ANN achieved a classification sensitivity of 100% for malignant cases with a false-positive rate of only 41%, whereas the average radiologist yielded a sensitivity of only 89% with a false-positive rate for classification of 60%.

We are also developing computer-aided methods for the interpretation of digital chest radiographs, such as in the detection of pulmonary nodules, interstitial infiltrates, pneumothorax and cardiomegaly (67,69-75). The computer-vision scheme for the detection of lung nodules is based on a difference-image approach, which (like the one described above for detection of clustered

microcalcifications) is novel in that it attempts to remove the structured anatomic background before applying feature-extraction techniques. After the difference between the signal-enhanced image and the signal-suppressed image is obtained, gray-level thresholding and feature-extraction techniques (involving the size, contrast and shape of the detected features) are performed by the computer to identify the locations of possible nodules. More recently, false-positive detections have been reduced by adding nonlinear filters to the difference-image step and additional feature-extraction techniques based on detailed analyses of the false positives.

The research team in the Rossmann Lab also has considerable experience in evaluation of factors affecting image quality and diagnostic accuracy in digital radiography. We have investigated basic imaging properties including the characteristic system response, spatial resolution properties and noise properties of various types of digital radiographic imaging systems (76-86). The effects of various physical parameters, such as detector system, sampling aperture, pixel size, number of quantization levels, exposure level and display aperture, were examined at various stages of the digital imaging chain (87-91). Knowledge gained in this research will be useful in understanding the effect of spatial resolution and noise on the performance of computer-assisted interpretation.

In developing methods for computer-assisted interpretations, it is crucial to employ appropriate means for evaluation. We have carried out various observer performance studies in comparing the detection capability of new techniques both with regard to simulated and clinical images. 18-alternative forced-choice observer studies were employed to examine the effect of pixel size on the threshold contrast of simple objects digitally superimposed on uniform background noise (92-94) and the effect of structured background on the detectability of simulated stenotic lesions (95). In an observer study with radiologists using clinical images, ROC analysis was employed in order to examine the effects of different display modalities (film and CRT) on diagnostic accuracy in digital chest radiography (96). Similar studies were performed to investigate the effect of data compression ratios on detectability (97), the comparison of computed radiography with conventional screen/film imaging (98), and the utility of computer-assisted interpretation in mammography (55) and chest (71). In addition, we have used ROC and FROC analyses to evaluate the performance level of the computerized schemes and the artificial

neural networks (99). This broad experience will provide the basis for developing similar methodology to evaluate the computer-vision modules for mammography proposed in this application.

Purpose of the present work

The main hypothesis to be tested is that given a dedicated computer-vision module for the computer-assisted interpretation of mammograms, the diagnostic accuracy for mammograhic interpretation will be improved, yielding earlier detection of breast cancer (i.e., a reduction in the number of missed lesions) and a reduction in the number of benign cases sent to biopsy.

Computer-aided diagnosis (CAD) can be defined as a diagnosis made by a radiologist who takes into consideration the results of a computerized analysis of radiographic images and uses them as a "second opinion" in detecting lesions and in making diagnostic decisions. The final diagnosis would be made by the radiologist. Although mammography is currently the best method for the detection of breast cancer, between 10-30% of women who have breast cancer and undergo mammography have negative mammograms (20-24). It has been suggested that double reading (by two radiologists) may increase sensitivity (100-102). Thus, one aim of CAD is to increase the efficiency and effectiveness of screening procedures by using a computer system, as a "second opinion or second reading," to aid the radiologist by indicating locations of suspicious abnormalities in mammograms.

If a suspicious region is detected by a radiologist, he or she must then visually extract various radiographic characteristics. Using these features, the radiologist then decides if the abnormality is likely to be malignant or benign, and what course of action should be recommended (i.e., return to screening, return for follow-up or send for biopsy). Many patients are referred for surgical biopsy on the basis of a radiographically detected mass lesion or cluster of microcalcifications. On average, only 10-20% of masses referred for surgical breast biopsy are actually malignant (20,28). Thus, another aim of CAD is to extract and analyze the characteristics of benign and malignant lesions in an objective manner in order to aid the radiologist by reducing the numbers of false-positive diagnoses of malignancies, thereby decreasing patient morbidity as well as the number of surgical biopsies performed and their associated complications

Methods of approach

The objective of the proposed research is to develop a dedicated computer-vision module for use in mammography in order to increase the diagnostic decision accuracy of radiologists and to aid in mammographic screening programs. The computer-aided diagnostic module will incorporate various novel computer-vision and artificial intelligence schemes already under development in the Rossmann Laboratories at the University of Chicago.

The specific objectives of the research to be addressed are:

(1) Further development of advanced computerized schemes for the detection and classification of masses and microcalcifications in digital mammograms. This part of the research involves quantitative analysis of the radiographic characteristics of masses and microcalcifications, and the decision-making processes used by radiologists in making a decision with respect to the likelihood of malignancy and in choosing the appropriate course of action.

(a) Further development of an advanced computerized detection scheme for masses that uses bilateral-subtraction techniques, gray-level thresholding, and analysis of various image features.

(b) Further development of an advanced computerized detection scheme for microcalcifications that uses linear and nonlinear spatial filters, spectral content analysis and various morphological filters for size, contrast and cluster analyses.

(c) Further development of advanced computerized classification schemes for masses and microcalcifications that use computer-vision techniques and artificial-intelligence techniques to calculate a probability of malignancy.

(2) Development of a dedicated module with man-machine interfaces appropriate for the effective and efficient use of the CAD schemes. Final diagnostic decisions will remain with the radiologists.

(a) Optimization of the CAD software.

(b) Examination of various methods of presenting the computer's results to the radiologist.

(c) Development of a prototype intelligent modular workstation using a high-speed (fast CPU & large-capacity memory) computer and a high-resolution, filmless CRT display.

(3) Evaluation of the efficacy and efficiency of the dedicated computer-vision module for mammography using a large clinical database. This part will use both film and filmless media for image acquisition and display.

BODY: Experimental methods and results to date

(1) Development of the computerized schemes for the detection and classification of masses and microcalcifications in digital mammograms.

Experimental methods

The computerized schemes for detection and classification are at various levels of development. These schemes will be used as aids by radiologists in the interpretation of mammograms. For the development and testing of these algorithms, we will collect 500 mammographic cases from the Department of Radiology. Initially, these cases will include screen/film mammograms that are currently acquired in the department. Later, the database will include digital images both from computed radiography (CR) units (stimulable phosphor) and from a CCD array detector that will be installed on our digital biospy unit (see Section 6 on Facilities).

(a) Development of the computerized detection scheme for masses.

The computer-vision scheme is based on deviations from the architectural symmetry of normal right and left breasts, with asymmetries indicating potential masses (60-62). Thus, we will continue investigating subtraction techniques as a means to increase the conspicuity of masses in mammograms. These techniques will be combined with analysis of individual mammograms. The input to the computerized scheme, for a given patient, are the four conventional breast images obtained in a routine screening examination: the right CC view, the left CC view, the right MLO view, and the left MLO view. Mammograms will be digitized using a laser scanner digitizer (2K by 2K matrix). In the initial detection stage, the digital image can be reduced to a 512 by 512 matrix (with an effective pixel size of

0.4 mm) due to the large size of masses relative to the pixel size. An automated alignment technique, which we have developed, will be used to align corresponding left and right breast images and also images of the same breast obtained over some time period. The automated alignment of two corresponding breast images will be performed in three stages: image segmentation, image feature selection and image registration. During image segmentation, the breast area will be isolated from the exterior region using a technique which combines multiple gray-level thresholding and morphological filtering. With image-feature selection, landmarks on each breast image will be determined. These landmarks are the breast border and the nipple position. Since the image features around the nipple often include a thicker skin line and greater subcutaneous parenchymal opacity, a band signature method will be employed to identify the nipple position along the breast border. During image registration, translation and rotation of one of the breast images relative to the other will be determined using a partial-border matching technique.

Once the two images are aligned relative to each other, the detection of possible asymmetries between the border-matched right and left breast images is achieved by correlation of the two mammograms, using a bilateral-subtraction technique. We are investigating linear and nonlinear subtraction methods. With linear subtraction, the two breast images are subtracted (using a left-minus-right convention) and then gray-level thresholding is performed in order to segment the image into possible locations of suspect masses. With the nonlinear technique, gray-level thresholding is performed prior to subtraction. This initial thresholding eliminates some normal anatomic background from further analysis. A selected number of images thresholded with different cutoff gray levels is obtained from the right breast image, and a corresponding number is obtained from the left breast image. Subtraction of ten sets of corresponding right and left breast images, each thresholded at ten different levels, is performed to generate ten bilateral-subtraction images (containing information on suspicious masses in the two original mammograms). A linking process then accumulates the information into two images, called runlength images, where the value of each pixel in each image indicates how often the corresponding location in the set of 10 subtraction images has gray levels above

or below a particular cutoff gray value. These images are next thresholded to yield the suspicious areas and submitted for feature extraction.

Feature-extraction techniques will be performed on both the runlength images and the original mammograms to reduce the number of false-positive detections. Initially, a morphological closing operation followed by a opening operation will be used to eliminate isolated pixels and merge small neighboring features. Next a size test will be used to eliminate features that are smaller than a predetermined cutoff size. A border test will be used to eliminate artifact features arising from any border misalignment that occurred during digitization and registration. On the original images, suspected regions will be subjected to region-growing techniques and then examined with respect to size, shape and contrast, in order to eliminate features of elongated shape and diffuse connective tissue.

Further analysis will be performed by correlating geometrically the information obtained from the CC mammographic pair and the MLO pair. Since the two views are obtained from the same breast image, the appearance of a mass in one view of a breast will be expected to exist in the other view of the same breast. This geometric correlation will need to take into account the different angles of projection of the 3-dimensional breast in forming the two 2-dimensional images and the possibly different amounts of physical compression applied to the breast in question during acquisition of the two views.

In addition to comparing the right and left breast images of a given view obtained at a given time, comparisons will be made between images of the same breast obtained at the same projection but at different times in order to note changes in the breast. This follows the methodology employed by mammographers when interpreting a case with previous examinations available. Similar subtraction techniques and feature-extraction methods will be employed. Use of histogram specification methods (103), however, may be necessary in order to match the gray-level distributions of the two images (that were obtained at different times) when there exists a large variation in the exposure techniques employed.

Results to date

Currently, the scheme employs two pairs of conventional screen-film mammograms (the right and left MLO views and CC views), which are digitized. After the right and left breast images in each pair are aligned, a nonlinear bilateral-subtraction technique is employed that involves linking multiple subtracted images to locate initial candidate masses. Various features are then extracted and merged using an artificial neural network in order to reduce false-positive detections resulting from the bilateral subtraction.

The features extracted from each suspected mass lesion include geometric measures, gradient-based measures and intensity-based measures. The geometric measures are lesion size, lesion circularity, margin irregularity, and lesion compactness. The gradient-based measures are the average gradient (based on a 3 by 3 Sobel operator) and its standard deviation calculated within the specified region of interest. The intensity-based measures are local contrast, average gray value, standard deviation of the gray values within the lesion, and the ratio of the average to the standard deviation. The features were normalized between 0 and 1 and input to the a back-propagation, feed-forward neural network. The ANN's structure consisted of 10 input units, one hidden layer with 7 hidden units and one output unit. In this task, the output unit ranged from 0 to 1, where 1 corresponded to the suspected lesion being an actual mass (i.e., a true-positive detection) and 0 corresponded to the suspected lesion being a false-posiitve detection (and thus, allowed to be eliminated as a suspect lesion-candidate). Based on the performances of the ANN as a function of iteration, in terms of self-consistency and round robin analyses, the optimal number of training iterations was determined.

ROC (receiver operating characteristic) analysis was applied to evaluate the output of the ANN in terms of its ability to distinguish between actual mass lesions and false-positive detections. The output values from the ANN for actual masses and for false-positive detections were used in the ROC analysis as the decision variable. Basically, the ROC curve represents the true-positive fraction and the false-positive fraction at various thresholds of the ANN output. ROC analysis was used a an index of performance in determining the "optimal" number of input features, the "optimal" number of hidden units, and the "optimal" number of training iterations of the ANN.

In the self-consistency analysis, the ANN achieved an Az of 1.0 and in the round-robin analysis, the ANN achieved an Az of 0.92 in distinguishing actual masses (true positives) from false-positive detectios. In an evaluation study using the 154 pairs of clinical mammograms (90 pairs with masses and 64 pairs without), the detection scheme yielded a sensitivity of 95% at an average of 2.5 false-positive detections per image. This was a substantial improvement from the previous year's performance of 85% sensitivity and 4 false-positive detections per image.

During the past year, we have even further reduced the number of false positives per image by expanding the types of gradient-based measures, and using them in addition to the features discussed above. In the feature extraction stage, the potential lesion was extracted from the parenchymal background using region growing techniques yielding the margin of the suspect mass. The gradient-based measures were calculated by first processing the region with a 3 by 3 Sobel filter yielding the maximum gradient and the angle of this gradient relative to the radial direction and a fixed (x-axis) at each pixel location. Cumulated gradient-weighted histograms were calculated for the maximum gradients across the various angles. From each histogram, various measures were calculated including full-width at half-maximum, average values, minima, heights, and standard deviations, which gave information such as the amount of spiculation and shape.

A three-level, feed-forward neural network, which utilizes a generalized delta rule in the training, was employed in this study. Fifteen features were chosen from an 91 initial features by analyzing the differences in the average and standard deviations of true positives (i.e., actual lesions) and false-positive detections. In addition, receiver operating characteristic (ROC) analysis was used to evaluate the individual performance of each feature in the task of distinguishing true positives from false-positive detections. The fifteen features included the three geometric measures and the three intensity-based measures, as well as nine of the gradient-based measures.

The parameters of the ANN, such as the number of hidden units, the learning rate, and the necessary number of training iterations, were determined empirically by evaluating the performance of the ANN as a function of each of the parameters. Area under the ROC curve was used to indicate performance. Both self consistency and round robin testing was employed.

Analysis of the ANN in distinguishing true positives (actual masses) from false positive detections yielded an Az of 0.99 and an Az of 0.97 in the consistency and robin round tests, respectively. This yielded a sensitivity of 90% at less than two false positives per image for the overall mass detection scheme using a database of 110 pairs of digital mammograms containing a total of 102 masses (54 malignant and 48 benign).

Also, during the past year, a new method for segmentation of the breast region in a mammogram was developed. The algorithm identifies unexposed and direct exposure image regions and generates a border surrounding the valid breast region, which can then be used as input for further image analysis and input to the CAD schemes. The program was tested on 740 digitized mammograms with the segmentation results being evluated by two experts mammograms and two medical physicsts. In 97% of the mammograms, the segmentation results were rated acceptable for use in computer-aided diagnosis schemes. Segmentation problems encountered inthe remaining 22 images (3%) were most often due to digitization artifacts or poor mammographic technique. The developed algorithm will be a valuable component of an "intelligent" workstation for computer-aided diagnosis.

(b) Development of the computerized detection scheme for microcalcifications.

Microcalcifications are a primary indicator of cancer and are often visible in the mammogram before a palpable tumor can be detected. Initially, clinical screen/film mammograms will be digitized using the laser scanner and analyzed in the 2048 by 2048 matrix format in order to retain the high spatial-frequency content of the microcalcifications. First, the original mammograms will be processed to enhance and suppress the signal of the microcalcifications, followed by calculation of a difference image. Both linear and nonlinear filters will be investigated for enhancement and suppression. Previous use of both linear and nonlinear filters in detecting lung nodules in digital chest images has shown that while both types of filters tended to detect nodules, locations of false positives differed. Thus, a combination of the results from each processing technique has the potential to yield high sensitivity and reduce the number of false-positive detections. Examples of filters for signal enhancement include a linear "matched" filter that matches the profile of a typical microcalcification and

a morphological open filter (to enlarge the appearance of microcalcifications). Morphological filtering (104) is basically a nonlinear filtering method that calculates the logical AND (erosion function) or OR (dilation function) of pixels within a kernel of some given size and shape. When extended to gray-scale images, the logical AND and OR operations can be replaced by minimum and maximum operations. By appropriately choosing the size and shape of the kernels, as well as the sequences of the AND and the OR, the filters can eliminate groups of pixels of limited size or merge neighboring pixels. Examples of filters for signal suppression include ring-shaped filters that yield either the average or median value of the surrounding normal anatomic background (54).

The difference image will then be subjected to various feature-extraction techniques to reduce further the number of false-positive detections. These techniques will test for size, contrast and spectral content of neighboring features. New methods for analyzing these features will involve the use of morphological filters. For example, we have found that the use of asymmetric morphological filters to eliminate features less than 3 pixels in size are more effective and efficient than use of a point-by-point analysis that involves counting the number of pixels in each remaining feature and comparing it to a size cutoff. In addition, the presence of clustering of the microcalcifications will be examined since singular microcalcifications are usually not cancerous. The morphological kernel for the clustering test will correspond to the size of a typical cluster (approximately 6 mm in diameter).

Results to date

The microcalcification detection scheme consists of three steps. First, the image is filtered so that the signal-to-noise ratio of microcalcifications is increased by suppression of the normal background structure of the breast. Second, potential microcalcifications are extracted from the filtered image with a series of three different techniques: a global thresholding based on the grey-level histogram of the full filtered image, an erosion operator for eliminating very small signals, and a local adaptive grey-level thresholding. Third, some false-positive signals are eliminated by means of a texture analysis technique, and a nonlinear clustering algorithm is then used for grouping the remaining signals.

In our computer detection scheme it is neccesary to group or cluster microcalcifications, since clustered microcalcifications are more clinically significant than are isolated microcalcifications. In the past we used a "growing" technique in which signals (possible microcalcifications) were clustered by grouping those that were within some predefined distance from the center of the growing cluster. In this paper, we introduce a new technique for grouping signals, which consists of two steps. First, signals that may be several pixels in area are reduced to single pixels by means of a recursive transformation. Second, the number of signals (non-zero pixels) within a small region, typically 3.2x3.2 mm, are counted. Only if three or more signals are present within such a region are they preserved in the output image. In this way, isolated signals are eliminated. Furthermore, this method can eliminate falsely detected clusters, which were identified by our previous detection scheme, based on the spatial distribution of signals within the cluster. The differences in performance of our CAD scheme for detecting clustered microcalcifications using the old and new clustering techniques was measured using 78 mammograms, containing 41 clusters. The new clustering technique improved our detection scheme by reducing the false-positive detection rate while maintaining a sensitivity of approximately 85%.

We also applied artificial neural networks to the differentiation of actual "true" clusters of microcalcifications from normal parenchymal patterns and from false positive detections as reported by a computerized scheme. The differentiation was carried out in both the spatial and spatial frequency domains. In the spatial domain, the performance of the neural networks was evaluated quantitatively by means of ROC analysis. We found that the networks could distinguish clustered microcalcifications from normal nonclustered areas in the frequency domain, and that they could eliminate approximately 50% of false-positive clusters of microcalcifications while preserving 95% of the positive clusters.

The number of false-positive detections was even further reduced when a shift-invariant artificial neural network (SIANN) was used to analyze the remaining suspected locations. The SIANN is a multilayer back-propagation neural network with local, shift-invariant interconnections. The advantage of the SIANN is that the result of the network is not dependent on the locations of the clustered microcalcifications in the input layer. The performance of the SIANN was evaluated by means of a

jack-knife method and ROC analysis using a database of 168 regions as reported by the CAD scheme. Approximately 55% of the false positives were eliminated without loss of any of the true-positive detections. These technique led to a performance of 85% sensitivity with less than 0.6 false-positive detections per image. In this study, we also examined the effect of the network structure on the performance of the SIANN.

During the past year, modifications were made to improve the performance of the SIANN. First, the preprocessing was removed because the result of background-rend correction is affected by the size of ROIs. Second, image-feature analysis was employed to the output of the SIANN in an effort to eliminate more of the false detections. In order to train the SIANN to detect microcalcifications and also to extract image features of microcalcifications, zero-mean-weight constraint and training-free-zone techniques were developed. A cross-validation training method was also applied to avoid the over-training problem. The performance of the SIANN was evaluated by means of ROC analysis using a database of 39 mammograms for training and 50 different mammograms for testing. The analysis yielded an average area under the ROC curve (Az) of 0.90 for the testing set. Approximately 62% of false-positive clusters detected by the rule-based scheme were eliminated without any loss of the true-positives clusters by using the improved SIANN with image feature analysis techniques.

(c) Development of computerized classification schemes.

Various feature-extraction techniques and artificial intelligence schemes will be investigated in order to distinguish malignant masses and/or microcalcifications from benign masses and/or microcalcifications. The database for this investigation will be obtained from the conventional four screening breast images, as well as special views such as spot compression.

In our previous work, we compiled a list of features that radiologists use in distinguishing between malignant and benign masses. These features include: margin spiculation (number of spiculations, length of spiculation, and difference between spicules and local linear features), shape (linear to spherical, geometrical to diffuse, and existence of satellite lesions), size (mean diameter), margin characteristics (complete to inseparable from surround, well-defined to indistinct, and presence of halo

sign), and pattern of interior (uniformity, presence of well-defined lucencies, and opacity relative to size). The analysis of spiculation will be based on a novel computer-vision method involving the Fourier analysis of the fluctuations around the margin of the mass in question (60). The computer-extracted margin used in the analysis for spiculation also contains information related to the number and length of spiculations. Also, prior to the analysis of spiculation, the mass is extracted from the normal anatomic background of the breast parenchyma. Currently, region-growing techniques are employed for this extraction. Once extracted, the shape and size of the mass can be easily calculated. The size will be defined as the effective diameter of a circle that has the same area as the extracted mass. The shape will be expressed by a degree of circularity, which will be defined as the ratio of the area of the mass within the equivalent circle to the total area of the mass. Masses with ill-defined margins are more likely to be malignant than those with relatively well-defined margins. Thus, a margin gradient test will be developed to measure the sharpness of the margin. This sharpness will be defined as the degree of density change across the margin and will be measured perpendicular to the margin at all points along the margin. The pattern of the interior will be quantitatively determined from the spectral content of the interior.

Features related to the classification of microcalcifications include: the shape of the individual microcalcifications (rounded to irregular, linear, and branched), uniformity of microcalcifications within a cluster (uniformity in size, shape, and density), distribution of the microcalcifications (diffuseness and shape of cluster) and presence of macrocalcifications. The size and shape of the individual microcalcifications will be determined by the computer using an effective diameter and a circularity measure, respectively, as described earlier. Uniformity within a cluster will be assessed by calculating the spread of values for a particular characteristic such as size. Once a cluster has been defined, its diffuseness will be given by the number of microcalcifications per unit area and its shape will be defined using a circularity measure.

These various computer-determined quantitative measures describing the mass or cluster of microcalcifications in question will be input to an artificial neural network that will merge the features into a probability of malignancy for use by radiologists. As mentioned in the Background section, our

work with a neural network in merging human-reported mammographic features into a malignant/benign decision has been extremely promising. The input data (corresponding to the computer-extracted features of the masses and microcalcifications) will be represented by numbers ranging from 0 to 1 and will be supplied to the input units of the neural network. The output data from the neural network is then provided from output units through two successive nonlinear calculations in the hidden and output layers. The calculation at each unit in a layer includes a weighted summation of all entry numbers, an addition of a certain offset number, and a conversion into a number ranging from 0 to 1 using a sigmoid-shape function such as a logistic function. Two different basic processes are involved in a neural network; namely, a training process and a testing process. The neural network will be trained by a back-propagation algorithm (105) using input data (i.e., computer-reported features) and the desired corresponding output data (i.e., biopsy or follow-up proven truth of the malignant or benign status of the mass or microcalcifications in question), for a variety of cases. Once trained, the neural network will accept computer-reported features of the mass or microcalcifications in question and output a value from 0 to 1 where 0 is definitely benign and 1 is definitely malignant. Based on the distribution of these values for various known cases, we will be able to determine what course of action (e.g., biopsy, follow-up or return to normal screening) should be recommended to the radiologist.

Results to date

Classification of masses

Our earlier work showed that a back-propagation, feed-forward artificial neural network could merge human-extracted features of mammographic lesions into a likelihood of maligannncy at a similar level of that of an expert mammographer. In the study presented here, however, ANN is used to merge computer-extracted features of mass lesions into a likelihood of malignancy.

The method takes as input the center location of a mass lesion in question. Next, the lesion is segmented from the breast parenchyma (background) using an automatic region growing technique and various features of the lesion are extracted. The automatic lesion segmentation involves the analysis of the size of the grown region as a function of the gray-level interval used for the region growing. Many of the extracted features are determined from a cumulative edge-gradient-orientation histogram analysis modified for orientation relative to a radial angle. Input to an ANN consists of four features from the gradient analysis along with the average gray value within the grown lesion. The gradient measures include the FWHM (full width at half max) of the cumulative edge-gradient-orientation histogram calculated from pixels within the lesion and its neighboring surround, and from just pixels along the lesion margin (see Figures 1-3). These measures correspond to the presence of spiculation, which is a sign of malignancy in the visual interpretation of mammographic masses. The ANN's structure consisted of 5 input units, one hidden layer with 4 hidden units and one output unit. In this task, the output unit ranged from 0 to 1, where 1 corresponded to the lesion being malignant and 0 corresponded to the lesion being benign. Use of ROC analysis with self-consistency testing and round-robin testing was employed as discussed in the previous section.

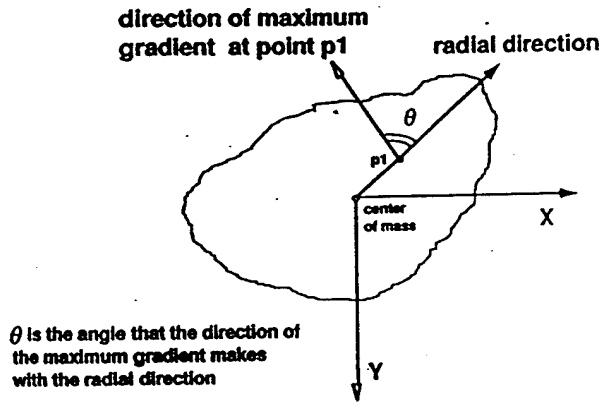


Figure 1. Schematic illustrating the maximum gradient and its direction relative to the radial direction indicated.

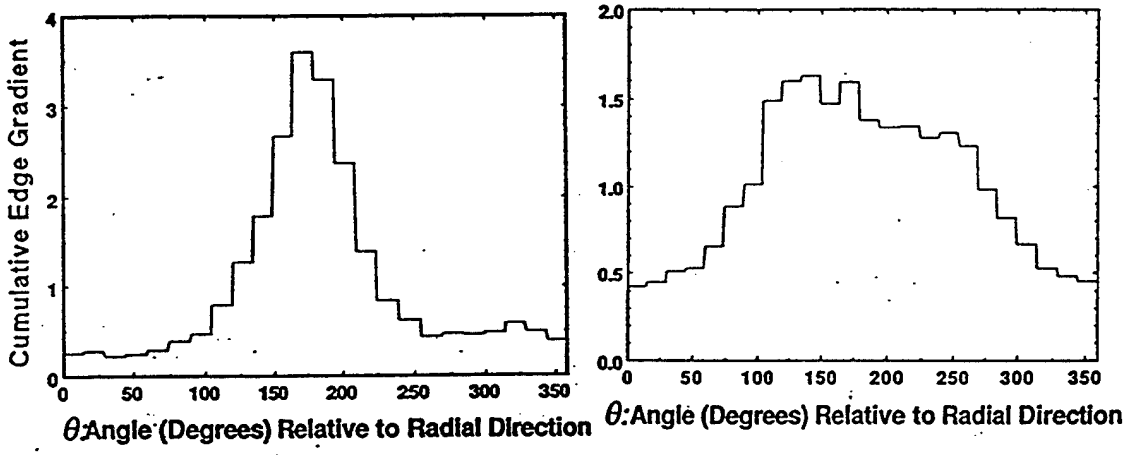


Figure 2. Cumulative edge-gradient-orientation histograms illustrating the presence of (a) a non-spiculated mass as indicated by the relatively narrow peak in the histogram and (b) a spiculated mass as indicated by the presence of a broader peak.

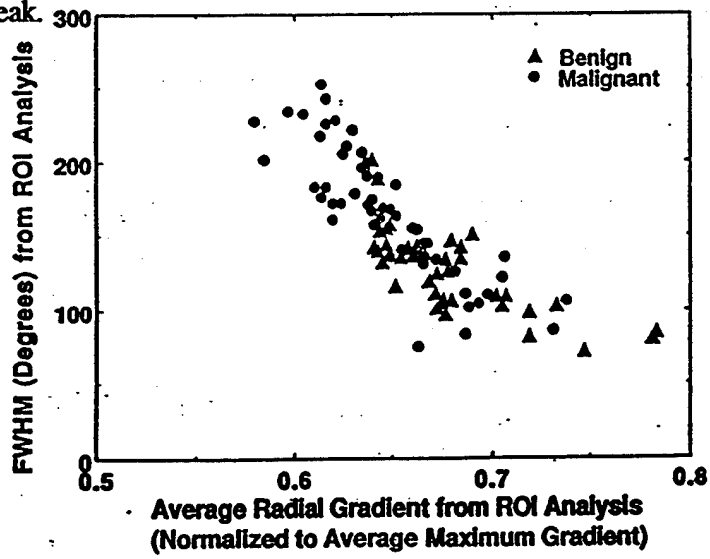


Figure 3. Relationship between FWHM of edge-gradient-orientation histogram and average radial gradient for malignant and benign masses.

The classification method was evaluated using a pathologically-confirmed database of 95 masses (57 malignant and 38 benign), of which all but one had been sent to biospy. The mammograms in the database had been digitized to a pixel size of 0.1 mm. Using the five input features, an Az (area under to the ROC curve) of 0.83 was obtained in the task of distinguishing benign from malignant masses using a round-robin method for evaluation. However, we found that by using a rule-based decision on one of the features (FWHM) based on its correspondence to visual interpretation methods, prior to use of the ANN, the performance increased yielding an Az of 0.90 for distinguishing between malignant and benign masses. We found that using a combination of the measurements from the four neigborhoods is superior in the classification of mammographic mass lesions.

Results to date

Classification of microcalcifications

The analysis of microcalcifications can be difficult to perform consistently for human observers leading to the poor positive predictive value. We have been investigating methods to identify computer-extracted quantitative features of microcalcifications and their clusters that can be used to classify malignant and benign clustered microcalcifications, and, to exam if a computer can make accurate differential diagnoses based on computer-extracted features. In this study, features of the microcalcifications and their clusters were automatically extracted from digitized conventional mammograms.

The microcalcifications were segmented using the following method, which is described in detail elsewhere. A third-degree polynomial was fitted to the pixel-value distribution in a ROI (region of interest) of the digitized mammogram in both horizontal and vertical directions to reduce the background structure of the breast parenchyma. The microcalcification was then delineated by region growing. The effective thickness of the microcalcification (physical dimension along x-ray projection line) was estimated from signal contrast (mean pixel value above background) of the isolated

microcalcification. This was done by first converting signal contrast in terms of optical density to contrast in terms of exposure using knowledge of the H&D curve of the screen-film system, and secondly converting contrast in terms of exposure to physical dimension using the exponential attenuation law assuming a "standard" model of the breast and the microcalcification. The standard model assumes (i) a 4-cm compressed breast composed of 50% adipose and 50% glandular tissues; (ii) a microcalcification composed of calcium hydroxyapatite with physical density of 3.06 g/mm^3; and (iii) a 20-keV monochromatic x-ray beam. Two contrast corrections were applied for better accuracy: compensation for blurring caused by the screen-film system and the digitization process, and compensation for x-ray scatter.

The usefulness of the features were evaluated using the distributions of the benign and malignant populations. Features capable of showing separation between benign clusters from the malignant population were chosen for the automated classification. Extracted features were based on the size, shape, contrast, and uniformity of individual microcalcifications; and the size and shape of microcalcification clusters. An artificial neural network was used to classify benign versus malignant clusters of microcalcifications using 8 computer-extracted features. The database consisted of 100 images, digitized at 100-mm pixel size and 10-bit grey-scale resolution, from 53 patients biopsied for suspicion of breast cancer based on clustered microcalcifications. The neural network correctly identified 69% of the benign patients, all of whom had biopsies, and 100% of the malignant patients.

In the past year, an observer study was performed which indicated that for the cases used, the performance of the computer method was statistically higher than that of radiologists ($p=0.03$). The observer study included three experienced mammographers and two radiology fellows.

(2) Development of a dedicated CAD module for use by radiologists.

Experimental methods

The various computer-vision and artificial intelligence schemes will be incorporated into a dedicated computer system (module) equipped with a high-speed computer and a digital image interface, as shown in Figure 4. The digital image interface will initially be to a film digitizer in order to test the CAD schemes using the large database of clinical mammograms available in our Radiology Department. Later, mammographic images will be obtained using the CR system or the CCD-based digital biopsy unit. The intelligent modular workstation will need to have sufficient computer power (CPU and large capacity memory) and display capabilities to allow for "real-time" computation and viewing of the computer-vision results. Thus, we plan to upgrade our current computer hardware and optimize our software to achieve high-speed and efficient computation of CAD results. Our target is to reduce the CPU time required for CAD computations from the current level of about 5 minutes per image to a few seconds. Also, appropriate man-machine interfaces will be needed for effective and efficient computer-assisted interpretations. This part of the research will involve the examination of various methods of presenting the computer-determined results to the radiologists. Important parameters include (a) the shape and size of the markers of the computer output that could represent the severity or confidence level (probability) of the lesion, (b) the optimal operating point of the CAD schemes (high sensitivity with an acceptable number of false positives), (c) the timing and duration of displaying the computer output, (d) the selection of the minimum number of inputs required for radiologists and (e) the user-friendliness of instructions and input entries.

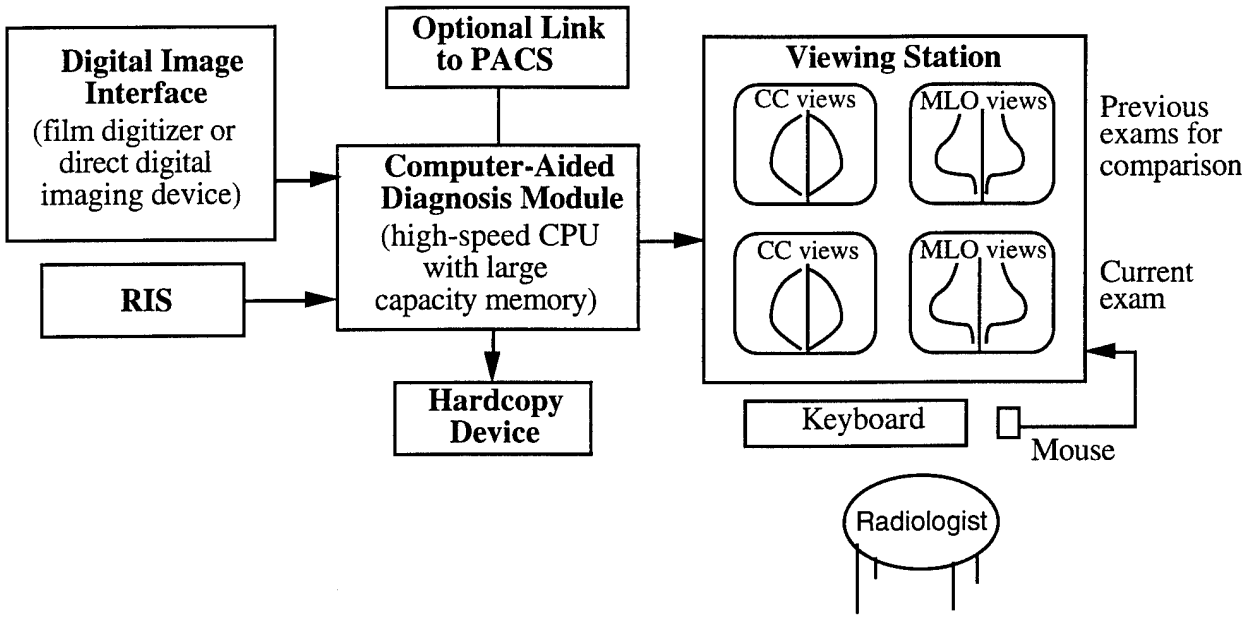


Figure 4. The intelligent modular workstation. The digital image interface will be the film digitizer and also direct digital imaging devices such as the CR system and the digital biopsy unit. The CAD module requires a high-speed CPU and large capacity memory. Hardcopy devices include the laser film printer and the economical thermal-paper image printer. Optional links to RIS (radiology information system) and PACS (picture archival and communication system) are included.

The development of the prototype modular system will be achieved in stages. In Phase 1, the introduction of the computer-vision aid to the radiologists will be implemented with minimum change in the current radiologist method of operating. This will allow for a gradual introduction in order to minimize any resistance to change. Thus, only computer-reported detection results will be presented to the radiologist, leaving all of the interpretation to the radiologist. Basically, the computer will serve as a "second opinion" indicating suspicious areas without critique as to their degree of malignancy. Original films will be digitized (2048 by 2048 digitization matrix) and analyzed, with the computer output then printed on either film or thermal paper. Radiologists will perform their normal reading using the original image and the computer results. It is believed that this introduction of CAD to radiologists will cause minimum modification to their normal reading patterns, thus allowing for a smooth and effective transition. During Phase 2, results from the classification schemes also will be included, using the methodology described for Phase 1. Howver, in this second phase the computer will serve as a "second opinion" for both the location and the interpretation of breast lesions.

During the first two phases, we will investigate the best markers for use by radiologists, who may prefer arrows or circles (icon-type symbols). It should be noted that the implementation of computer vision in mammographic screening using the methods described above is not limited to fully digital (PACS) departments but can be incorporated in a general film-based radiology department or in a mobile, filmless mammography unit (i.e., a limited PACS environment).

Once the use of computer vision is shown to be useful, beneficial and efficient, we will incorporate high-resolution, state-of-the-art monitors into the dedicated computer system (Phase 3) as shown in Figure 1. The "intelligent" module will be interfaced to our department's RIS (radiology information system) to link the demographic and medical history information with the CAD output. In order for the radiologist to examine the entire breast image, the display monitor will need to have 2K by 2K capability. In mammography, each breast image usually can be digitized adequately into a 2K by 1K image. Thus, in order to view all four breast images (left and right CC views and left and right MLO views), two high-resolution 2K by 2K monitors are needed. However, in the practice of radiology, films (images) from previous examinations play an important role in the current exam due to the need for comparison in order to detect subtle changes. Thus, the display requirements are four 2K by 2K monitors (in a 2 by 2 arrangement), allowing the top two monitors to be used for sequencing through previous exams of the patient in question. In this phase, the radiologists will do their reading of the mammographic cases from the high-resolution monitors. Due to the dynamic nature of the display, the computer-reported results can be presented in a toggle format where the radiologist can press a button to either show or remove the computer-reported results. In addition, the computerized schemes can be configured to allow for the radiologist to control the tradeoff between the sensitivity and specificity of the computer output, because more true-positive detections always can be achieved at the cost of a larger number of false-positive findings, and vice versa. This tradeoff would be adjusted by the radiologist, depending on the nature of the case material and personal preference. For example, a radiologist might choose a computer output with high sensitivity for examining high-risk patients, whereas a lower sensitivity and correspondingly lower false-positive rate might be preferred for patients at low risk for cancer. It should be noted, however, that increasing the number of interactive choices available to the

radiologist will lengthen the reading time per case. Therefore, we will investigate optimization of the module's human interface by studying the relationship between achievable diagnostic accuracy and required reading time.

Results to date

The computerized image analysis software has been integrated into a user friendly interface based on UNIX, XWINDOWS and Motif and operated on an IBM RISC 6000 Series 570 computer workstation. The prototype (hardware & software) was demonstrated at the 1994 annual meeting of the Radiological Society of North America (RSNA) and was well received by the many radiologists in attendance. Currently, arrows (red for masses and yellow for clustered microcalcifications) are used to indicate the computer-detected location of lesions. The input to the system can be either a film that is digitized and then analysed automatically or a computer file containing a digital image. The prototype system is interfaced to a Konica laser film digitizer which enables digitization of the mammograms to approximately 2K by 2K matrices. Video output of the IBM monitor is connected to a low-resolution thermal printer (approximately 1K by 1K) for hardcopy reporting of the CAD results.

In the past year, our prototype workstation was placed in the clinical mammography reading area of the Department of Radiology. Since Nov. 8, 1994, we have analyzed over 1000 screening cases. Results are discussed in the next section.

(3) Evaluation procedure using large clinical databases

Experimental methods

As described in the previous section, the computer-vision methods for mammography will be developed in phases. Plans include testing the computer-vision system at the end of each phase in order to demonstrate the effect of the various modes of presentation on the accuracy, efficiency and acceptability of the mammographic aid. The system will be evaluated using clinical mammograms obtained from both a low-risk population and a high-risk population. The low-risk population will be

obtained from The University of Chicago mammography screening program. The high-risk population will be drawn from examinations referred to our Department of Radiology, since The University of Chicago serves as a tertiary medical center. Initially, performance studies will be done using a database of preselected mammographic cases that have a distribution of subtle cases of normal, benign and malignant areas of either masses or microcalcifications. Later studies will be performed using a more representative database of consecutive mammographic cases obtained from four weeks worth of screening. "Truth" concerning the presence and malignancy of masses and microcalcifications will be established with the aid of expert mammographers, follow-up reports and surgical biopsy reports. Normal cases will be selected from patients who have had normal follow-up exams. Performance studies will be done using cases involving the four conventional mammograms (left and right CC views, and left and right MLO views), since these are the usual images obtained in screening.

At the detection stage of the computer-vision system, performance will be examined by calculating the fraction of lesions detected (true-positive rate) and the number of falsely-reported areas per case. At the classification stage of the computer-vision system, performance will be examined by calculating the fraction of malignant cases correctly classified (true-positive classification rate) and the number of benign cases that are reported by the computer as being malignant (false-positive classification rate). The clinical database for these performance evaluations will contain 180 cases (60 normal, 30 with benign masses, 30 with malignant masses, 30 with benign microcalcifications, and 30 with malignant microcalcifications).

Observer studies will be performed to examine the usefulness of the computer-assisted interpretation process in enhancing radiologists' performance levels, as compared to the unaided performance by radiologists. During phases 1 and 2, the database cases will be printed with the computer-vision results on each film. These database cases will then be used in observer performance studies. Stratified sampling (106) will be used in choosing subtle cases in order to avoid problems associated with either "too easy" or "too difficult" cases (107). Twelve attending radiologists and senior residents will act as observers. Then, for the 180 cases in the database, three "reading methods" will be tested; (a) the original cases without the computer-vision aid, (b) the cases with the detection-

results reported (phase 1 computer locations of suspicious areas) and (c) the cases with both the detection and classification results reported (phase 2 computer locations with probability of malignancy). Each observer will be asked to perform two tasks: (1) locate and rate suspicious areas as to the presence of an abnormality (rating scale of 0 to 100) and (2) indicate an overall level of certainty as to the presence of cancer using a 5-point rating scale where 1=definitely benign and 5=definitely malignant. This five-point scale is the same as that being recommended by the American College of Radiology for routine use by clinical mammographers. The dual-task observer study will allow for evaluation of the utility of both the computer-vision detection and classification results. (In addition, questionnaires will be given to each observer in order to obtain subjective information with regard to the efficiency and acceptability of the computer-vision mammography system.) In the analysis of the observer study results, maximum likelihood estimation (108) will be used to fit a binormal ROC (receiver operating characteristics) curve to each observer's confidence-rating data from each diagnostic method. The index Az, which represents the area under a binormal ROC curve, will be calculated for each fitted curve. To represent the average performance of the observers for each diagnostic method, the composite ROC curves will be calculated by averaging the slope and intercept parameters of the individual observer-specific ROC curves. The statistical significance of apparent differences between pairs of diagnostic methods will then be analyzed by applying a "two-tailed" t-test for paired data to the observer-specific Az index values.

Free-response ROC (FROC) analysis (109) and FROC-AFROC analysis (110) will be used in analyzing the data pertaining to localization of the abnormality. The ordinates of both FROC curves and AFROC curves are the fraction of lesions (masses or microcalcifications) that are correctly localized by the observer. However, the abscissa of an FROC curve is the average number of false positives per image, whereas the abscissa of an AFROC curve is the probability of obtaining a false-positive image (i.e., an image containing one or more false-positive responses).

After phase 3, another observer study will be performed in which four weeks' worth of mammographic cases will collected and interpreted by six radiologists with and without the computer-vision results of location and classification. Although this database lacks the control over the subtlety of

the cases that the earlier mentioned study has, it represents a more typical clinical situation. Half of the radiologists will read the first two weeks of cases without aid and the second two weeks of cases with the mammographic aid; and the other half of the radiologists will read the first two weeks of cases with the aid and the second two weeks of cases without the aid. Rating methods and analyses will be the same as mentioned above.

Results to date

FROC analysis and ROC analysis has been used extensively for the intermediate testing results of the various detection and classification methods. Constant collection of the database is ongoing. Investigators have developed a case reporting sheet for organizing the new cases on a Macintosh computer using FileMakerPro software. The various databases being collected include pathologically-proven mass and clustered microcalcification cases. In addition, a "missed lesion" database is being digitized in order to test the detection methods in the upcoming grant period. This database includes lesions that were seen in retrospect, i.e., after the cancer was detected at a later date. This database will demonstrate the ability of the detection schemes to increase the sensitivity of detection in a screening program. In a preliminary study (presented at the RSNA 94) in which 26 "missed lesion" cases were analyzed, the computerized detection schemes achieved a sensitivity of 50%. (Note that these "missed lesion" cases can be thought of yielding a sensitivity of 0% when they had been read by the radiologists).

We have been tabulating the performance of the clinical intelligent mammography workstation. In this prospective study, the results of the computer output have been quite promising. Since the study is prospective, we do not know "truth" yet, although we are currently following the workups and biopsy results. Approximately 70% of the cases deemed suspicious by the study radiologist have been detected by the computer. To date, the missed cases have either been found to be benign or are still in workup/biopsy. In fact, to date, a confirmation of a cancer (malignant case) has not been made. In two cases, a cluster of microcalcifications was located by the computer but not by the radiologists. A large number of screening cases need to be analyzed by the workstation prior to assessment of its performance and contribution in the mammographic interpretation process, since

with screening mammography, only 5 to 10 cancers are found for every 1000 patients. The false-positive rates are, on average, 0.92 false clusters per image and 1.4 false masses per image. Many of the false clusters are due to calcified vascular structures and many of the false masses are due to nodular-like structures. We found that the study radiologist can easily learn to recognize typical false positives and disregard them in her assessment of the presence of a lesion.

CONCLUSIONS

Substantial inprovements in the performances of the computer-aided diagnosis methods for the detection of masses and clustered microcalcifications have been achieved during the past funding period. For the detection of masses, the sensitivity remained constant, while the false-positive rate per image reduced to less than 2 per image. For the detection of clustered microcalcifications, the false-positive rate was reduced from 2 per image to approximately 0.7 per image, without loss in sensitivity. Constant collection of the database is ongoing. Investigators have developed a case reporting sheet for organizing the new cases on a Macintosh computer using FileMakerPro software. The various databases being collected include pathologically-proven mass and clustered microcalcification cases. Databases for both mammograms containing mass lesions and mammograms containing microcalcifications have both increased in size and some have been digitized on more than one digitizer in order to observe the affect of digitization on detection performance. In addition, a "missed lesion" database is being digitized in order to test the detection methods in the upcoming grant period. This database includes lesions that were seen in retrospect, i.e., after the cancer was detected at a later date. This database will demonstrate the ability of the detection schemes to increase the sensitivity of detection in a screening program. In a preliminary study (presented at the RSNA 94) in which 26 "missed lesion" cases were analyzed, the computerized detection schemes achieved a sensitivity of 50%. (Note that these "missed lesion" cases can be thought of yielding a sensitivity of 0% when they had been read by the radiologists).

With regard to the classification of mammographic lesions as an aid in distinguishing between malignant and benign cases, the initial performances for both masses and microcalcifications has been quite promising. In the classsification of masses, an Az (area under the ROC curve) of 0.90 was obtained from the ROC analysis of the output from the neural network, which was used to merge the extracted features of the lesions. In the classification of clustered microcalcifications a neural network correctly identified 69% of the benign patients, all of whom had biopsies, and 100% of the malignant patients. We conclude that a computer is capable of distinguishing benign from malignant clustered microcalcifications even at 100-mm pixel size.

The computerized image analysis software has been integrated into a user friendly interface based on UNIX, XWINDOWS and Motif and operated on an IBM RISC 6000 Series 570 computer workstation. The prototype (hardware & software) was demonstrated at the 1994 annual meeting of the Radiological Society of North America (RSNA) and was well received by the many radiologists in attendance. The input to the system can be either a film that is digitized and then analysed automatically or a computer file containing a digital image. The prototype system will be transferred to the clinical reading area for the next phase of development and testing.

We are very optimistic about the continuing success of our research. We will continue to improve the detection and classification performance of our algorithms. The mammographers in the clinical reading area of the department are pleased with the prototype. Weekly meetings are held between the basic science and clinical researchers in order to ensure a smooth integration of the workstation in the clinical arena. The results with the clinical prototype are promising and a full clinical trial begins next month.

REFERENCES

1. Silverberg E, Boring CC, Squires TS: Cancer Statistics, 1990. CA 40: 9-27, 1990.
2. Tabar L, Dean PB: Basic principles of mammographic diagnosis. Diagn. Imag. Clin. Med. 54: 146-157, 1985.
3. American Cancer Society: CA Cancer J Clin 33: 255, 1983.
4. Baker L: CA Cancer J Clin 32: 194, 1982.
5. NCRP Report No. 85: Mammography (National Council on Radiation Protection: Washington, D.C., 1986).
6. Shapiro S, Venet WS, Strax PH, et al.: Ten to fourteen-year effect of screening on breast cancer mortality. JNCI 69: 349-355, 1982.
7. Verbeek ALM, Hendricks JH, Holland R, et al: Reduction of breast cancer mortality through mass screening with modern mammography. Lancet 1: 1222-1224, 1984.
8. Collette HJA, Day NE, et al.: Evaluation of screening for breast cancer in a non-randomized study (the DOM project) by means of a case-control study. Lancet 1: 1224, 1226, 1984.
9. Tabar L, Gad A, Holmberg LH, et al.: Reduction in mortality from breast cancer after mass screening with mammography. Randomized trial from the Breast Screening Working Groups of the Swedish National Board of Health and Welfare. Lancet 1: 829-832, 1985.
10. Andersson I, Aspegren L, Janzon L, et al.: Mammographic screening and nortality from breast cancer: The Malmo mammographic screening trial. Br. Med. J. 297:943,1988.
11. Feig SA: Decreased breast cancer mortality through mammographic screening: Results of clinical trials. Radiology 167: 659-665, 1988.
12. Black JW, Young B: A radiological and pathological study of the incidence of calcifications in diseases of the breast and neoplasms of other tissues. Br. J. Radiol. 58: 596-598, 1965.
13. Wolfe JN: Analysis of 462 breast carcinomas. AJR 121: 846-853, 1974.
14. Sickles EA: Mammographic detectability of breast microcalcifications. AJR 139: 913-918, 1982.
15. Fisher ER, Gregorio RM, Fisher B, et al.: The pathology of invasive breast cancer. Cancer 36: 1-84, 1975.
16. Millis RR, Davis R, Stacey AJ: The detection and significance of calcifications in the breast: A radiological and pathological study. Br. J. Radiol. 49: 12-26, 1976.
17. Murphy WA, DeSchryver-Kecskemeti K: Isolated clustered microcalcifications in the breast: Radiologic-pathologic correlation. Radiology 127: 335-341, 1978.
18. Muir BB, Lamb J, Anderson TJ: Microcalcification and its relationship to cancer of the breast: Experience in a screening clinic. Clin. Radiol. 149: 193-200, 1983.
19. Sickles EA: Mammographic features of 300 consecutive nonpalpable breast cancers. Am J Rad 146: 662-663, 1986.
20. Bassett LW, Gold RH: Breast cancer detection: Mammography and other methods in breast imaging, Grune and Stratton (New York), 1987.
21. Baines CJ, Miller AB, Wall C, McFarlane DV, et al.: Sensitivity and specificity of first screen mammography in the Canadian National Breast Screening Study: A preliminary report from five centers. Radiology 160: 295-298, 1986.
22. Pollei SR, Mettler FA, Bartow SA, Moradian G, Moskowitz M: Occult breast cancer: Prevalence and radiographic detectability. Radiology 163: 459-462, 1987.
23. Andersson I: What can we learn from interval carcinomas? Recent Results in Cancer Research 90: 161-163, 1984.
24. Martin JE, Moskowitz M, Milbrath JR: Breast cancers missed by mammography. AJR 132: 737, 1979.
25. Buchanann JR, Spratt JS, Heuser LS: Tumor growth, doubling times, and the inability of the radiologist to diagnose certain cancers. Radiologic Clinics of North America 21: 115, 1983.
26. Holland T, Mrvunac M, Hendriks JHCL, et al.: So-called interval cancers of the breast. Pathologic and radiographic analysis. Cancer 49:2527, 1982.
27. Tabar L, Dean PB: Teaching Atlas of Mammography, George Thieme Verlag (Stuttgart, New York), 1983.

28. Moskowitz M: Screening for breast cancer: How effective are our tests? A critical review. Ca-A Cancer Journal for Clinicians 33: 26-39, 1983.
29. Winsberg F, Elkin M, Macy J, Bordaz V, Weymouth W: Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis. Radiology 89: 211-215, 1967.
30. Spiesberger W: Mammogram inspection by computer. IEEE Transactions on Biomedical Engineering 26: 213-219, 1979.
31. Kimme C, O'Loughlin BJ, Sklansky J: Automatic detection of suspicious abnormalities in breast radiographs. In: Data Structures, Computer Graphics, and Pattern Recognition, edited by A. Klinger, K. S. Fu, T. L. Kunii (Academic Press, New York), 1975, pp. 427-447.
32. Hand W, Semmlow JL, Ackerman LV, Alcorn FS: Computer screening of xeromammograms: A technique for defining suspicious areas of the breast. Computers and Biomedical Research 12: 445-460, 1979.
33. Semmlow JL, Shadagoppan A, Ackerman LV, Hand W, Alcorn FS: A fully automated system for screening xeromammograms. Computers and Biomedical Research 13: 350-362, 1980.
34. Ackerman LV, Gose EE: Breast lesion classification by computer and xeroradiography. Cancer 30: 1025-1035, 1972.
35. Ackerman LV, Mucciardi AN, Gose EE, Alcorn FS: Classification of benign and malignant breast tumors on the basis of 36 radiographic properties. Cancer 31: 342-352, 1973.
36. Wee WG, Moskowitz M, Chang N-C, Ting Y-C, Pemmeraju S: Evaluation of mammographic calcifications using a computer program. Radiology 116:717-720, 1975.
37. Fox SH, Pujare UM, Wee WG, Moskowitz M, Hutter RVP: A computer analysis of mammographic microcalcifications: Global approach. Proc. IEEE 5th International Conf. on Pattern Recognition: 624-631, 1980.
38. Magnin IE, Cluzeau F, Odet CL, Bremond A: Mammographic texture analysis: an evaluation of risk for developing breast cancer. Optical Engineering 25: 780-784, 1986.
39. Caldwell CB, Stapleton SJ, Holdsworth DW, Jong RA, Weiser WJ, Cooke G, Yaffe MJ: Characterization of mammographic parenchymal pattern by fractal dimensions. Phys. Med. Biol. 35: 235-247, 1990.
40. Fam BW, Olson SL, Winter PF, Scholz FJ: Algorithm for the detection of fine clustered calcifications on film mammograms. Radiology 169: 333-337, 1988.
41. Olson SL, Fam BW, Winter PF, Scholz FJ, Lee AK, Gordon SE: Breast calcifications: Analysis of imaging properties. Radiology 169: 329-331, 1988.
42. Davies DH, Dance DR: Automatic computer detection of clustered calcifications in digital mammograms. Phys. Med. Biol. 35: 111-118, 1990.
43. Astley S, Taylor C, Boggis C, Wilson M, Ellison T: Automated detection of abnormalities on screening mammograms. Radiology 177(P):288, 1990.
44. Grimaud M, Muller S, Meyer F: Automated detection of microcalcifications in mammograms. Radiology 177(P):288, 1990.
45. Jin H-R, Matsumoto K, Kobatake H: Automatic diagnosis of breast cancer with structural line analysis and mathematical morphology. Radiology 177(P) 319, 1990.
46. Karssemeijer N: A stochastic method for automated detection of microcalcifications in digital mammograms. Information Processing in Medical Imaging, Springer-Verlag (New York), pp.227-238, 1991.
47. Lai SM, Li X, Bischof WF: On techniques for detecting circumscribed masses in mammograms. IEEE Transactions on Medical Imaging 8: 377-386, 1989.
48. Brzakovic D, Luo XM, Brzakovic P: An approach to automated detection of tumors in mammograms. IEEE Transactions on Medical Imaging 9: 233-241, 1990.
49. Gale AG, Roebuck EJ, Riley P, Worthington BS, et al.: Computer aids to mammographic diagnosis. British Journal of Radiology 60: 887-891, 1987.
50. Getty DJ, Pickett RM, D'Orsi CJ, Swets JA: Enhanced interpretation of diagnostic images. Invest. Radiol. 23: 240-252, 1988.
51. Swett HA, Miller PA: ICON: A computer-based approach to differential diagnosis in radiology. Radiology 163: 555-558, 1987.

52. Swett HA, Fisher PR, Cohn AI, Miller PI, Mutalik PG: Expert system controlled image display. Radiology 172: 487-493, 1989.
53. Chan HP, Doi K, Galhotra S, Vyborny CJ, MacMahon H, Jokich PM: Image feature analysis and computer-aided diagnosis in digital radiography. 1. Automated detection of microcalcifications in mammography. Med Phys 14: 538-548, 1987.
54. Chan HP, Doi K, Vyborny CJ, Lam KL, Schmidt RA: Computer-aided detection of microcalcifications in mammograms: Methodology and preliminary clinical study. Invest Radiol 23: 664-671, 1988.
55. Chan HP, Doi K, Vyborny CJ, Schmidt RA, Metz CE, Lam KL, Ogura T, Wu Y, MacMahon H: Improvement in radiologists' detection of clustered microcalcifications on mammograms: The Potential of computer-aided diagnosis. Invest Radiol 25: 1102-1110, 1990.
56. Nishikawa RM, Doi K, Giger ML, Yoshimura H, Wu Y, Vyborny CJ, Schmidt RA, Chan HP: Use of morphological filters in the computerized detection of microcalcifications in digital mammograms. Medical Physics 17: 524, 1990.
57. Nishikawa RM, Giger ML, Doi K, Schmidt RA, Vyborny CJ: Automated detection of microcalcifications on mammograms: New feature-extraction techniques with morphologic filters. Radiology 177(P): 288, 1990.
58. Nishikawa RM, Giger ML, Doi K, Vyborny CJ, Schmidt RA: Computer-aided detection of clustered microcalcifications on digital mammograms. Medical and Biological Engineering and Computing (submitted), 1992.
59. Wu Y, Doi K, Giger ML, Nishikawa RM: Computerized detection of clustered microcalcifications in digital mammograms: applications of artificial neural networks. Radiology (in press) 1992.
60. Giger ML, Yin F-F, Doi K, Metz CE, Schmidt RA, Vyborny CJ: Investigation of methods for the computerized detection and analysis of mammographic masses. Proc. SPIE 1233: 183-184, 1990.
61. Yin F-F, Giger ML, Doi K, Metz CE, Vyborny CJ, Schmidt RA: Computerized detection of masses in digital mammograms: Analysis of bilateral-subtraction images. Medical Physics 18: 955-963, 1991.
62. Yin F-F, Giger ML, Doi K, Vyborny CJ, Schmidt RA: Computerizd detection of masses in digital mammograms: Investigation of feature-extraction techniques. Med Phys (submitted) 1992.
63. Wu Y, Giger ML, Doi K, Vyborny CJ, Schmidt RA, Metz CE: Application of neural networks in mammography: Applications in decision making in the diagnosis of breast cancer. Radiology (submitted) 1992.
64. Giger ML, Nishikawa RM, Doi K, Yin FF, Vyborny CJ, Schmidt RA, Metz CE, Wu Y, MacMahon H, Yoshimura H: Development of a "smart" workstation for use in mammography. Proc. SPIE 1445: 101-103, 1991.
65. Doi K, Giger ML, MacMahon H, Hoffmann KR, et al.: Computer-aided diagnosis: development of automated schemes for quantitative analysis of radiographic images. Seminars in Ultrasound, CT and MR (in press).
66. Giger ML, Doi K, MacMahon H, Nishikawa RM, Hofmann KR, et al.: An "intelligent" workstation for computer-aided diagnosis". RadioGraphics (submitted) 1992.
67. MacMahon H, Doi K, Chan HP, Giger ML, Katsuragawa S, Nakamori N: Computer-aided diagnosis in chest radiology. J Thoracic Imaging 5: 67-76, 1990.
68. Metz CE: ROC methodology in radiologic imaging. Invest Radiol 21: 720-733, 1986.
69. Giger ML, Doi K, MacMahon H: Image feature analysis and computer-aided diagnosis in digital radiography. 3. Automated detection of nodules in peripheral lung fields. Med. Phys. 15: 158-166, 1988.
70. Giger ML, Ahn N, Doi K, MacMahon H, Metz CE: Computerized detection of pulmonary nodules in digital chest images: Use of morphological filters in reducing false-positive detections. Med Phys 17: 861-865, 1990.
71. Giger ML, Doi K, MacMahon H, Metz CE, Yin FF: Computer-aided detection of pulmonary nodules in digital chest images. RadioGraphics 10: 41-52, 1990.
72. Yoshimura H, Giger ML, Doi K, MacMahon H, Montner SM: Computerized scheme for the detection of pulmonary nodules: A nonlinear filtering technique. Invest Radiol 27: 124-129, 1992.

73. Katsuragawa K, Doi K, et al.: Image feature analysis and computer-aided diagnosis in digital radiography: Effect of digital parameters on the accuracy of computerized analysis of interstitial disease in digital chest radiographs. Med Phys 17: 72078, 1990.
74. Katsuragawa S, Doi K, et al.: Quantitative computer-aided analysis of lung texture in chest radiographs. RadioGraphics 10: 257-269, 1990.
75. Nakamori N, Doi K, Sabeti V, MacMahon H: Image feature analysis and computer-aided diangosis in digital radiography: Automated analysis of sizes of hearts and lung in digital chest images. Med Phys 17: 342-350, 1990.
76. Fujita H, Doi K, Giger ML, Chan HP: Investigation of basic imaging properties in digital radiography. 5. Characteristic curves of I.I.-TV digital systems. Medical Physics 13: 13-18, 1986.
77. Giger ML: Film digitization—technical requirements. Invited for presentation. In: Proceedings of Chest Imaging Conference '87, (Eds.) Peppler W, Alter A, Medical Physics Publishing Corp. Madison, Wisconsin, pp. 92-100, 1988.
78. Giger ML, Doi K: Investigation of basic imaging properties in digital radiography. 1. Modulation transfer function. Medical Physics 11: 287-295, 1984.
79. Fujita H, Doi K, Giger ML: Investigation of basic imaging properties in digital radiography. 6. MTFs of I.I.-TV digital systems. Med Phys 12: 713-729, 1985.
80. Fujita H, Giger ML, Doi K: Investigation of basic imaging properties in digital radiography. 12. Effect of matrix configuration on system resolution. Medical Physics 15: 384-390, 1988.
81. Fraser RG, Sanders C, Barnes GT, MacMahon H, Giger ML, Doi K, Templeton AW, Cox GG, Dwyer SJ, Merritt C, Jones J: Digital imaging of the chest: state of the art. Radiology 171: 297-307, 1989.
82. Cook LT, Giger ML, Batnitzky S, Wetzel LH, Murphey MD: Digitized film radiography. Investigative Radiology 24: 910-916, 1989.
83. Montner S, Xu X-W, Tsuzaka M, Doi K, MacMahon H, Yoshimura H, Sanada S, Giger ML, Yin F-F: Evaluation of basic imaging properties of a new digital chest system. Proc. SPIE 1231: 390-393, 1990.
84. Sanada S, Doi K, Xu X-W, Yin F-F, Giger ML, MacMahon H: Comparison of imaging properties of a computed radiography system and screen-film systems. Medical Physics (submitted).
85. Yin F-F, Giger ML, Doi K: Measurement of the presampling MTF of film digitizers using a curve fitting technique. Medical Physics (in press).
86. Yin FF, Giger ML, Doi K, Yoshimura H, Xu XW, Nishikawa RM: Evaluation of imaging properties of a laser film digitizer. Phys. Med. Biol. 37: 273-280, 1992.
87. Giger ML, Doi K, Metz CE: Investigation of basic imaging properties in digital radiography. 2. Noise Wiener Spectrum. Medical Physics 11: 797-805, 1984.
88. Giger ML, Doi K, Fujita H: Investigation of basic imaging properties in digital radiography. 7. Noise Wiener spectra of I.I.-TV digital imaging systems. Medical Physics 13: 131-138, 1986.
89. Giger ML, Ohara K, Doi K: Effect of quantization on digitized noise and detection of low-contrast objects. Proc. SPIE 626: 214-224, 1986.
90. Kume Y, Doi K, Ohara K, Giger ML: Investigation of basic imaging properties in digital radiography. 10. Structure mottle of I.I.-TV digital imaging systems. Medical Physics 13: 843-849, 1986.
91. Giger ML: Image quality: Effects of digitization, matrix size and noise. Invited for presentation at the 1987 AAPM Annual Summer School, In: Image Communication and Image Analysis. (Mulvaney, Haus, Windham, eds.), New York, AIP (in press) .
92. Giger ML, Doi K: Investigation of basic imaging properties in digital radiography. 3. Effect of Pixel Size on SNR and Threshold Contrast. Med. Phys. 12: 201-208, 1985.
93. Giger ML, Doi K: Effect of pixel size on detectability of low-contrast signals in digital radiography. J of the Optical Society of America A 4: 966-975, 1987.
94. Ohara K, Chan HP, Doi K, Giger ML, Fujita H: Investigation of basic imaging properties in digital radiography. 8. Detection of simulated low-contrast objects in DSA images. Medical Physics 13: 304-311, 1986.

95. Ohara K, Doi K, Metz CE, Giger ML: Investigation of basic imaging properties in digital radiography. 13. Effect of structured noise on the detectability of simulated stenotic lesions. Medical Physics 16:14-21, 1989.
96. MacMahon H, Metz CE, Doi K, Kim T, Giger ML, Chan H-P: The effect of display format on diagnostic accuracy in digital chest radiography: A comparison of hardcopy, video, and reversed grey scale. Radiology 168: 669-673, 1988.
97. MacMahon H, Doi K, Sanada S, Montner SM, Giger ML, et al.: Effect of data compression on diagnostic accuracy in digital chest radiography. An ROC study. Radiology (submitted).
98. MacMahon H, Sanada S, Doi K, Giger ML, Xu X-W, Yin F-F, Montner SM, Carlin M: Direct comparison of conventional and computed radiography with a dual image recording technique. RadioGraphics (submitted).
99. Asada N, Doi K, MacMahon H, Montner S, Giger ML, Abe C, Wu Y: Potential usefulness of artificial neural network for differential diagnosis of interstitial lung diseases: a pilot study. Radiology (in press).
100. Murphy WA Jr, Destouet JM, Monsees BS: Professional quality assurance for mammography screening programs. Radiology 175: 319-320, 1990.
101. Bird RE: Professional quality assurance for mammography screening programs. Radiology 177: 587, 1990.
102. Brenner RJ: Medicolegal aspects of breast imaging: variable standards of care relating to different types of practice. AJR 156: 719-723, 1991.
103. Pratt WK: Digital Image Processing (Wiley, New York, 1978).
104. Serra J: Image Analysis and Mathematical Morphology, (Academic Press, New York), 1982.
105. Rumelhart DE, Hinton GE, Williams RJ: Learning internal representations by error propagation. In: Rumelhart DE, McClelland JL, PDP Research Group. Parallel distributed processing: Explorations in the microstructure of cognition. Cambridge: MIT Press 1: 318-362, 1986.
106. Kendall M, Stuart A: The Advanced Theory of Statistics, Vol. 3 (Hufner, New York), 1976.
107. Metz CE: Some practical issues of experimental design and data analysis in radiological ROC studies. Invest. Radiol. 24: 234-245, 1989.
108. Dorfman DD, Alf E: Maximum likelihood estimation of parameters of signal detection theory and determination of confidence intervals-rating method data. J Math Psych 6: 487-498, 1969.
109. Bunch PC, Hamilton JF, et al.: A free response approach to the measurement and characterization of radiographic observer performance. Proc. SPIE 127: 124-135, 1977.
110. Chakraborty D: Maximum likelihood analysis of free-response receiver operating characteristic (FROC) data. Med. Phys. 16: 561-568, 1989.

APPENDIX

Automated Segmentation of Digitized Mammograms

Ulrich Bick, MD, Maryellen L. Giger, PhD, Robert A. Schmidt, MD, Robert M. Nishikawa, PhD, Dulcy E. Wolverton, MD, Kunio Doi, PhD

Rationale and Objectives. Fast and reliable segmentation of digital mammograms into breast and nonbreast regions is an important prerequisite for further image analysis. We are developing a segmentation algorithm that is fully automated and can operate independent of type of digitizing system, image orientation, and image projection.

Methods. The algorithm identifies unexposed and direct-exposure image regions and generates a border surrounding the valid breast region, which can then be used as input for further image analysis. The program was tested on 740 digitized mammograms; the segmentation results were evaluated by two expert mammographers and two medical physicists.

Results. In 97% of the mammograms, the segmentation results were rated as acceptable for use in computer-aided diagnostic schemes. Segmentation problems encountered in the remaining 22 images (2.9%) were most often caused by digitization artifacts or poor mammographic technique.

Conclusion. The developed algorithm can serve as a component of an "intelligent" workstation for computer-aided diagnosis in mammography.

Key Words. Computer-aided diagnosis; digital mammography; image segmentation; image processing; digitization.

The advent of digital projection radiography, either as a direct digital modality (e.g., computed radiography) or as film digitization, has opened a variety of new opportunities including digital image processing, digital image storage and transfer, and computer-aided image analysis [1]. For any type of automatic image analysis, it is necessary to first identify a region of interest (ROI; e.g., the breast region in a mammogram). In many previous studies of computer-aided diagnosis (CAD) in mammography, analysis was based on manually selected ROIs [2–6]. Semmlow et al. [7] described a method that automatically detects the breast skin line in xeromammograms with the use of edge detection. However, because of the different image characteristics of xeromammograms, this method is not directly applicable to screen-film mammograms. As part of our CAD scheme in mammography, we previously developed a method for identifying the breast region in mammo-

From the Kurt Rossmann Laboratories for Radiologic Image Research, Department of Radiology, the University of Chicago, Chicago, IL.

Address reprint requests to U. Bick, MD, who is now at the Department of Radiology, University of Münster, Albert-Schweitzer-Str. 33, 48129 Münster, Germany.

Received March 29, 1994, and accepted for publication after revision September 15, 1994.

Acad Radiol 1995;2:1–9

grams on the basis of a global histogram analysis [8, 9]. Histogram-derived upper and lower thresholds are used to separate the breast region pixels from the dark direct-exposure and bright unexposed image background. A similar approach has been used by Lau et al. [10]. However, a method that is based on global histogram thresholding alone is critically dependent on the threshold selection process and may not be robust. Problems commonly encountered with this method include the following: overlap of the direct-exposure and object pixel values attributable to a nonuniform background (e.g., the heel effect in mammograms); misclassification of intermediate density pixels in the transition zone between direct-exposure and unexposed image areas; and erroneous exclusion of bright object areas if the image does not contain any unexposed image region. The latter may occur in mammography if the cone size is equal to or larger than the film size and the histogram peak corresponding to the bright pectoralis muscle region is mistaken for an unexposed region peak.

Davies and Dance [11] used a histogram-derived threshold in conjunction with a mode filter to exclude uniform background areas from the image analysis. Chen et al. [12] described an algorithm that detects the skin line edge on the basis of a combination of histogram analysis and a Laplacian edge detection operator. Karssemeijer [13] used a fixed threshold in change-coupled device-camera digitized mammograms to identify the skin line and an edge filter to detect the chest side border of the breast region. However, to our knowledge, none of these methods has been tested on a large number of randomly selected clinical mammograms or on images digitized on a variety of film digitizers. The purpose of our study was to develop a new, fully automated segmentation algorithm that is able to reliably identify the image breast region independent of digitization system, image orientation, and image projection.

MATERIALS AND METHODS

Digitized routine clinical screen-film mammograms were used in this study. These films were collected from three different institutions over an 8-year period (1985–1993). Images were digitized with three different digitizers: one optical drum scanner (system A; FIP-II, Fuji, Kyoto, Japan) and two laser scanners (system B [KFDR-S, Konica, Tokyo, Japan] and system C [LD-4500, Konica, Tokyo, Japan]). Sampling distance, quantification, matrix sizes, and dynamic range of the different digitizers are shown in Table 1. Details concerning the imaging properties of these digitizers have been described elsewhere [14, 15].

As input to the segmentation algorithm, digital images, which were subsampled with a default sampling distance of roughly 2 mm (corresponding to matrix sizes ranging from 128 × 128 to 128 × 162) and a 10-bit gray-value resolution (high pixel values representing low optical densities in the original image), were generated. The program was implemented in C on a high-speed RISC workstation (IBM Powerstation 570, RISC 6000 Series, IBM, Austin, TX).

Segmentation Algorithm

The individual steps of the segmentation algorithm are outlined in Table 2. Elimination of digitizer line and isolated pixel artifacts, as well as an overall noise reduction, was achieved by an initial 3 × 3 median filtering step.

On the basis of a local gray-value range and modified histogram analysis, we classified each pixel in the image into one of the following categories: (1) unexposed image region (part of the image outside the radiation cone);

TABLE 1: Description of Digitizer Systems

Parameter	System A	System B	System C
Model	FIP-II (Fuji, Kyoto, Japan)	KFDR-S (Konica, Tokyo, Japan)	LD-4500 (Konica, Tokyo, Japan)
Type	Optical drum scanner	Laser scanner	Laser scanner
Matrix size	2000 × 2500 (20.32 × 25.4 cm) 2500 × 3000 (25.4 × 30.48 cm)	2000 × 2600 (20.32 × 25.4 cm) 1880 × 2270 (25.4 × 30.48 cm)	2048 × 2580 (20.32 × 25.4 cm) 2048 × 2472 (25.4 × 30.48 cm)
Sampling distance	100 μm	87 μm (20.32 × 25.4 cm) 131 μm (25.4 × 30.48 cm)	96 μm (20.32 × 25.4 cm) 121 μm (25.4 × 30.48 cm)
Quantification	10 bit	10 bit	10 bit
Dynamic range	0.2–2.75 OD	0–2.8 OD	0–3.5 OD

OD = Optical density.

TABLE 2: Outline and Performance of Segmentation Algorithm

Algorithm Step	CPU Time (sec)[a]
Noise filter	0.3
Calculation of local gray-value range	0.7
Modified global histogram analysis	0.1
Classification of image pixels	0.1
Region growing	0.3
Morphologic filtering	0.2–0.4
Determination of object contour	0.7–1.0
Total performance time	2.4–2.9

CPU = central processing unit.

[a]CPU time on an IBM 570 for 128 × 160 subsampled matrix excluding image data input and output.

(2) directly exposed image region; or (3) potential object (in this case, the breast) pixel (Fig. 1B–D). The local range operator used in our algorithm was based on a 7-pixel-wide ring of 16 pixels. From this neighborhood, the local maximum and minimum pixel values were calculated. A modified, "selective" histogram [16] was constructed including only pixels with a small local range (local maximum minus local minimum), as shown in Figure 2. For a pixel to be classified as a direct-exposure pixel, the following criteria had to be fulfilled: A direct-exposure peak exists in the modified global histogram; the pixel value is close to this direct exposure peak; and

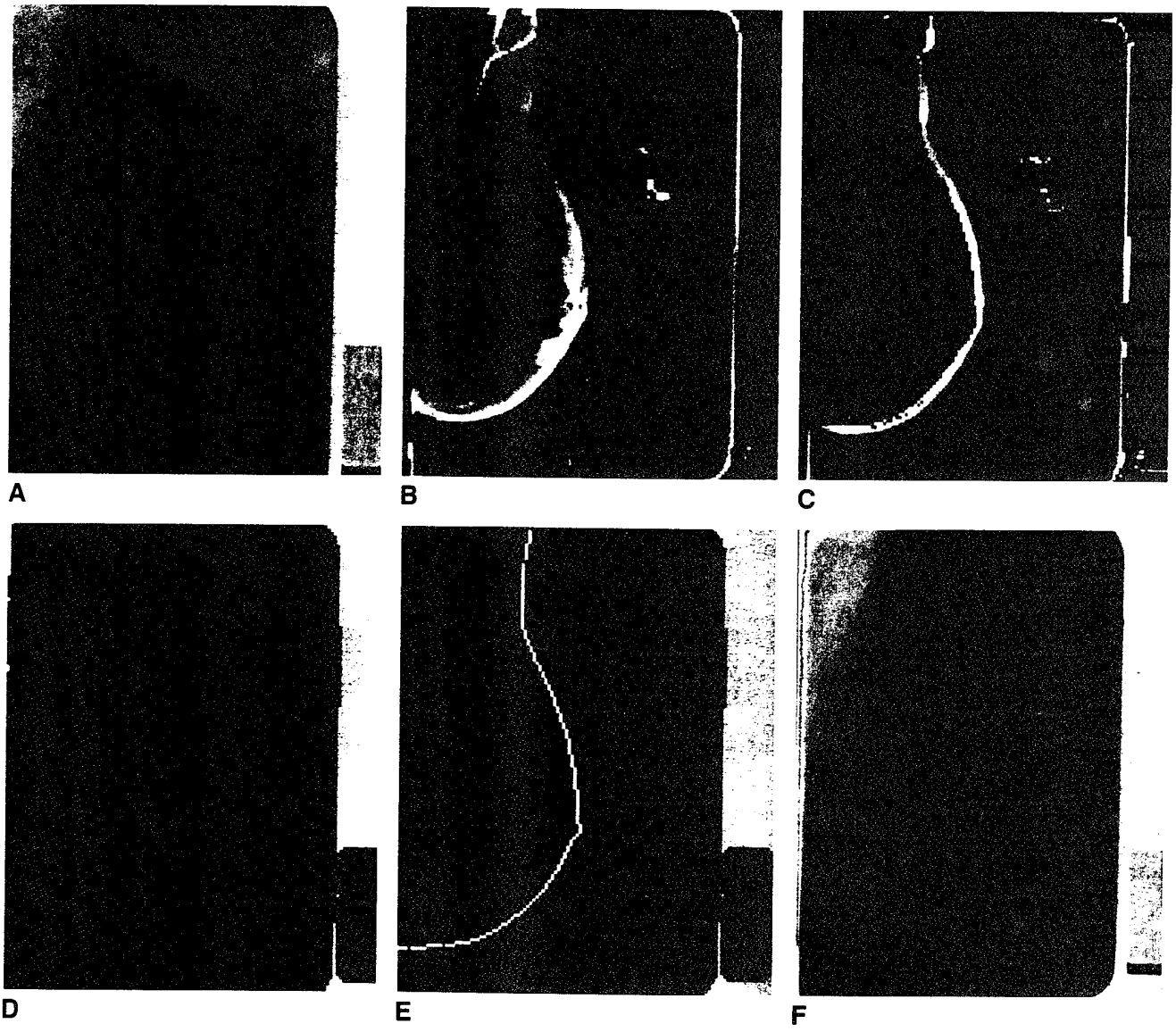


FIGURE 1. Segmentation of digital mammograms. ***A***, Original digital mammogram. ***B***, Local gray-value range (local maximum minus local minimum) image. ***C***, Range image with intermediate density pixels inside the breast already identified as object pixels by the modified global histogram analysis shown as dark gray. ***D***, Image after initial pixel classification based on local gray-value range and modified global histogram analysis. *Black* = direct exposure, *gray* = potential object pixels, and *white* = unexposed image region. Note that there is a transition zone of gray potential object pixels along the edge between the direct-exposure and unexposed image region. ***E***, Computer-generated breast border. *Arrowheads* mark the connection points from the internal object border (between object and direct-exposure region) to the external object border (between object and unexposed image region). ***F***, Computer-generated breast border superimposed on the original image.

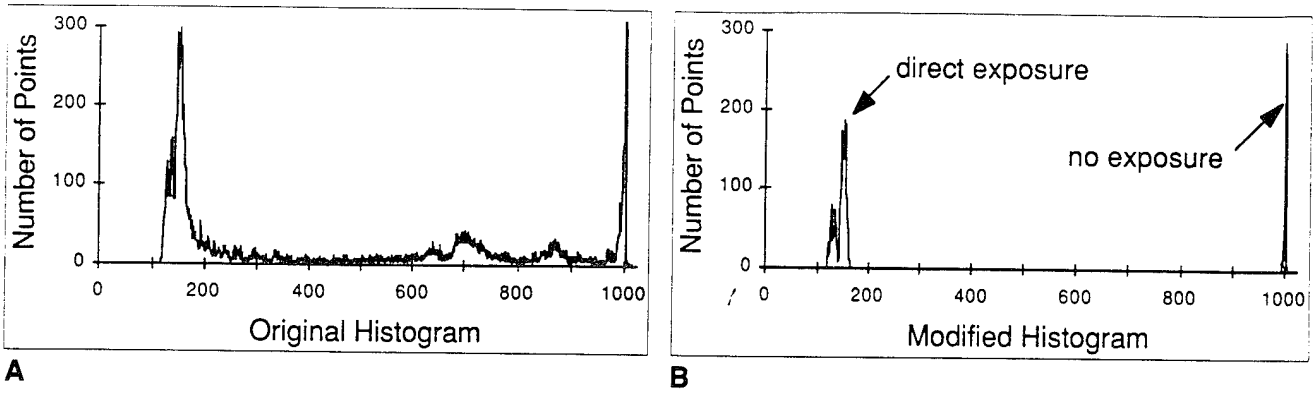


FIGURE 2. Modified global histogram analysis. ***A***, Global histogram of entire mammogram shown in Figure 1A. ***B***, Modified global histogram after exclusion of all pixels with a large local gray-value range. Peaks from the uniform direct-exposure and unexposed image regions are clearly identified with almost no contribution from the intermediate density breast pixels.

the difference between the center pixel value and the local minimum pixel value is below a certain threshold. Accordingly, to be classified as a pixel belonging to an unexposed image region, the following criteria had to met: There is a corresponding modified global histogram peak; the pixel value is close to this histogram peak value; and the difference between the local maximum pixel value and the center pixel value is below a certain threshold. The remaining pixels, not classified as belonging to either the direct-exposure or the unexposed image region, were considered as potential breast pixels. Figure 1B–D illustrates how pixel classification was achieved as a combination of local range and global modified histogram analysis. By using region growing with 4-point connectivity [17], we could identify the different image regions. Small isolated areas of either direct-exposure or unexposed pixels inside the breast region were assumed to be misclassified and were therefore removed.

We next used a morphologic filtering step, which eliminates minor irregularities along the outside breast contour and typical artifacts resulting from background inhomogeneity. Such artifacts may be seen as a band of misclassified pixels up to 3 pixels wide along a sharp edge between direct-exposure regions of different densities (Fig. 3). The maximum artifact width of 3 pixels depends on the specific configuration of the gray-value range operator with a width of 7 pixels. Using two perpendicular 5×1 pixel structuring elements, thin lines of object pixels up to 3 pixels in width in either the x or y direction with direct-exposure pixels on both sides were eliminated and reclassified as direct exposure (Fig. 3).

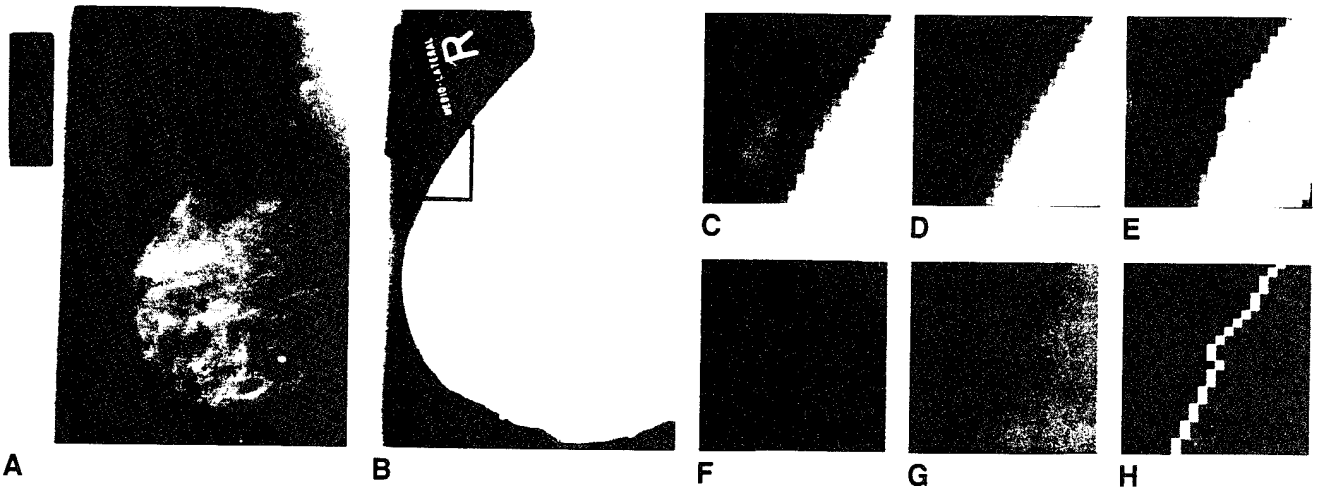


FIGURE 3. Illustration of the use of morphologic filtering to eliminate artifacts caused by background inhomogeneity. Mammogram digitized with system C laser scanner displayed with normal (***A***) and narrow direct-exposure (***B***) window setting. Note band of darker direct-exposure background pixels in scanning direction behind the identification label. ***C–H***, Enlarged region of interest: original image (***C***), after median filtering (***D***), range image (***E***), after pixel classification (***F***), 3-pixel-wide artifact along the edge between direct-exposure areas of different density removed by morphologic filtering step (***G***), and computer-generated breast border (***H***).

In the final step, a closed, 8-point connected border defining the breast region was generated (Fig. 1E). Because, in most cases, a transition zone of intermediate density object pixels is found along the edge between direct-exposure and unexposed image regions, the border generation algorithm has to identify certain "connection" points, where the object border is allowed to connect from the internal object border (between object and direct-exposure region) to the external object border (between object and unexposed region). Potential connection points are identified as points that fulfill the following two criteria: (1) a short connected path of object pixels exists between the connection point along the internal object border and the outside unexposed region and (2) the internal object border forms a concave angle at the connection point, which is smaller than a certain threshold. If more than one isolated object region exists in an image (additional "objects" may represent, for example, letters or the identification label), the breast region can be identified easily as the largest region of connected object pixels. The generated breast border is then expanded by linear interpolation to the original image matrix and smoothed using a running average of the border coordinates. Figure 1F shows the final computer-generated breast border superimposed on the original mammogram.

Evaluation

The testing database consisted of 740 routine clinical screen-film mammograms, including 373 mediolateral oblique and 367 craniocaudal views. One hundred twenty-one images were digitized with the optical drum scanner system A, 350 images with the laser scanner system B, and 269 images with the newer laser scanner system C (for a description of the digitizers, see Table 1). The program was run on all 740 images with a fixed default parameter setting. The computer-generated breast border was superimposed on the original image and displayed on a computer monitor. The segmentation results were subjectively rated by two expert mammographers and two medical physicists and were categorized as follows: (1) optimal—deviations of the computer-generated border from the "true" breast border of less than the sampling distance of 2 mm; (2) minor localized deviations; (3) readily visible deviations—however, results still acceptable for CAD purposes (e.g., no breast parenchymal tissue excluded); (4) substantial deviations—however, overall segmentation is still correct (may influence results of CAD schemes); and (5) complete failure of segmentation (likely to influence CAD results). Examples of minor (category 2) and acceptable (category 3) deviations are shown in Figure 4. During the evaluation, the observers were able to choose between different default window settings as well as manually adjust the window in order to better assess the performance of the segmentation. A chi-square test was used for statistical analysis of the results.

RESULTS

Results, shown in Figure 5, indicate that in more than 97% of the cases, the segmentation results were rated as acceptable for CAD purposes (category 1, 2, or 3). No significant differences in rating ($p = .12$) were found between mammographers and physicists (Fig. 5B). In 22 images (2.9%), the segmentation results were considered unsatisfactory (rated as category 4 or 5 by at least two observers). The most common causes of seg-

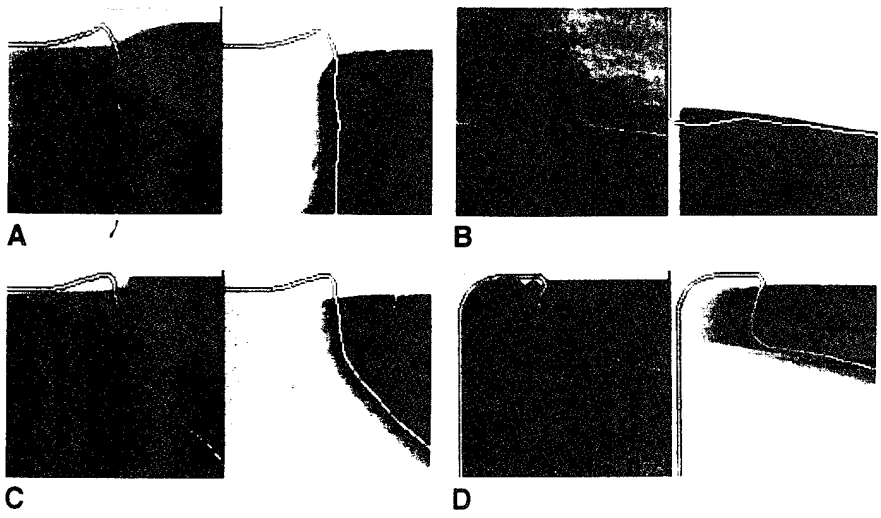


FIGURE 4. Evaluation of segmentation results. ***A*** and ***B***, Examples of minor localized deviations from the "true" breast border (category 2). ***C*** and ***D***, Deviations considered acceptable for computer-aided diagnostic purposes (category 3). All images are displayed in two different window settings with a "normal" wide window (left side) and a second narrow window (right side) showing the dark peripheral breast portion. Note that minor deviations along the skin line can be assessed only on the narrow window image.

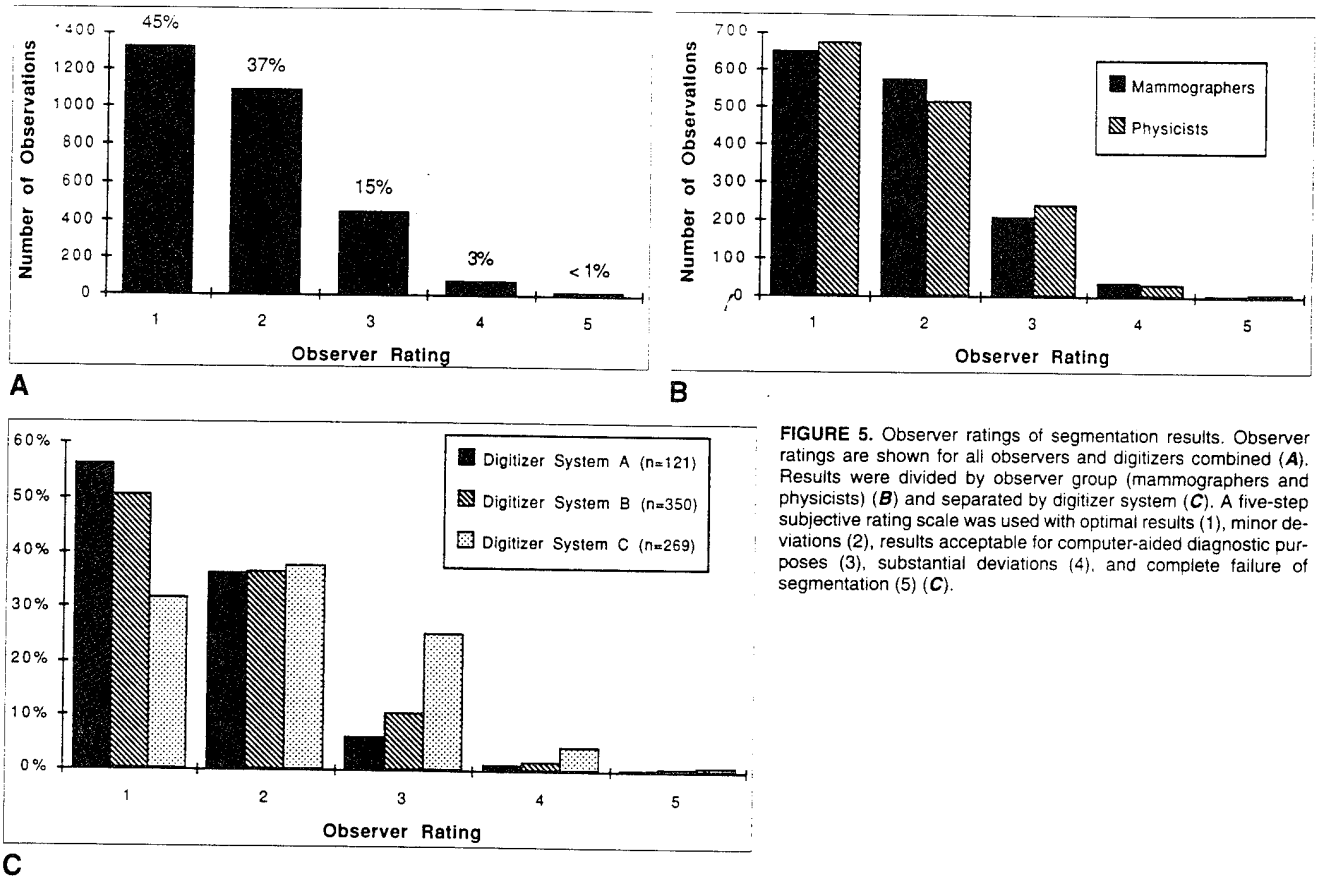


FIGURE 5. Observer ratings of segmentation results. Observer ratings are shown for all observers and digitizers combined (***A***). Results were divided by observer group (mammographers and physicists) (***B***) and separated by digitizer system (***C***). A five-step subjective rating scale was used with optimal results (1), minor deviations (2), results acceptable for computer-aided diagnostic purposes (3), substantial deviations (4), and complete failure of segmentation (5) (***C***).

mentation problems were overlying foreign material (e.g., identification letters), digitizer artifacts, and poor mammographic technique, which alone accounted for 18 of the 22 found (Table 3). Two examples of segmentation problems caused by poor mammographic technique are shown in Figure 6.

Significant differences in segmentation results were observed among the different digitizers ($p < .001$). Segmentation problems (both minor deviations as well as number of segmentation failures) were more common in images digitized with the newer system C laser scanner (Fig. 5C). This scanner, which was the only one with a

TABLE 3: Unsatisfactory Segmentation Results Found in 22 of 740 Images

Type of Segmentation Problem	*n*	Possible Explanation
Inclusion of overlying foreign material	8	Poor positioning
Poor definition of skin line	5	Loss of skin line during digitization
Inclusion of outside material along the edges	5	Small incorrect cone
Exclusion of pectoralis muscle	1	Gray-value range threshold too large
Exclusion of breast tissue along the edges	3	Error of connecting algorithm

Results were considered unsatisfactory if they were rated as category 4 or 5 by at least two observers.

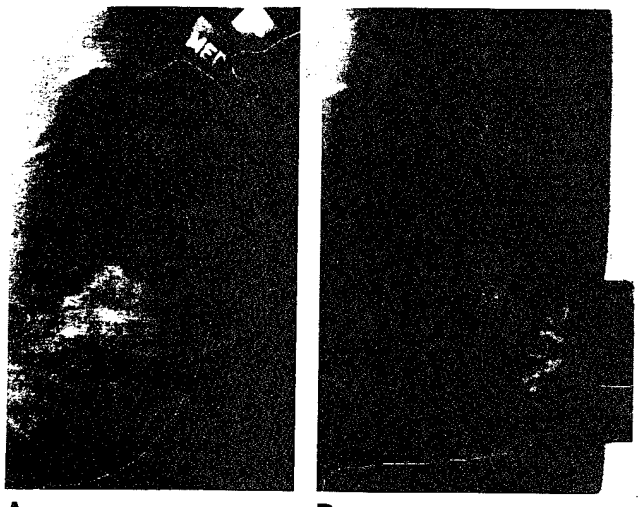


FIGURE 6. Illustration of segmentation problems caused by poor mammographic technique. Segmentation algorithm distracted by placement of letters along the partially included arm (***A***) and a small cone with identification label touching the breast contour (***B***).

dynamic range extending above 3 optical density units, created two typical artifacts that often interfered with the segmentation. In almost all images, a band of pixels with a higher signal intensity, up to 200 pixels (2 cm) in width, was found along the posterior edge of the direct-exposure area (Fig. 7). In some instances, this artifact was so severe that it completely masked the adjacent breast border. This problem was overcome by allowing the final border generation step to connect through such a transition zone of intermediate pixels with a certain maximum width, as described earlier (Fig. 7F). The second artifact was a region of lower signal intensity pixels in the direct-exposure area, which was found only in scanning lines that included the relatively dark identification label (Fig. 3). This led to misclassification of pixels with a large local gray-value range along the edge of this darker direct-exposure region. However, in most instances, these misclassified pixels could be eliminated by the morphologic filtering step (Fig. 3). Such artifacts were not found in the older system B laser scanner or in the system A optical drum scanner. In both of these latter digitizers, however, the small dynamic range often led to poor definition or complete loss of the skin line.

DISCUSSION

To be integrated into an automated, real-time radiographic CAD system, a segmentation algorithm must be fully automated, fast, reliable, and independent of the specific imaging condition (e.g., imaging system, type of image object, image orientation, and exposure conditions). Our proposed algorithm—a combination of a modified global histogram analysis, a gray-value range operator, and region growing—has been shown to fulfill these conditions. With a central processing unit time of 2–3 sec (Table 2), it is fast enough to be implemented in a real-time system. The program does not require any user interaction, and the only prior information necessary for operation is the image pixel size, which is usually included in the image file header after digital image acquisition. In our study of 740 routine clinical mammograms from different sources, 97% of the segmentation results were rated as acceptable for CAD purposes. When analyzing these results, one must remember that the described default program parameters (filter and range operator kernel size, segmentation image matrix), which were held constant throughout the testing, are a compromise between speed and accuracy.

Current mammographic screen-film systems with background optical densities approaching 4 [18, 19] pose a considerable challenge for film digitization systems. Only recently have new laser scanners been developed for medical imaging that are capable of digitizing film with optical densities of 3 or more [20]. In older systems, the dark peripheral parts of the breast and the skin line are often lost or indistinct because of the small dynamic range and a significant increase in digitizer noise in dark image areas [21–24]. Among the digitizer systems used in our study, only the newer system C laser scanner had a dynamic range including optical densities of more than 3 (Table 1). However, this was coupled with typical artifacts in dark image areas, which frequently interfered with the segmentation process (Figs. 3 and 7). These problems may be overcome with new improved digitizer systems [20] or by direct digital mammography [25, 26].

Our algorithm creates an initial raw segmentation of the image and is designed to operate in conjunction with an automatic evaluation of the segmentation results and an optional local contour optimization as shown in Figure 8.

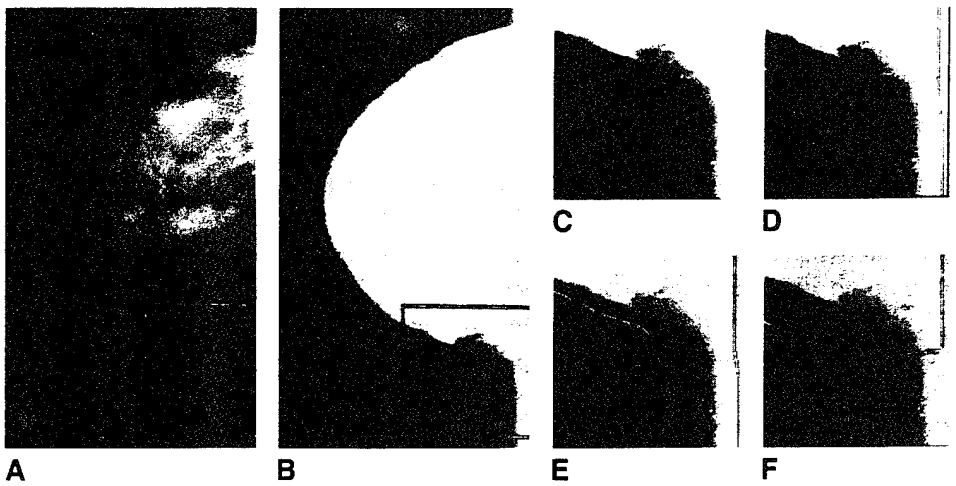


FIGURE 7. Typical example of a system C digitizer artifact. Digitized mammogram displayed in normal (***A***) and narrow direct-exposure (***B***) window setting. Note band of pixels with increased density along the posterior edge of the direct-exposure area. ***C–F***, Enlarged region of interest: enlarged original (***C***), border generated with default parameter setting but without being allowed to connect through artifact area (***D***), increased gray- value range threshold (***E***), and after use of the connecting algorithm (***F***). Because of the higher edge strength of the artifact, an increase of the gray-value range threshold led to loss of the skin line (***E***) before the artifact area was eliminated.

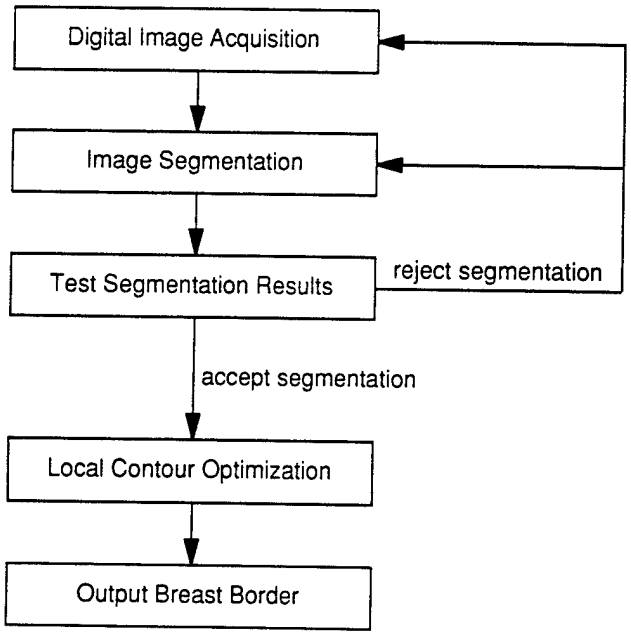


FIGURE 8. Flowchart demonstrating the use of the segmentation algorithm in conjunction with a segmentation result evaluation step and an optional local contour optimization step.

By analyzing the shape and smoothness of the computer-generated object contour, and by analyzing the resulting histogram containing only object pixels, it is possible to automatically evaluate the segmentation results. If the unsatisfactory results are caused by poor image technique, the image acquisition may be repeated or the program may be run with a different parameter set. Once the segmentation is acceptable, the object border information can be used in different CAD schemes, such as for bilateral alignment of breast images in the detection of breast masses [8, 9]. Prior to this, the object contour also may be locally optimized using a Laplacian of Gaussian or other second-derivative operator. This is necessary, for example, in the analysis of skin abnormalities because an accuracy of ±2 mm (the default pixel size used in our segmentation) is inadequate for this purpose, considering the normal skin thickness of 2–4 mm [27, 28].

Apart from CAD, the image segmentation provided by our algorithm is also useful in enhancement for digital image display. For viewing radiographic images on a computer monitor, it is useful to eliminate the bright, unexposed outside image areas, which may interfere with the detection of subtle density abnormalities in the displayed radiographic images [29, 30]. In some radiographic workstations, it is therefore possible to manually shade the bright outside image areas to focus on the actual object area. With our algorithm, which automatically identifies the object border, it is possible to set all outside pixels to a specified dark pixel value. By excluding background pixels from the histogram evaluation, this algorithm may also increase the reliability and accuracy of histogram-based automatic exposure correction or image enhancement schemes [31–33].

ACKNOWLEDGMENTS

The authors are grateful to Charles E. Metz, PhD, for his constructive contributions regarding the statistical analysis. The authors are also grateful to Carl J. Vyborny, MD, for his helpful discussions and to Ping Lu, MS, for her assistance during film digitization and computer evaluation.

This work is being done as part of the National Digital Mammography Development Group study, which was established and is funded in part by the National Cancer Institute and the National Institutes of Health. This research was supported in part by the American Cancer Society (grant FRA-390), the United States Public Health Service (grants CA48985 and CA24806), and the U.S. Army Medical Research and Development Command (grant 92153010). The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of any of the supporting organizations.

REFERENCES

1. Doi K, Giger ML, Nishikawa RM, et al. Digital radiography: a useful tool for computer-aided diagnosis by quantitative analysis of radiographic images. *Acta Radiol* **1993**;34:426–439.
2. Ackerman LV, Gose EE. Breast lesion classification by computer and xeroradiograph. *Cancer* **1972**;30:1025–1035.
3. Davies DH, Dance DR. The automatic computer detection of subtle calcifications in radiographically dense breasts. *Phys Med Biol* **1992**;37:1385–1390.
4. Dengler J, Behrens S, Desaga JF. Segmentation of microcalcifications in mammograms. *IEEE Trans Med Imag* **1993**;12:634–642.
5. Dhawan AP, Chitre Y, Moskowitz M. Artificial neural network based classification of mammographic microcalcifications using image structure features. *Proc Soc Photo-Op Instrum Eng* **1993**;1905:820–831.
6. Magnin IE, El Alaoui M, Bremond A. Automatic microcalcifications pattern recognition from X-ray mammographies. *Proc Soc Photo-Op Instrum Eng* **1989**;1137:170–175.
7. Semmlow JL, Shadagopappan A, Ackerman LV, Hand W, Alcorn FS. A fully automated system for screening xeromammograms. *Comput Biomed Res* **1980**;13:350–362.
8. Yin F-F, Giger ML, Doi K, Metz CE, Vyborny CJ, Schmidt RA. Computerized detection of masses in digital mammograms: analysis of bilateral subtraction images. *Med Phys* **1991**;18:955–963.
9. Yin F-F, Giger ML, Doi K, Vyborny CJ, Schmidt RA. Computerized detection of masses in digital mammograms: automated alignment of breast images and its effect on bilateral-subtraction technique. *Med Phys* **1994**;21:445-452.
10. Lau T-K, Bischof WF. Automated detection of breast tumors using the asymmetry approach. *Comput Biomed Res* **1991**;24:273–295.
11. Davies DH, Dance DR. Automatic computer detection of clustered calcifications in digital mammograms. *Phys Med Biol* **1990**;35:1111–1118.
12. Chen J, Flynn MJ, Rebner M. Regional contrast enhancement and data compression for digital mammographic images. *Proc Soc Photo-Op Instrum Eng* **1993**;1905:752–758.

13. Karssemeijer N. Recognition of clustered microcalcifications using a random field model. *Proc Soc Photo-Op Instrum Eng* **1993**;1905:776–786.
14. Ishida M, Kato H, Doi K, Frank PH. Development of a new digital radiographic image processing system. *Proc Soc Photo-Op Instrum Eng* **1982**; 347:42–48.
15. Yin F-F, Giger ML, Doi K, Yoshimura H, Xu X-W, Nishikawa RM. Evaluation of imaging properties of a laser film digitizer. *Phys Med Biol* **1992**;37: 273–280.
16. Weszka JS. A survey of threshold selection techniques. *Comput Graph Image Process* **1978**;7:259–265.
17. Zucker SW. Region growing: childhood and adolescence. *Comput Graph Image Process* **1976**;5:382–399.
18. Haus A, Gray JE, Daly TR. Evaluation of mammographic viewbox luminance, illuminance, and color. *Med Phys* **1993**;20:819–821.
19. Haus AG. Technologic improvements in screen-film mammography. *Radiology* **1990**;174:628–637.
20. Whiting BR, Muka E, Kocher TE, Flynn MJ. High-resolution, high-performance radiographic film scanner. *Proc Soc Photo-Op Instrum Eng* **1990**; 1231:295–305.
21. Fredfeldt KE, Christensen E, Conradsen K, Ersbøll B, Stedstrup S. Automatic screening of plain film mammography. *Semin Ultrasound CT MR* **1992**;13:135–139.
22. Vranckx J, Strul B. Performance evaluation of a laser digitizer. *Proc Soc Photo-Op Instrum Eng* **1987**;767:524–528.
23. Lo S-CB, Gaskill JW, Mun SK, Krasner BH. Contrast information of digital imaging in laser film digitizer and display monitor. *J Digit Imaging* **1990**;3: 119–123.
24. Halpern EJ, Esser PD, Nickoloff EL, Alderson PO. A quality-control phantom for digitization of radiographs. *J Digit Imaging* **1990**;3:42–48.
25. Nishikawa RM, Mawdsley GE, Fenster A, Yaffe MJ. Scanned-projection digital mammography. *Med Phys* **1987**;14:717–727.
26. Jarlman O, Samuelsson L, Braw M. Digital luminescence mammography: early clinical experience. *Acta Radiol* **1991**;32:110–113.
27. Pope TL, Read ME, Medsker T, Buschi AJ, Brenbridge AN. Breast skin thickness: normal range and causes of thickening shown on film-screen mammography. *J Can Assoc Radiol* **1984**;35:365–368.
28. Bick U, Giger ML, Huo Z, et al. Automated detection of skin thickening in mammograms. In: Lemke HU, Inamura K, Jaffe CC, Felix R, eds., *Computer assisted radiology: proceedings of the international symposium.* Berlin: Springer, **1993**:461–465.
29. McNitt-Gray MF, Pietka E, Huang HK. Image preprocessing for a picture archiving and communication system. *Invest Radiol* **1992**;27:529–535.
30. Blume H, Roehrig H, Browne M, Ji TL. Comparison of the physical performance of high resolution CRT displays and films recorded by laser image printers and displayed on light-boxes and the need for a display standard. *Proc Soc Photo-Op Instrum Eng* **1990**;1232:97–114.
31. Sezan MI, Tekalp AM, Schaetzing R. Automatic anatomically selective image enhancement in digital chest radiography. *IEEE Trans Med Imaging* **1989**;8:154–162.
32. McAdams HP, Johnson GA, Suddarth SA, Ravin CE. Histogram-directed processing of digital chest images. *Invest Radiol* **1986**;21:253–259.
33. Yoshimura H, Xu X, Doi K, et al. Development of a high quality film duplication system using laser digitizer: comparison with computed radiography. *Med Phys* **1993**;20:51–58.

Preprint: Proc. SPIE 2434, 1995

Initial Experience with a Prototype Clinical "Intelligent" Mammography Workstation for Computer-Aided Diagnosis

Robert M Nishikawa, Regina C. Haldemann, John Papaioannou, Maryellen L Giger, Ping Lu,
Robert A Schmidt, Dulcy E Wolverton, Ulrich Bick, and Kunio Doi

Kurt Rossmann Laboratories for Radiologic Image Research,
Department of Radiology, The University of Chicago, Chicago IL 60637

ABSTRACT

This paper reports on the preliminary results of an on-going prospective evaluation of an "intelligent" mammography workstation. This workstation can provide to radiologists a "second opinion" as to the location of suspicious lesions on mammograms. The workstation consists of a high speed computer, film digitizer, image archive, and both hard and soft copy output. Running on the workstation are automated computerized schemes for the detection of breast masses and clustered microcalcifications. In the current study, all screening mammograms are digitized on the workstation and then analyzed by the computerized schemes. The preliminary results for the first 37 days (573 patients) have been analyzed. Although follow-up to establish truth has not been done for all patients, the two schemes detected the lesion in 10 of the 14 patients who had a 'suspicious' lesion present mammographically. Three of the lesions missed by the computer were found to be benign either at biopsy or after further work-up, and fourth one is scheduled for further work-up. For two patients, a cluster of microcalcifications was detected by the computerized scheme that was initially missed by the radiologist. The false positive rate was 1.2 false masses and 0.87 false clusters per image. Over 70% of the false positive masses were caused by nodular densities and approximately 50% of the false cluster included obviously benign calcifications. The results from this ongoing study will be used to plan a full-scale clinical study.

Keywords: computer-aided diagnosis, breast cancer, digital mammography, image analysis, artificial neural network, detection, masses, microcalcifications, workstation

1. INTRODUCTION

Over the past 10 years, our group has been developing automated schemes to assist radiologists in interpreting mammograms. We have performed extensive testing on our schemes which are designed for the detection of masses and clustered microcalcifications. To date, these tests were performed retrospectively on a selected set of mammograms and we have obtained results that indicated that our schemes have potential to be used as an effective aid for radiologists. We are now at the stage in development of our CAD (computer-aided diagnosis) program to prospectively test our schemes on a large number of clinical mammograms. On November 8th, 1994, we implemented an "intelligent" mammography workstation and began the first test of our schemes on clinical mammograms obtained in the mammography section of our department. This paper reports preliminary results from the first 37 days of this prospective testing on a prototype clinical "intelligent" mammography workstation.

2. MATERIALS

The workstation hardware consists of an IBM RISC 6000 PowerStation Model 590, a Konica LD4500 laser film digitizer (0.1-mm pixels, 10 bit), an Alphatronix Inspire 40-GB magneto-optical jukebox, two Imlogix 1024-line monitors, and a Seikosha VP4500 video printer for hard copy.

The "intelligence" of the workstation comes from automated detection schemes for masses and clustered microcalcifications. We have reported extensively on our methods in the past,[1, 2] which are outlined in flowchart form in Fig. 1. The breast area is first segmented from its background using a combination of grey-level thresholding, morphological erosion, and region growing,[3] and then the two detection schemes are run as described below.

The mass detection program uses asymmetries between the left and right breasts as a basis for the identification of breast masses. The mass program works with pairs of images, the left and right mediolateral-oblique views or the left and right craniocaudal views. Each image is subsampled down to an equivalent pixel size of 0.5 mm. The left and right images are aligned based on the skinline and the nipple,[4] and then a non-linear bilateral subtraction technique is applied.[5] By examining pairs of images that are first thresholded and then subtracted, asymmetries between left and right views can be enhanced. Adaptive region growing, where the size of the grown region depends on the grey-level interval used for the region growing, is used to extract the lesion from its background. Next, various features are extracted both from the original and the processed image. The features are of 3 types: geometric measures (e.g., shape), gradient-based measures (e.g. average gradient using a 3 x3 Sobel operator), and intensity-based measures (e.g., contrast). These features are used as input to a back-propagation, feed-forward artificial neural network, which can distinguish true masses from false positives. Using this method, 92% of masses at a false positive rate of approximately 2 per image was obtained on a database of 154 pairs of mammograms -- 90 with masses and 64 normal. The masses had a mean diameter of less than 2 cm and a mean contrast of 0.49 in terms of film optical density.

The automated scheme for the detection of clustered microcalcifications consists of four basic steps, as shown in the flowchart in Fig. 1. The images have 0.1-mm pixels and 10-bit grey-scale resolution. The four different steps are: (1) Difference Image Technique: Two linear filters are used to increase the signal-to-noise ratio of microcalcifications. By subtracting a signal-suppressed image from a signal-enhanced image, the normal background structure of the breast is

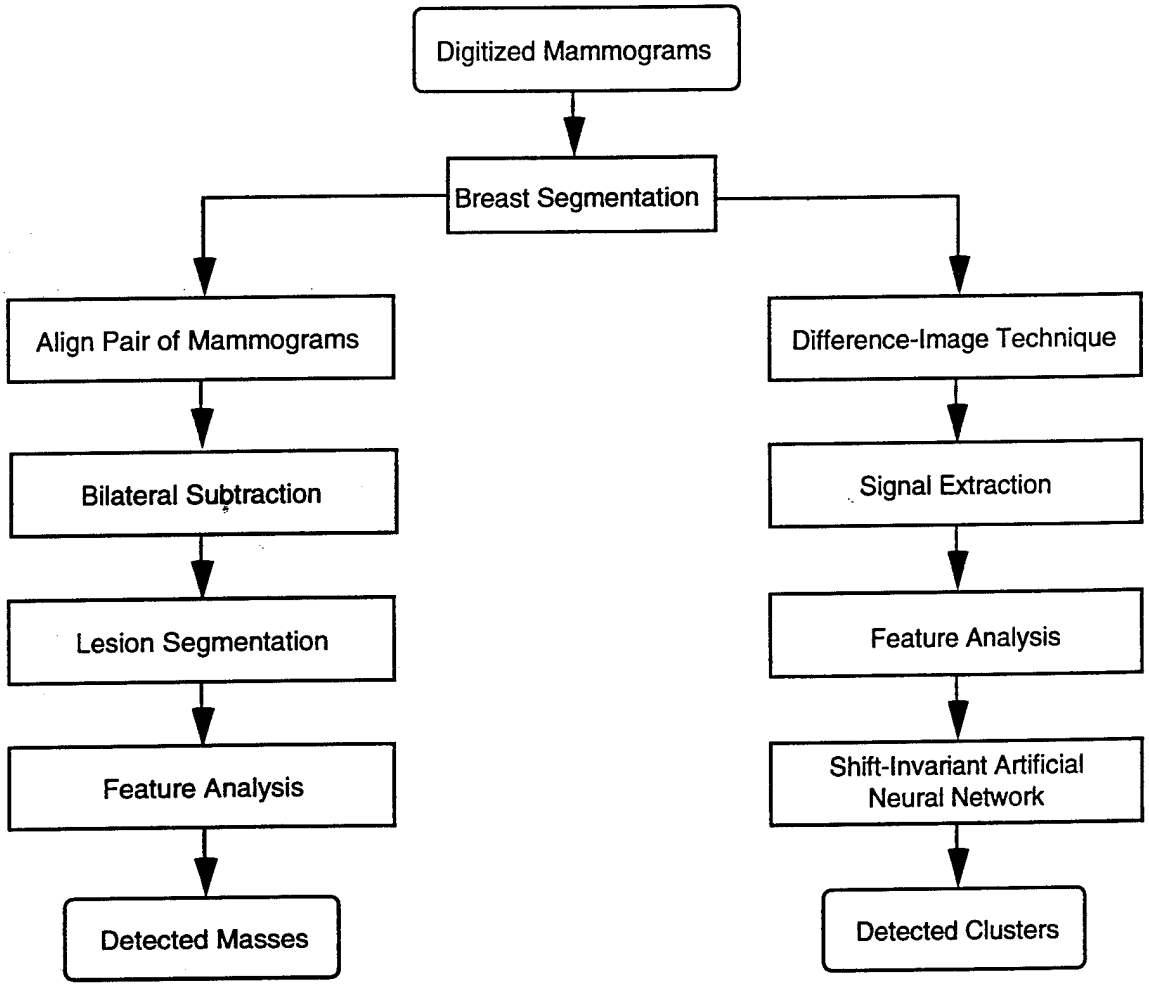


Figure 1. Flowcharts outlining our computerized schemes for the automated detection of masses and clustered microcalcifications.

suppressed and as a result microcalcifications appear enhanced. (2) Signal extraction technique: a global and local adaptive grey-level thresholdings are applied to the processed difference image.[6] In addition, a morphological erosion operator is used to eliminate very small signals (1 or 2 pixels in size). (3) Feature analysis: four different features (texture, area, contrast, and spatial distribution) are extracted from the image and are used in a series of rule-based thresholding to reduce the number of falsely identified signals.[7-9] (4) Artificial neural network: After feature analysis, the detected clusters are used as input into a shift-invariant artificial neural network (SIANN).[10] The SIANN is a multilayer back-propagation neural network with local, shift-invariant interconnections. The SIANN was trained to detect individual microcalcifications. A cluster was considered positive if there were at least 2 microcalcifications detected in the ROI. Negative ROIs were removed from the image. The resulting output image contained the detected clustered signals. This scheme has a sensitivity of 85% with a false-positive rate of 0.6 clusters per image, when tested on 78 clinical mammograms -- half with clusters, half without. It should be noted that the mean size and mean contrast of the microcalcifications in our database are smaller than those reported by most other investigators.

3. METHOD

Each day all screening mammograms are digitized, 4-views per case. As the films are being digitized, the microcalcification detection program is run on-line in parallel. The mass detection is run off-line overnight, since the films are not reviewed until the next afternoon. After all four films have been analyzed, the results of the microcalcification detection program are displayed in a single 1024x1024 image as a collage of four, 512x512 images with arrow(s) annotated to the image indicating the computer results. This image serves two purposes. First, it allows for the verification that the films were properly digitized and second a thermal-paper image is made. This hardcopy is used by the reviewing radiologist when reading the original films, in order to simulate CAD reading conditions. The results of the mass detection program are printed using the same format the next morning. In the future, two 1024-line CRT monitors will be used to display the computer results to the radiologist. A full case, four films, can be processed in less than 5 minutes.

For this initial study, the results of the computer analyses are not used in deciding proper patient care. Instead, all screening mammograms are read twice, once by our study radiologist and once by a radiologist who dictates the official clinical report (clinical radiologist). The study and the clinical radiologists are experienced in reading mammograms. The study radiologist reads the mammograms without the computer results and makes an interpretation. Then the computer results are reviewed and a revised interpretation, if necessary, is given. The performances of the CAD scheme alone, the study radiologist without CAD, the study radiologist with CAD, and the clinical radiologist (who does not use the computer results) are recorded.

Since this is a prospective study, evaluation of "truth" is very difficult at present. In this paper we have used as truth the opinion of the study radiologist. Any lesion that she considered suspicious for malignancy was considered a true lesion. For masses, if the lesion had a greater than 50% probability of being malignant, in the opinion of the study radiologist, it was considered to be a true mass. A lesion with a 20-50% probability for malignancy was considered to be a "true" benign lesion. A lesion with less than a 10% probability of malignancy was not considered to be a true lesion, neither malignant nor benign. For clustered microcalcifications, any cluster with greater than 30% probability of being malignant was scored as a true cluster by the study radiologist. Benign appearing calcifications were not considered a true lesion. Computer detection of such calcifications was scored as a false-positive detection in this study. Correlation of the study radiologist's opinion with the opinion of the clinical radiologist and with pathology are being made.

4. RESULTS

Tables 1 and 2 summarize the computer schemes' performances. To date, 573 cases or 2292 films have been analyzed by both schemes. However, in 80 pairs of mammograms, the left and right views were not aligned properly and the program automatically passed over the mass detection scheme. In total, the two computerized schemes correctly identified 10 of 14 suspicious lesions as judged by the study radiologist. There were 4 lesions that were considered suspicious by the study radiologist that the computer schemes missed. Of the 3 missed masses: two were benign, while the other patient is scheduled for further work-up. One suspicious cluster of microcalcifications was missed by the detection program, but upon further

5. DISCUSSION

In this study, we have used the study radiologist's opinion as truth. Because this is a subjective assessment, it is useful to compare the clinical and the study radiologists' opinions. To do this, though, one must first consider what is "truth" (since biopsy results are not yet available). If we were to use as truth all patients who had further work-up, then the number of "true" lesions would be 104. However, in clinical practice, work-up is often used to show that an indeterminate lesion is just normal tissue, for example, overlapping of normal tissue mimicking a spiculated mass. Therefore, using lesions that require additional work-up as truth would result in a large fraction of truly normal patients being included as abnormal (over 50% in this study). Similarly, biopsy is often used to confirm that a probably benign lesion is indeed benign. So using biopsy as truth would include a substantial number of benign lesions, as much as 80% based on the published literature.

While it is important that CAD schemes do not miss any cancers, it is not clear whether CAD schemes should identify any types of benign lesions. It may seem desirable to have the computerized schemes detect only malignant lesions, but this is probably not the case. For the radiologist to have confidence in the computer results, all lesions that the radiologist believes are malignant, should be identified by the computer. The other extreme is to have the computer identify all lesions,

Table 3. Causes of false positives in mass detection program.

Cause	Number
Nodular	1986 (71%)
Dense	30 (0%)
Artifacts	1 (0%)
Nipple	18 (0%)
Unknown	765 (27%)
Total	2800 (100%)

Table 4. Causes of false positives in clustered microcalcification detection program.

Cause	Number (percentage)
Vascular Calcifications	736 (37%)
Obvious Benign Calcifications	197 (10%)
Artifacts	16 (0%)
Unknown	1056 (53%)
Total	2003 (100%)

benign or malignant, that are present. In this situation, most computer findings would be benign lesions that would not require any treatment. The radiologist may get annoyed with the computer indicating trivial findings -- most breasts will show either small benign nodular densities, such as lymph nodes, or scattered benign calcifications. Therefore, it is not clear where in between these two extremes CAD should perform. Furthermore, in its current status, our computerized schemes only detect lesions present on a mammogram -- they are not designed to distinguish benign from malignant lesions. So it is not reasonable for the CAD schemes to identify only malignant lesions, at the present time. Computerized schemes are being developed in our laboratory to classify benign and malignant lesions,[1, 2] but these have not yet been implemented on the workstation.

Therefore, we have chosen to use 'suspicious' lesions as judged by the study radiologist as truth. While this is an arbitrary and subjective method, it does insure that most "true" lesions are probably malignant. Out of the 573 patients, there were only 4 cases in which the study radiologist considered a lesion not to be suspicious, when the clinical radiologist recommended a biopsy. One of the lesions had only a 30% probability of being malignant as judged by the clinical radiologist (i.e., the clinical radiologist ordered a biopsy to verify the mass is benign). The three others were mass cases in which the study radiologist thought the lesions were a cyst, radial scar, and normal tissue. Biopsies have not yet been performed, so it is not known at present, which if any of these lesions are truly cancer. Conversely, from Table 2, in 6 of the 14 lesions considered suspicious by the study radiologist, the clinical radiologist recommended a biopsy. Four of the 14 are still in work-up, and in 2 of the 14 six-month follow-up was recommended by the clinical radiologist. Overall the agreement between study and clinical radiologists has been quite good. The study radiologist, whose opinion we are using as truth, is not undercalling or missing cancers.

Finally, the false-positive rates in this study have been relatively low. Therefore, we can choose to adjust the two detection programs so that their sensitivities are increased while increasing the false-positive rates slightly, but maintaining a low false-positive rate overall (less than 3 false detections per image for both schemes combined).

6. SUMMARY

We have begun prospective testing of our CAD schemes for the detection of masses and clustered microcalcifications in digital mammograms. Preliminary results on the first 573 patients are very encouraging. Both the false-negative and the false-positive rates are low; a substantial fraction of missed lesions have been benign. Furthermore, there have been two cases where the computer identified a cluster of microcalcifications that was initially missed by the study radiologist. We plan to use the results of this on-going study to plan a full clinical evaluation of our "intelligent" mammography workstation.

7. ACKNOWLEDGMENTS

The authors thank K.J. Munn, Y. Jiang, Z. Huo, Y. Mao, R.H. Nagel, W.K. Zouras, W. Cai, W. Zhong, M. Carlin, M. Zhang, J. Shiraishi, and H. Sumida for their assistance on this project.

This work was done as part of the National Digital Mammography Development Group (NDMDG), which was established and funded in part by the NCI/NIH. Members of the NDMDG are: the Universities of Chicago, Toronto and North Carolina (Chapel Hill), General Electric Corporation, Thomas Jefferson University and Massachusetts General Hospital.

The participation of the University of Chicago in the NDMDG is funded by the NCI (CA 60187 and CA 24806). Additional funding for this work was provided by the Whitaker Foundation, the U.S. Army (DAMD 92153010), and the American Cancer Society (FRA 390), two additional grants from the NIH/NCI (CA 48985 and T32 CA09649), Toshiba Corporation, Eastman Kodak Company, and R2 Technology, Inc. The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of any of the supporting organizations.

work-up, this cluster was considered to be benign by the clinical radiologist. In total, 10 of the 14 patients with a suspicious lesion were correctly identified by the computerized scheme. Three of the four missed lesions were benign.

The computer was able to detected two clusters of microcalcifications initially missed by the study radiologist. After being alerted to the area containing the cluster, the study radiologist thought the two clusters were suspicious. These two clusters were also missed by the clinical radiologist. These clusters have not yet been worked-up, so it is not known at this time whether either are real.

The false-positive rates were 1.31 false masses and 0.87 false clusters per image. We consider these false-positive rates to be very acceptable. There have been 26 (5%) images in which a normal mammogram had no false masses detected and 116 (20%) normal mammograms had no false clusters detected. The causes of the false positives, as determined by the study radiologist, are listed in Tables 2 and 3. Over 70% of false masses are caused by nodular density patterns. Approximately half of the false clusters were caused by benign calcifications.

Table 1. Summary of results of the computerized detection schemes for the first 573 patients.

Detection Program	Number of Films	Sensitivity*^	False Positives per Image	Computer "True" Detections Missed by Radiologist*
Masses	2132	6/9	1.31	0
Clustered Microcalcifications	2292	4/5	0.87	2
Totals	2292	10/14	2.18	2

* On a per patient bases as opposed to per image
^ A patient with a suspicious lesion was scored as a true positive if the computer detected the lesion in either of the two views.

Table 2. Clinical outcomes for patients whom the study radiologist considered suspicious for malignancy. Each entry gives the number of patients for which the computer detected a lesion (numerator) and the total number of patients in that category (denominator).

Type of Lesion	Cases Considered Suspicious for Malignancy by Study Radiologist			
	Patients with Positive Work-up (biopsy recommended)	Patients with Negative Work-up (biopsy not recommended)	Patients not yet having Completed Work-up	Totals
Masses	4/5	1/2	1/2	6/9
Clustered Microcalcifications	1/1	1/2	2/2*	4/5
Totals	5/6	2/4	3/4	10/14

* These two cases were detected by the computer scheme, but were initially missed by both the study and the clinical radiologists.

8. REFERENCES

1. M. L. Giger, P. Lu, Z. Huo, U. Bick, K. Doi, C. J. Vyborny, R. A. Schmidt, W. Zhang, C. E. Metz, D. E. Wolverton, R. M. Nishikawa and W. C. Zouras, "CAD in digital mammography: Computerized detection and classification of masses," in *Digital Mammography,* edited by A. G. Gale, S. M. Astley, D. R. Dance and A. Y. Cairns (Elsevier Science B.V., Amsterdam, 1994), pp. 281-288.

2. R. M. Nishikawa, Y. Jiang, M. L. Giger, K. Doi, R. A. Schmidt, C. J. Vyborny, W. Zhang, T. Ema, J. Papaioannou, D. E. Wolverton, U. Bick, R. H. Nagel and Y. Mao, "Performance of Automated CAD schemes for the detection and classification of clustered microcalcifications," in *Digital Mammography,* edited by A. G. Gale, S. M. Astley, D. R. Dance and A. Y. Cairns (Elsevier Science B.V., Amsterdam, 1994), pp. 13-20.

3. U. Bick, M. L. Giger, R. A. Schmidt, R. M. Nishikawa, D. E. Wolverton, P. Lu, C. J. Vyborny and K. Doi, "Automated segmentation of digitized mammograms," Academic Radiology 2, 1-9 (1994).

4. F.-F. Yin, M. L. Giger, K. Doi, C. J. Vyborny and R. A. Schmidt. Computerized detection of masses in digital mammograms: Automated alignment of breast images and its effect on bilateral-subtraction technique. 1992: .

5. F.-F. Yin, M. L. Giger, K. Doi, C. E. Metz, C. J. Vyborny and R. A. Schmidt, "Computerized detection of masses in digital mammograms: Analysis of bilateral-subtraction images," Med. Phys. 18, 955-963 (1991).

6. H.-P. Chan, K. Doi, C. J. Vyborny, C. E. Metz, H. MacMahon, P. Jokich and S. Galhotra, "Digital mammography: Development of a computer-aided system for detection of microcalcifications," Proc. SPIE 767, 367-370 (1987).

7. H.-P. Chan, K. Doi, C. J. Vyborny, K. L. Lam and R. A. Schmidt, "Computer-aided detection of microcalcifications in mammograms: Methodology and preliminary clinical study," Invest. Radiol. 23, 664-671 (1988).

8. Y. Jiang, R. M. Nishikawa, M. L. Giger, K. Doi, C. J. Vyborny and R. A. Schmidt, "Method of extracting microcalcifications' signal area and signal thickness from digital mammograms," Proc. SPIE 1778, 28-36 (1992).

9. R. M. Nishikawa, M. L. Giger, K. Doi, C. J. Vyborny, R. A. Schmidt, C. E. Metz, Y. Wu, F.-F. Yin, Y. Jiang, Z. Huo, P. Lu, W. Zhang, T. Ema, U. Bick, J. Papaioannou and R. H. Nagel, "Computer-aided detection and diagnosis of masses and clustered microcalcifications from digital mammograms," Proc. SPIE 1905, 422-432 (1993).

10. W. Zhang, K. Doi, M. L. Giger, R. M. Nishikawa and Y. Wu, "A shift-invariant artificial neural network for detecting clustered microcalcifications in digital mammograms," Med. Phys. 21, 517-24 (1994).

clustered microcalcifications in digital mammograms," Med. Phys. 21, 517-24 (1994).

Analysis of Spiculation in the Computerized Classification of Mammographic Masses*

Zhimin Huo, M.Sc
Maryellen L. Giger, Ph.D
Carl J. Vyborny, M.D, Ph.D
Ulrich Bick, M.D
Ping Lu, M.Sc
Dulcy E. Wolverton, M.D
Robert A. Schmidt, M.D

Kurt Rossmann Laboratories for Radiologic Image Research
Department of Radiology
5841 South Maryland Avenue
The University of Chicago
Chicago, Illinois 60637

Corresponding Author:
Maryellen L. Giger, Ph.D
Dept. of Radiology, MC2026, The University of Chicago
5841 S. Maryland Ave, Chicago, IL 60637
312-702-6778, 312-702-0371 (Fax)

*Presented in part at the 1993 annual meeting of the RSNA, Chicago, Illinois and the 1994 annual meeting of AAPM, Anaheim, CA. Supported by the American Cancer Society (FRA-390), USPHS grant CA 48985 and US Army grant DAMD17-93-J-3021

ABSTRACT:

Spiculation is a primary sign of malignancy for masses detected by mammography. In this study, we developed a technique that analyses patterns and quantifies the degree of spiculation present. Our current approach involves (1) automatic lesion extraction using region growing and (2) feature extraction using radial edge-gradient analysis. Two spiculation measures are obtained from an analysis of radial edge-gradients. These measures are evaluated in four different neighborhoods about the extracted mammographic mass. The performance of each of the two measures of spiculation was tested on a database of 95 mammographic masses using ROC analysis that evaluates their individual ability to determine the likelihood of malignancy of a mass. The dependence of the performance of these measures on the choice of neighborhood was analyzed. We have found that it is only necessary to accurately extract the principal outlines of a mass lesion for the purposes of this analysis since the choice of a neighborhood that accommodates the thin spicules at the margin allows for the assessment of margin spiculation with the radial edge-gradient analysis technique. The two measures performed at their highest level when the surrounding periphery of the extracted region is used for feature extraction, yielding Az values of 0.83 and 0.85, respectively, for the determination of malignancy. These are similar to that achieved when a radiologist's ratings of spiculation (Az=0.85) are used alone. However, a combination of the measurements from the four neighborhoods is superior in the classification of mammographic mass lesions.

Key words: spiculation
digital mammogram
radial edge-gradient analysis
ROC analysis
computer-aided diagnosis
computer vision

I. INTRODUCTION

X-ray mammography has been proven to be the most effective method for the detection of early breast cancer. However, mammographic findings of benign and malignant masses often overlap[1]. At many centers, only 10-20% of detected masses removed by surgical breast biopsy are malignant[2,3]. A computer scheme capable of providing objective information may aid radiologists in their classification of masses, thus preventing unnecessary biopsies. Computer aids have already been shown to improve the detection performance of radiologists[4,5].

The shape, margin and density of a mass are used by radiologists to characterize masses[1,6,7]. The margin characteristics of a mass observed mammographically are very important indicators of its benign or malignant status. The margin of a mass can be categorized as circumscribed, lobulated, obscured, indistinct or spiculated with a spiculated margin being the strongest sign for malignancy[1,6,7].

Various investigations[8-13] have attempted to classify breast lesions or to detect spiculated masses based on computer-extracted features characterizing either the margin, shape or density of a mass. Ackerman *et al* [8] extracted four features of malignancy, calcification, spiculation, roughness and shape from lesions identified by radiologists on xeroradiographs and then merged the four features to classify those lesions. Brzakovic *et al* [9] classified detected abnormalities into non-tumor, benign tumor and malignant tumor using measures of size, shape and intensity change. Kegelmeyer[10] used the analysis of edge orientation histograms to detect stellate lesions. Kilday *et al*[11] segmented lesions with a simple thresholding technique and used linear discriminant analysis to merge several shape-related features to distinguish between fibroadenomas, cysts, and carcinomas. Other investigators have used only a single computer-extracted feature related to either margin, shape or density as an indicator of malignancy. Burdett *et al*[12] applied a fractal analysis to quantify the degree of surface roughness as a single indicator of malignancy.

Claridge *et al*[13] analyzed a small set of malignant lesions by measuring the lesion edge blurriness. In addition, many investigators[14-18] have taken advantage of the ability of radiologists to extract mammographic features, which are subsequently merged by rule-based, discriminant analysis or neural networks into a final determination of the likelihood of malignancy.

Previously we developed a classification method that involved the extraction of lesions using a manual region-growing technique and the extraction of two features containing margin information. These were merged by an artificial neural network to quantify the degree of spiculation[19]. The database in that study contained 28 benign and 25 malignant masses. The result showed that the mammographic features extracted and merged in this way yielded measures of spiculation comparable to those obtained by an expert mammographer.

In this study, we developed a new spiculation-sensitive pattern-recognition technique, "radial edge-gradient analysis." Prior to the feature extraction, we employed an automatic lesion segmentation to accurately extract the lesion. The radial edge-gradient analysis was then performed on various neighborhoods of the extracted lesion. Two new measures of spiculation are generated with this technique. Their performances in classifying malignant and benign lesions were evaluated using ROC analysis. The dependence of the two spiculation measures on the different neighborhoods about the extracted lesion was also investigated. It should be noted that the input to the classification scheme, that is the initial detection of a mass, could come from either a radiologist or a computer detection method[20,21,22].

II. MATERIALS

The database used in this study consisted of 95 clinical mammograms (Kodak MinR screen/OM-1 film, Eastman Kodak, Rochester, NY), each containing a mass. Of the ninety

to characterize our database, an experienced mammographer (CJV) rated each mass with respect to spiculation and shape on a scale from 1 to 5, with 1 corresponding to definitely smooth or definitely circular and 5 corresponding to definitely spiculated or definitely ovoid. The distribution of the masses in terms of spiculation and shape are shown in Figures 1(a) and 1(b), respectively.

The screen/film mammograms were digitized with an optical drum scanner (FIP II, Fuji Film, Tokyo, Japan) at a sampling distance of 0.1 mm and at 10-bit quantization. The classification analysis was performed within a 512 by 512 pixels region of interest centered about the mass in question as illustrated in Figure 2, which shows examples of (a) malignant and (b) benign mammographic masses.

III. METHODS

Our current approach consists of two major steps: (1) automated segmentation of mammographic mass lesions from the surrounding parenchyma and (2) automated feature extraction using radial edge-gradient analysis.

A. Automated Extraction of Lesion from Surrounding Breast Parenchyma

Lesion extraction starts with a region of interest (ROI) of size 512x512 centered about the abnormality in question (Figures 2(a) and 2(b)) that excludes areas outside the breast[20]. The region is then processed using background trend correction and histogram equalization prior to the automatic extraction of the lesion using a modified region growing technique[21].

A two-dimensional background trend correction is employed to correct the nonuniformity of the background optical density (pixel value) distribution upon which the mass is superimposed since overlying normal structures may cause the mass to be either undergrown or overgrown. The background trend is estimated by fitting a two-dimensional

(second-order polynomial) surface to the gradual change in the background pixel values within the 512x512 ROI.

Histogram equalization[22] is then performed within the same region in order to enhance contrast, which in turn enhances the mass margin and allows the region growing technique to more accurately extract the margin of the mass. Figure 2 shows (c) the malignant mass and (d) the benign mass after background trend correction and histogram equalization processing.

Gray-level region growing is subsequently performed in the processed 512x512 ROI. For each mass, the starting pixel for region growing is selected as the pixel with the highest gray level within a 15-pixel distance (x,y) from the location of the identified center. Region growing is then performed as described below for various gray-level intervals in increments of 3 pixel values.

The size of a grown region is defined as the effective diameter of the equivalent circle (whose area is the same as the area of the grown region) of the grown region and the circularity is defined as the ratio of the area of the grown region within the circle to the area of the grown region. The "transition point" indicating the terminating gray level for the region growing is determined by employing a multiple "transition point" technique. In this technique, two potential "transition point" candidates are searched and examined. The first potential "transition point" candidate is determined based on criteria involving the derivative of size as a function of gray-level interval, since abrupt increases in size correspond to merging of the grown mass and the background. The second potential "transition point" candidate is determined in much the same way as the first. However, if the second potential "transition point" corresponds to the grown region having the highest value in circularity or a larger drop in circularity, the second rather than the first is retained as the "transition point", because decreases in circularity also correspond to merging of the mass and background. Determined transition points are indicated in Figures 2(e) and 2(f), which are the diagrams of

size and circularity of the grown region as functions of gray-level, for the malignant mass and benign mass in Figures 2(a) and 2(b), respectively.

This multiple "transition point" technique has overcome the difficulties encountered in region growing for masses that overlie certain high contrast normal structures since the criteria used for the second "transition point" candidate, which restrain the circularity to be either the maximum value of the regions grown from all the gray-level intervals or a larger drop in circularity than that of the first potential "transition point" candidate, avoid having the gray-level region growing terminated prematurely inside the mass. Our results show that the multiple "transition point" technique for region growing is successful in correctly identifying the margins of mammographic masses. Extracted margins overlayed on the original images are shown in Figure 2 for (g) the malignant and (h) the benign masses, respectively. The two image processing techniques help assure that the gray-level region growing technique correctly identifies the margins of masses, thus, allowing for subsequent feature extraction and analysis.

B. Radial Edge-Gradient Analysis: Spiculation Measures

Once lesions are accurately extracted, a radial edge-gradient analysis technique is applied within various neighborhoods about the grown region to quantitate the marginal spiculation of a mass. The neighborhoods, shown schematically in Figure 3 (a-d), are A) within the grown region, B) along the extracted margin, C) within a rectangular segment containing the mass in question, and D) in the surrounding periphery, respectively. The hatched area in each neighborhood is the actual region used in the analysis. The size of the rectangular segment in neighborhood C and D is determined by allowing an additional 10 pixels on each side of the grown region. The smooth curve within the grown region in neighborhood D is obtained by applying a morphological open filter [23] on the grown region with a circular kernel (the radius of the kernel being 20% of the effective diameter of the

equivalent circle of the grown mass). Due to the difficulty of capturing into the grown region all of the thin spicules radiating from the margin of a mass, neighborhoods C and D were introduced to include spicules that remain outside the grown region. The effect of the four different neighborhoods on the radial edge-gradient analysis used to quantify the spiculation of a mass will be discussed later.

In the radial edge-gradient analysis, the maximum gradient at each pixel location of a particular neighborhood is calculated with a 5x5 Sobel filter[22] and the angle of this gradient relative to its radial direction is determined. Figure 4 illustrates the definition of this angle relative to the radial direction, which is referred to as "radial angle" in this paper. The radial direction for point p1 is the direction pointing from the geometric center of the grown mass to p1. The angle θ between the direction of the maximum gradient at the pixel p1 and its radial direction is the angle relative to the radial direction or the "radial angle". Note that θ is not the angle the maximum gradient makes with the x-direction. Analysis relative to the x-axis yields information only on whether a lesion is circular or not[5,21], i.e. it can only be used to distinguish circular patterns from linear patterns, not circular patterns from spiculated patterns. Rather, our analysis was developed in order to distinguish spiculated masses from circular or oval masses with smooth margins, since spiculation is an important indicator of malignancy.

In each neighborhood, the maximum gradients having the same radial angle are summed for each radial angle, resulting in a cumulated edge-gradient distribution relative to the radial angle. The cumulated edge-gradient distribution is then normalized by the average maximum gradient of the particular neighborhood, enabling comparison of cumulated edge-gradient distributions between various lesions. Normalization is performed such that the area under the normalized distribution curve is one. Figures 5c and 5d show the normalized cumulated edge-gradient distributions relative to the radial angle obtained using neighborhood B (margin) for the (a) smooth, round-shape benign mass and (b) spiculated, round-shape

malignant mass, respectively, of Figure 5. It should be noted that the benign mass yields a narrow peak at 180 degrees, since the angle of the maximum gradient relative to the radial gradient of a smooth, round-shape mass is approximately 180 degrees at pixels along the margin of the mass. The spiculated mass yields a broad peak distribution, since the angle of the maximum gradient relative to the radial direction of the mass varies greatly along the margin of the mass. The full width at half maximum (FWHM), in terms of degrees, of the normalized cumulated edge-gradient distribution relative to the radial angle is used to characterize the shape of the distribution of each mass in each neighborhood type. As discussed above, a spiculated mass has a broader normalized cumulated edge-gradient distribution than a smooth mass. Thus, by determining the FWHM of a distribution, a measure of spiculation can be determined.

It should be noted that the geometric shape of a mass has an impact on the normalized cumulated edge-gradient distribution of the mass. This effect will be discussed in the next section.

Another feature arising from radial edge-gradient analysis that can be used to quantify the degree of spiculation is the "normalized radial gradient". A radial gradient at a pixel (p1) is defined as the projection of the maximum gradient at the pixel (p1) along its radial direction (Figure 4). A normalized radial gradient for an entire neighborhood is the summation of the radial gradients from all the pixels in the neighborhood divided by the summation of the magnitude of the maximum gradients from all the pixels in the same neighborhood or

$$\frac{\sum_{N}|G_i cos\theta_i|}{\sum_{N}G_i} \quad ,$$

where θ_i is the radial angle at pixel i, G_i is the magnitude of the maximum gradient at pixel i and N is the total number of pixels in the neighborhood. The value of the normalized radial gradient is between zero and one, with one corresponding to a round mass. Generally,

smooth masses have larger values of radial gradient than spiculated masses, since the maximum gradients along the margin of a smooth mass have larger projections along the radial direction than they do for a spiculated mass.

C. Geometric Shape Correction for the Spiculation Measures

The aim of radial edge-gradient analysis is to distinguish spiculated lesions from nonspiculated lesions. However, the geometric shape of a mass will affect the spiculation measures obtained from the radial edge-gradient analysis. For this reason, a simulation study was performed to quantify the effect of lesion shape on the spiculation measures, and a simple correction on the spiculation measures was made based on the simulation results, as discussed below.

A smooth round mass and a number of smooth oval-shaped masses having different long-to-short axis ratios were simulated as shown in Figure 6. In addition, two spiculated lesions were simulated--one slightly spiculated and one highly spiculated. Radial edge-gradient analysis was then performed on the simulated masses including the calculation of the two spiculation measures as shown in Table 1. The effect of shape on the two spiculation measures is shown in Figure 7 for the smooth simulated masses. As the long-to-short axis ratio of the smooth simulated masses increases, the FWHM measure increases, while the radial gradient measure decreases. Due to the discrete pixel size of the digital image, the values of the FWHM measure and the radial gradient measure for the simulated round-shaped mass are 27 and 0.98, instead of zero and one, respectively. The effect of bin size was eliminated in the simulation study, as shown in Figure 7, by reducing the bin size of the cumulated edge-gradient distribution from fifteen degrees to one degree.

The two spiculation measures for the simulated spiculated masses are also given in Table 1 and shown in Figure 7. It should be noted that the value of the FWHM measure for the slightly spiculated mass is larger than most of the smooth masses, but less than the ones

with long-to-short axis ratios larger than 1.7. This will result in misclassification between spiculated masses and smooth oval-shaped masses. In clinical practice, smooth round or oval masses are usually classified as benign or probably benign[7]. Of these two, smooth oval masses have the even greater likelihood of being benign. In order to improve the specificity of our approach without loss in sensitivity, the FWHM measures must be corrected for geometric shape. Precautions against over-correction of FWHM need to be taken when a correction is made, however, since the "cost" of a missed cancer is much greater than misclassification of a benign case. The effect of geometric shape on the radial gradient measure can be considered negligible as seen from Figure 7b, which shows this measure for the simulated smooth masses (circular and oval-shaped) and the simulated spiculated masses.

As discussed above and as seen in Figure 7a, the effect of geometric shape on the FWHM measure is not negligible and a correction for the FWHM measure is necessary. We have found that a single value correction on the FWHM measure can be used in practice. In order to correct the FWHM measure, the grown region of a mass is first smoothed by running mean filtering of the pixel locations of the extracted margin. The size of the running mean is equal to 5% of the total margin length in terms of number of pixels. Next, an ellipse is generated based on the smoothed margin and the long-to-short axis ratio of the ellipse is calculated. The correction of the FWHM measure is made according to the calculated long-to-short axis ratio. To prevent misclassifying spiculated masses as nonspiculated masses by over-correcting the FWHM measures, the FWHM correction is made only for the masses having a long-to-short axis ratio larger than 1.8, since the simulation study shows the overlap of the FWHM measures exists only between spiculated masses and smooth masses having a long-to-short axis ratio greater than 1.7. For masses having a long-to-short axis ratio greater than 1.8, the correction value is chosen such that it is smaller than that needed to correct for the simulated masses; but large enough so that the corrected FWHM measures are less than the FWHM measures of the simulated spiculated masses. Thus, the correction was chosen to

be a reduction by 36 degrees of the FWHM measure for masses having a long-to-short axis ratio greater than 1.8 .

IV. RESULTS

Figure 8 shows the relationship between the corrected FWHM and the normalized radial gradient measures within the rectangular segment (neighborhood C) for the 95 masses. It is apparent that most of the malignant masses have large values of FWHM and small values of normalized radial gradient. For example, by setting a threshold at 160 degrees for the FWHM measure, 75% of the malignant masses can be correctly identified with only 4 out of 38 benign masses being misclassified (Figure 8).

ROC analysis[24,25] was undertaken to evaluate the abilities of each of the two spiculation measures determined for the four neighborhoods in distinguishing between benign and malignant masses. The area under the ROC curve (Az) was calculated as an index for the performance of each feature as shown in Table 2. Figures 9(a-c) show the individual performance of the two spiculation measures for each neighborhood type. The performances of the uncorrected FWHM and normalized radial gradient measures in classifying the 95 masses for each neighborhood are similar as shown in Figures 9a and 9b. It is apparent that the choice of neighborhoods will affect the performance level also as illustrated in Figure 9a and 9b. The effect of the four neighborhoods on the two spiculation measures shows the same trend for each measure, with the grown region (neighborhood A) yielding the lowest Az value, the margin (neighborhood B) the second lowest, the encompassing region (neighborhood C) a higher value, and the surrounding periphery (neighborhood D) the highest Az value. When ROC analysis is performed on the FWHM measure after geometric shape correction, there is a consistent improvement in the performance of the FWHM measures for all four neighborhoods as shown in Figure 9c.

This improvement is limited due to the relatively small number of oval-shaped masses in the clinical database, however (Figure 1b).

As an approach to maximize the sensitivity of the FWHM measures, only the greatest value of the corrected FWHM measures from the four neighborhoods for each mass was chosen and ROC analysis performed. The Az obtained using the maximum value of the corrected FWHM from the four neighborhoods is 0.88 (Figure 10).

As mentioned earlier, spiculation is one of the most important determinants used by radiologists in determining the benign or malignant status of a mass. In order to compare the performance of our computer-based spiculation measures to those extracted by a human observer, ROC analysis using the radiologist's spiculation rating alone (Figure 1a) for determining the likelihood of malignancy was performed. This yields an Az of 0.85 as shown in Figure 10. The performance of the corrected FWHM measure obtained from neighborhood C (Az=0.83) or neighborhood D (Az=0.85) is comparable with the human visual assessment of the marginal spiculation of a mass. However, with the combination of the corrected FWHM measures from all four neighborhoods, the computer-based spiculation measure (FWHM) achieves higher Az values (Az=0.88) than that based on the spiculation ratings from an human observer. Of course, with the use of additional mass-related features such as opacity or shape, the performances of both the computer-based measures and human assessment would be expected to improve.

V. DISCUSSION

In order to maximize the extraction of the margin spiculation information from a mass, four different neighborhoods about the grown region were introduced for feature extraction. As described earlier, neighborhoods A and B rely entirely on the grown region, whereas neighborhoods C and D introduce regions surrounding the grown mass in order to include thin, short spicules radiating from the margin of a mass, which could not be delineated by the

gray-level region growing technique. The size of the region introduced in neighborhoods C and D is only large enough to accommodate thin, short spicules. Since the four neighborhoods are determined from the grown region, the accuracy of the lesion segmentation prior to feature extraction is important in the success of subsequent feature analysis. However, the regions introduced in neighborhoods C and D make the subsequent analysis less dependent on the grown region.

Results show that spiculation analysis within neighborhoods C and D yield higher Az values than that within neighborhoods A and B. This demonstrates the usefulness of introducing a zone around the extracted lesion to accommodate potential margin spiculation. The Az values of the two spiculation measures obtained from margin (B) and surrounding periphery (D), which exclude most of the interiors of the grown (A) and encompassing (C) regions, are higher than the Az values obtained from the grown (A) and encompassing (C) regions themselves, respectively. This demonstrates that mainly using the margin information increases the "signal-to-noise" ratio, and thus, optimizes the radial edge-gradient analysis technique in the extraction of the margin spiculation.

Thus, with the radial edge-gradient analysis technique, we found that a lesion can be extracted devoid of its spicules and still be accurately analyzed for spiculation if the proper neighborhood is chosen. That is, by studying the periphery (neighborhoods C and D) around a grown mass, it is not necessary to require that the grown region include fine spicules.

In the application of radial edge-gradient analysis in classifying mammographic masses, both FWHM and radial gradient measures yield useful spiculation information and can be used to differentiate smooth masses from spiculated masses. However, in general, both features could be used to differentiate smooth, round-shape patterns from other patterns, i.e. either spiculated or linear patterns. Our simulation study shows that the FWHM measure tends to be more sensitive to linear patterns than is the radial gradient.

Since margin characteristics are the major discriminants of the benign or malignant status of a mass, future investigation of computer-based features that could characterize the margin of a mass into more specific categories (circumscribed, lobulated, obscured, indistinct or spiculated) will also be useful. In addition, we are developing techniques to extract features related to the shape and density of mass lesions. The margin features along with the features characterizing the shape and density of a mass will eventually be merged by artificial neural networks into likelihoods of malignancy[26].

VI. CONCLUSION

In this study, a radial edge-gradient analysis technique for pattern recognition of mass lesion seen on mammography is presented. Its application is demonstrated in its ability to differentiate between smooth masses and spiculated masses, which could aid radiologists in distinguishing between benign and malignant abnormalities. The characterization of spiculation resulting from the computer analysis was, in fact, comparable to that of an experienced mammographer. Further, the use of appropriately selected neighborhoods for the analysis is shown to decrease the need for highly detailed segmentation of spicules at margin of mass lesions. The results support the reliability of automated feature extraction in the analysis of one of the most important predictive imaging features for breast malignancies.

ACKNOWLEDGMENTS

The authors thank Kunio Doi, Ph.D., Charles E. Metz, Ph.D. and Robert M. Nishikawa, Ph.D. for their helpful discussions. This work was supported in parts by the American Cancer Society (FRA-390), USPHS grants CA48985 and CA24806, and US Army grant DAMD 17-93-J-3021.

REFERENCES

[1]D. B. Kopans, *Breast Imaging*, (Lippincott, Philadelphia, 1989).

[2]L. W. Bassett, R. H. Gold, *Breast Cancer Detection: Mammography and Other Methods in Breast Imaging*, (Grune &Stratton, New York, 1987).

[3]M. Moskowitz, "Screening for breast cancer: how effective are our test? A critical review," CA Cancer J. Clin. **33**, 26-39 (1983).

[4]H. P. Chan, K. Doi, C. J. Vyborny, *et al.* "Improvement in radiologists' detection of clustered microcalcifications on mammograms: the potential of computer-aided diagnosis," Invest. Radiol. **25**, 1102-1110 (1990).

[5]W. P. Kegelmeyer, J. M. Pruneda, P. D. Bourland, A. Hillis, M. W. Riggs, and M. L. Nipper, "Computer-aided mammographic screening for spiculated lesions," Radiology **191**, 331-337 (1994).

[6]D. B. Kopans, "Standandard mammography reporting," Radiol. Clin. North Am. **30**, 257-261 (1992).

[7]C. J. D'Orsi and D. B. Kopans, "Mammographic feature analysis," Seminars in Roentgenology **XXVIII**, 204-230 (1993).

[8]L. V. Ackerman and E. E. Gose, "Breast lesion classification by computer and xeroradiograph," Cancer **30**, 1025-1035 (1972).

[9]D. Brzakovic, X. M. Luo, and P. Brzakovic, "An approach to automated detection of tumors in mammograms," IEEE Trans. on Med. Imaging **9**, 233-241 (1990).

[10]W. P. Kegelmeyer, "Computer detection of stellate lesions in mammograms," Proc. SPIE **1660**, 446-454 (1992).

[11]J. Kilday, F. Palmieri, and M. D. Fox, "Classifying mammographic lesions using computerized image analysis," IEEE Trans. on Med. Imaging **12**, 664-669 (1993).

[12]C. J. Burdett, H. G. Longbotham, M. Desai, W. B. Richardson, and J. F. Stoll, "Nonlinear indicator of malignancy," **Proc. SPIE 1905**, 853-860 (1993).

[13]E. Claridge, J. H. Richter, P. Stonelake, "Measuring edge blur in mammographic lesions," in: H. U. Lemke, K. Inamura, C. C. Jaffe, and R. Felix (eds), Computer Assisted Radiology, 612-617 (1993).

[14]L.V Ackerman, A. N. Mucciardi, E. E. Gose, and F. S. Alcorn, "Classification of benign and malignant breast tumors on the basis of 36 radiographic properties," Cancer **31**, 342-352 (1973).

[15]A. G. Gale, E. J. Roebuck, P. Riley, and B. S. Worthington, "Computer aids to mammographic diagnosis," Br. J. Radiol. **60**, 887-891 (1987).

[16]D. J. Getty, R. M. Pickett, C. J. D'Oris, and J. A. Swets, "Enhanced interpretation of diagnostic images," Invest. Radiol. **23**, 240-252 (1988).

[17]H. M. Cook, M. D. Fox, "Application of expert systems to mammographic image analysis," Am. J. Physiol. Image **4**, 16-22 (1989)

[18]Y. Wu, M. L. Giger, K.Doi , C. J. Vyborny, R. A. Schmidt, and C. E. Metz, "Application of neural networks in mammography: Application in decision making in the diagnosis of breast cancer," Radiology **87**, 81-87 (1991).

[19]M. L. Giger, C. J. Vyborny, and R. A. Schmidt, "Computerized characterization of mammographic masses: analysis of spiculation," Cancer Letters **77**, 201-211 (1994).

[20]U. Bick, M. L. Giger, R. A. Schmidt, R. M. Nishikawa, D. E. Wolverton, and K. Doi, "Automated segmentation of digitized mammograms," Academic Radiology (in press).

[21]T. Matsumoto, H. Yoshimura, K. Doi, M. L. Giger, A. Kano, H. MacMahon, K. Abe, and S. M. Montner, "Image feature analysis of false-positive diagnoses produced by automated detection of lung nodules," Invest. Radiol. **27**, 587-597 (1992).

[22]J. C. Russ, *The Image Processing Handbooks*. (CRC Press, Boca Raton, 1992).

[23]J. Serra, *Image Analysis and Mathematical Morphology.* (Academic Press, New York, 1982).

[24]C. E. Metz, "ROC methodology in radiologic imaging," Invest. Radiol. **21**,720-733 (1986).

[25]C. E. Metz, "Some practical issues of experimental design and data analysis in radiological ROC studies," Invest. Radiol., **24**, 234-245 (1989).

[26]Z. Huo, M. L. Giger, C. J. Vyborny, *et al.* " Computerized classification of mammographic masses," Radiology (to be submitted).

Figure captions:

Figure 1: Characteristic of masses in our database as rated by an experienced radiologist based on spiculation (a) and shape (b).

Figure 2: 512x512 ROIs centered about an original malignant mass (a) and benign mass (b). The processed images of the malignant (c) and benign (d) masses after background trend correction and histogram equalization. Diagrams of size and circularity of the grown region as functions of gray-level interval (contrast) with the automatically determined "transition point" indicated for the malignant (e) and benign (f) masses. The computer extracted margins overlayed on the malignant (g) and benign (h) masses.

Figure 3 : Neighborhoods used in the radial edge-gradient analysis: A) grown region, B) margin, C) encompassing region and D) surrounding periphery.

Figure 4: Illustration defining the radial angle θ as the angle between the direction of the maximum gradient and its radial direction which is the direction pointing from the center of mass to the point p1, and the radial gradient as the projection of the maximum gradient along the radial direction.

Figure 5: A mammographic circular, smooth mass (a) and its corresponding cumulated edge-gradient distribution (c). A mammographic spiculated mass (b) and its corresponding cumulated edge-gradient distribution (d).

Figure 6: Simulated smooth masses with long-to-short axis ratios of (a) 1:1, (b) 10:6, and (c) 10:5, and simulated spiculated masses (d) slightly spiculated and (e) highly spiculated with long-to-short axis ratio of 1:1 and 1:0.9, respectively.

Figure 7: Dependence of FWHM measure (a) and normalized radial gradient (b) on the long-to-short axis of the simulated smooth masses. Also shown are the values obtained for the two simulated spiculated masses.

Figure 8: Relationship between the corrected FWHM and the normalized radial gradient obtained within the encompassing region (neighborhood C) for the 95 mammographic masses.

Figure 9: ROC curves for the normalized radial gradient measures (a), the FWHM measures (b) and the corrected FWHM measures (c) on a database of 95 mammographic masses for the four neighborhoods showing the performance in classifying malignant and benign masses.

Figure 10: ROC curve of the computer-determined spiculation measures (FWHM) and that of an experienced radiologist's rating of spiculation in terms of their ability to distinguish malignant from benign masses.

Table 1

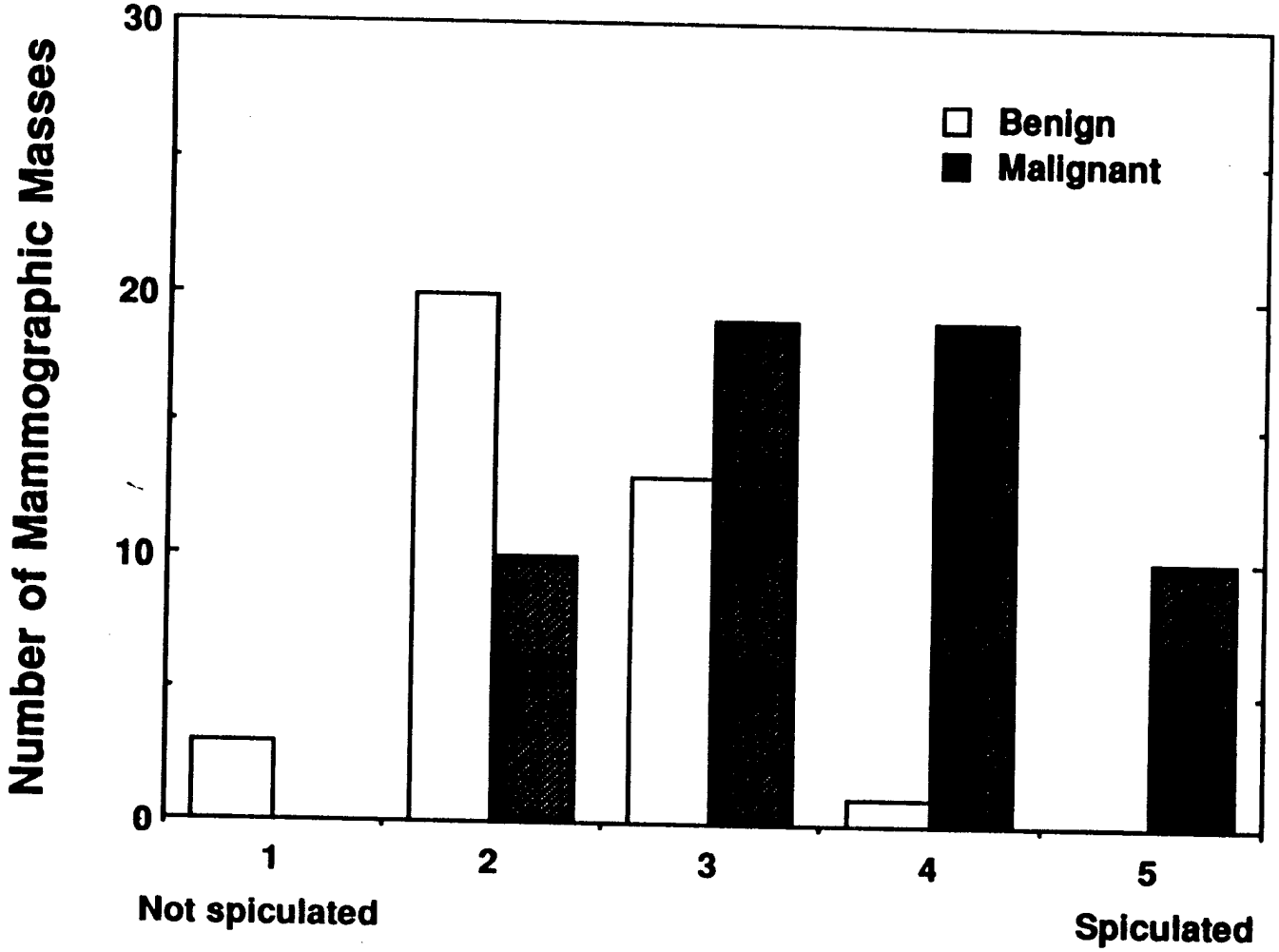
Values of the FWHM and normalized radial gradient for simulated smooth and spiculated masses

Shape (long axis : short axis)	FWHM measure with bin size of 1 degree	FWHM measure with bin size of 15 degrees	normalized radial gradient measure
smooth (1:1)	27	45	0.98
smooth (10:9)	27	45	0.98
smooth (10:8)	37	45	0.97
smooth (10:7)	55	75	0.95
smooth (10:6)	67	75	0.92
smooth (10:5)	87	105	0.88
smooth (10:4)	105	135	0.80
slightly spiculated(1:1)	70	101	0.79
highly spiculated(1:0.9)	133	145	0.72

Table 2

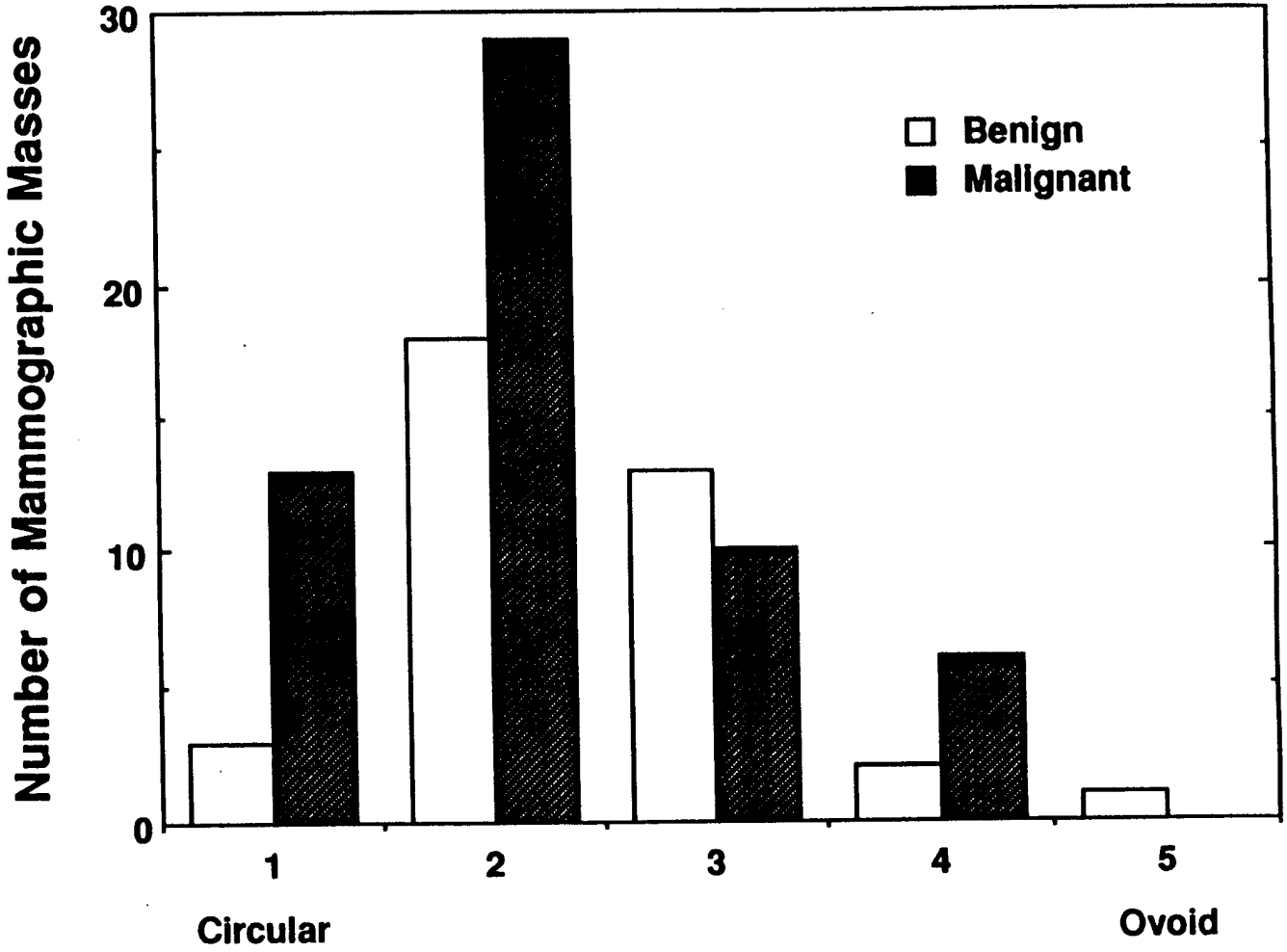
Az values of the two spiculation features extracted in the four neighborhoods (95 mammographic masses)

Measures	Neighborhoods (A) Margin	(B) Grown region	(C) Encompassing region	(D) Surrounding periphery
normalized radial gradient	0.70	0.75	0.80	0.80
uncorrected FWHM	0.70	0.75	0.80	0.83
corrected FWHM	0.73	0.77	0.83	0.85


Number of Mammographic Masses
30
20
10
0
Benign
Malignant
1
2
3
4
5
Not spiculated
Spiculated

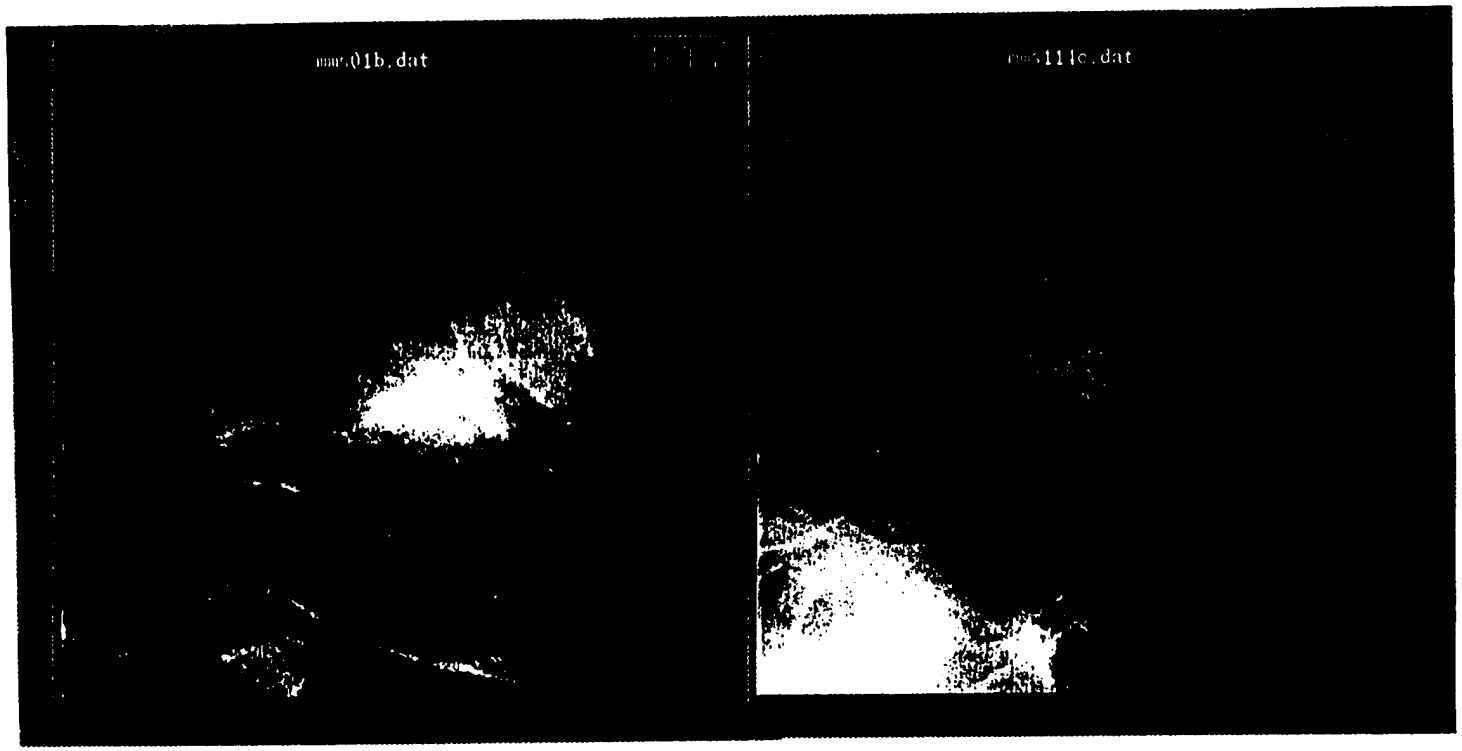
(a)

Fig 1


Number of Mammographic Masses
30
20
10
0
Benign
Malignant
1
2
3
4
5
Circular
Ovoid

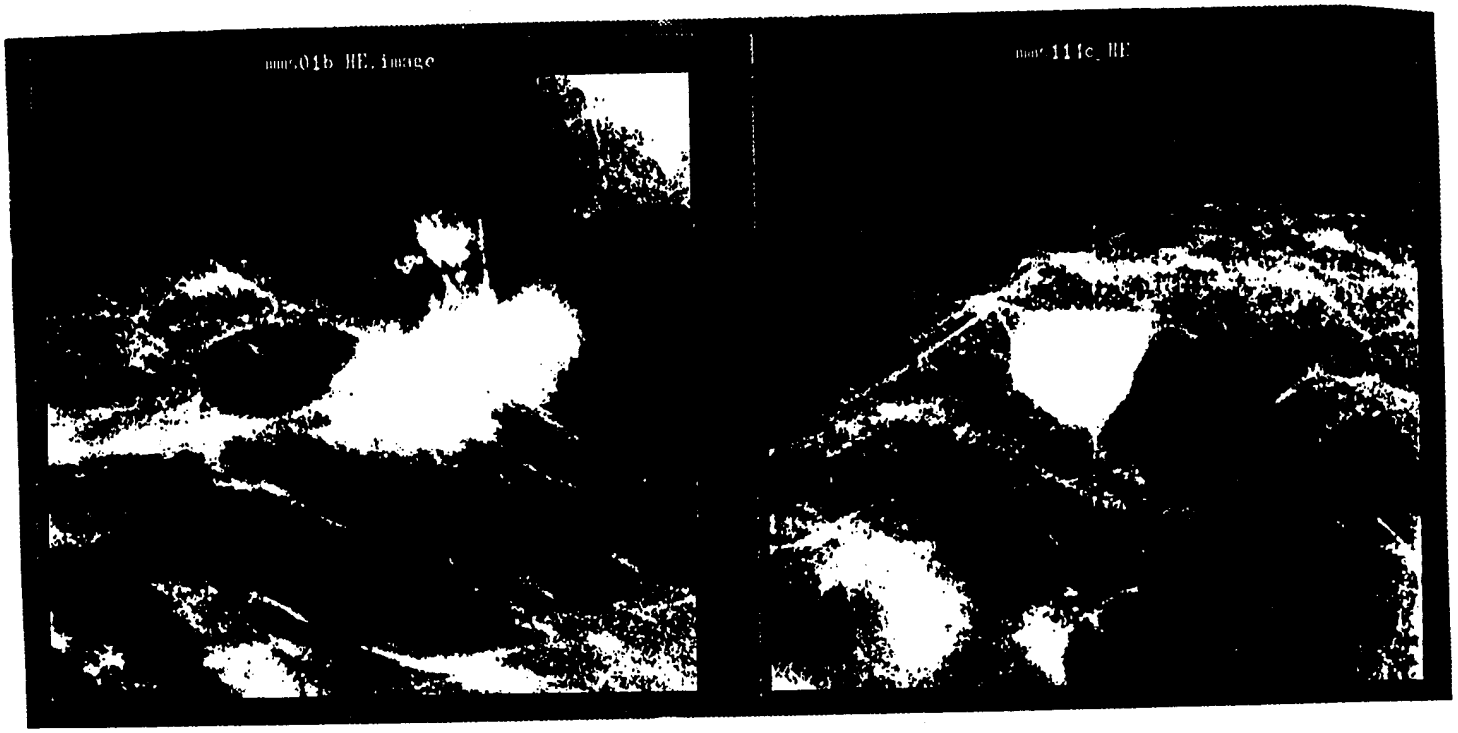
(b)

Fig 1



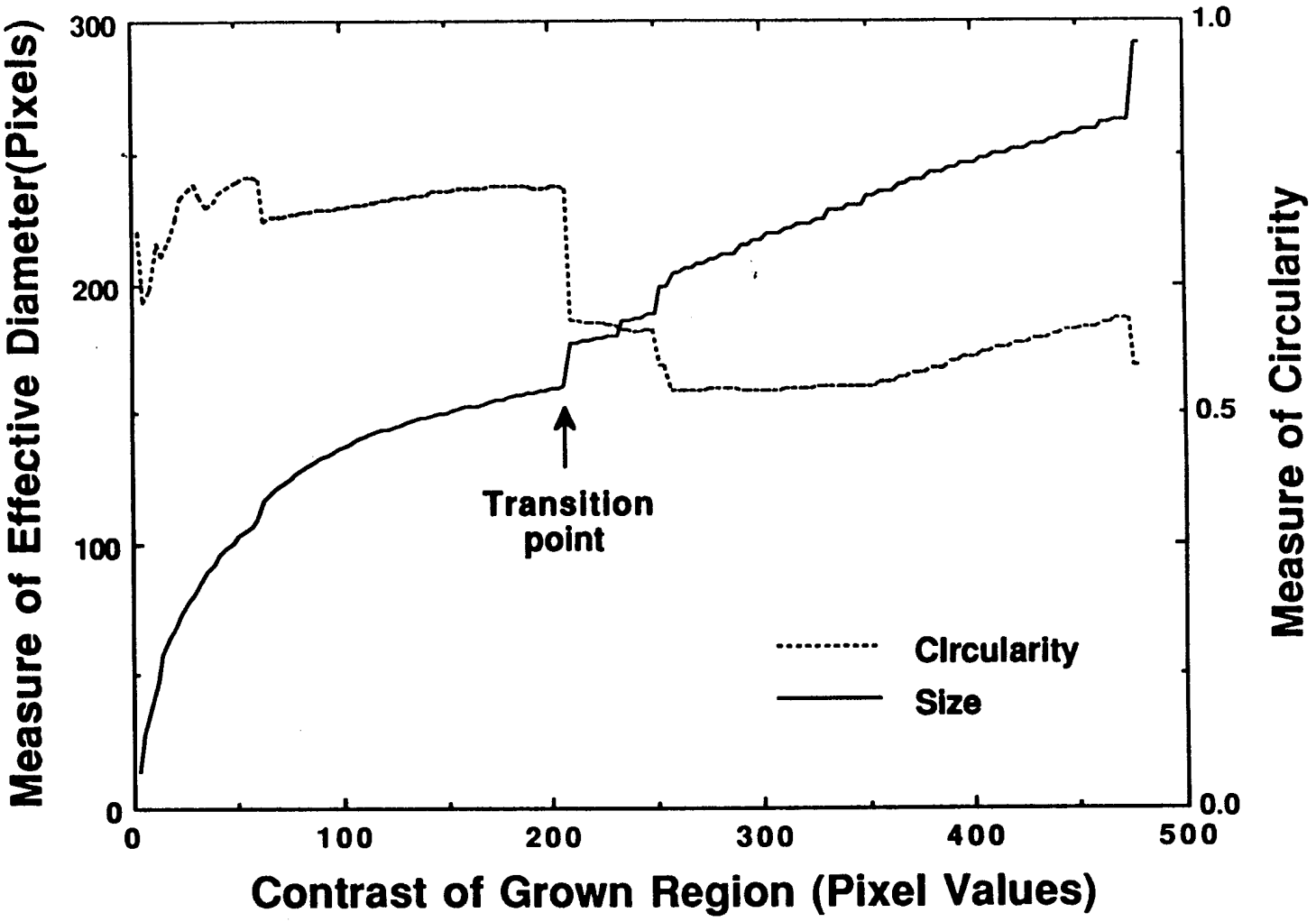
(a) (b)

Fig 2



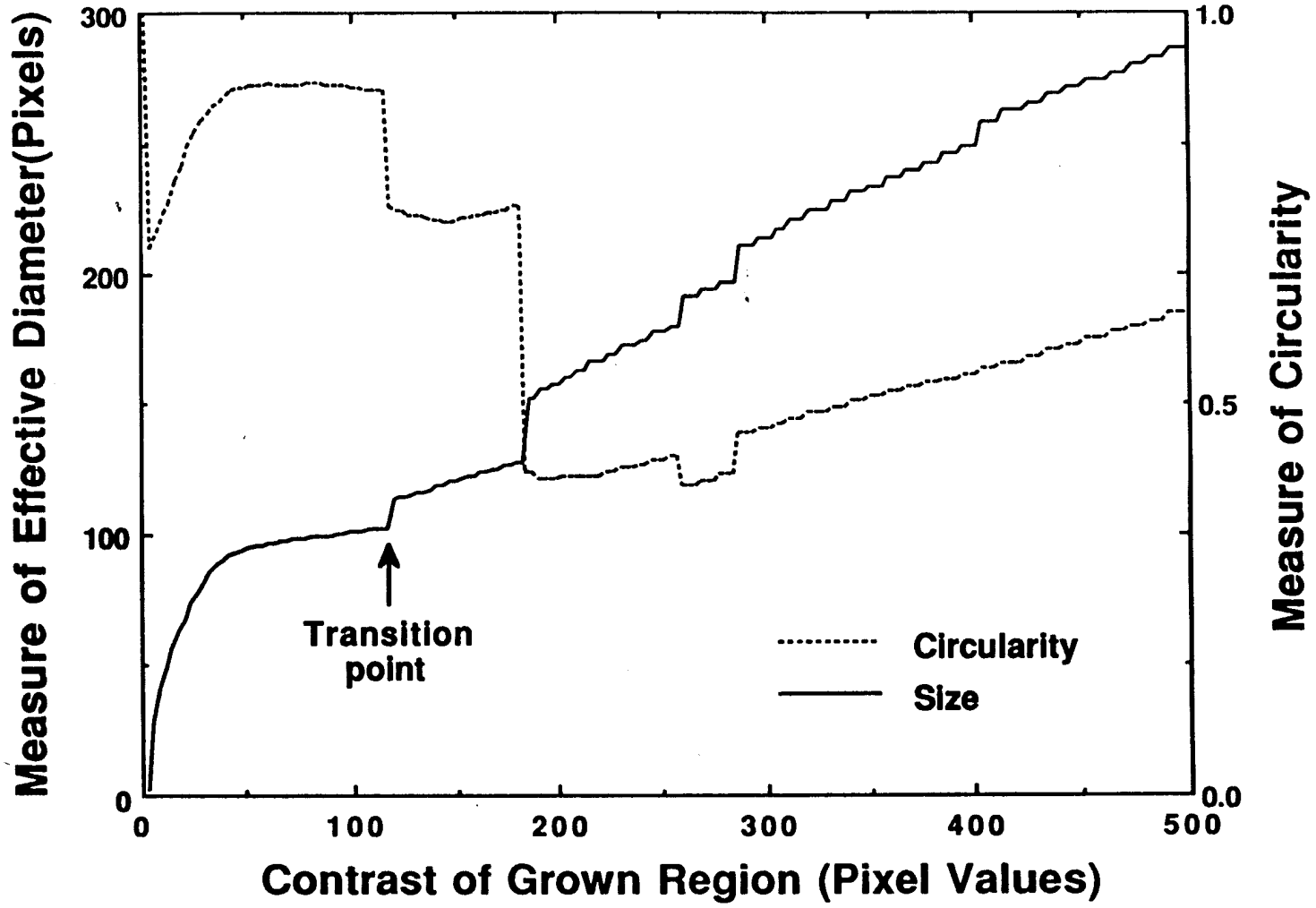
(c) (d)

Fig 2


Measure of Effective Diameter(Pixels)
300
200
100
0
1.0
0.5
0.0
Measure of Circularity
Transition point
Circularity
Size
0
100
200
300
400
500
Contrast of Grown Region (Pixel Values)

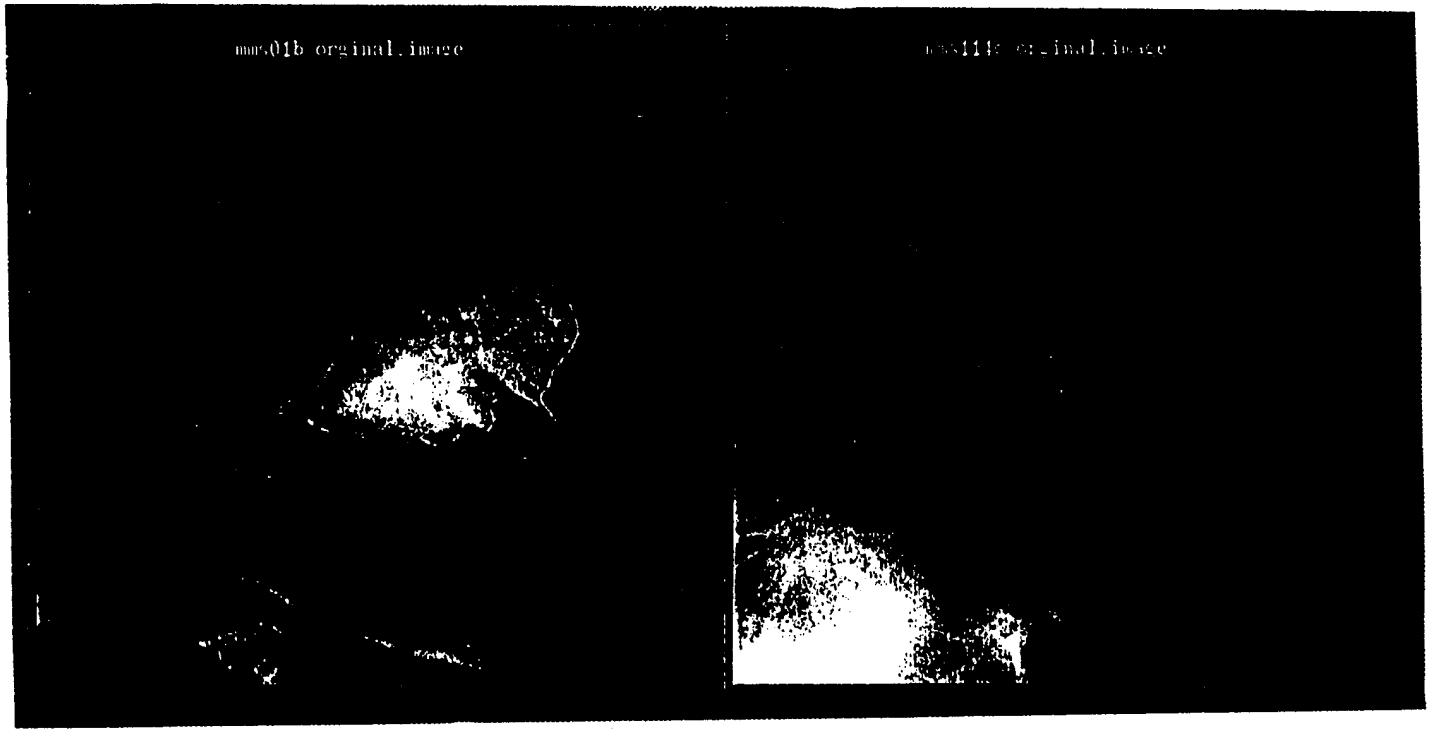
(e)

Fig 2


Measure of Effective Diameter(Pixels)
300
200
100
0
1.0
0.5
0.0
Measure of Circularity
0
100
200
300
400
500
Contrast of Grown Region (Pixel Values)
Transition point
Circularity
Size

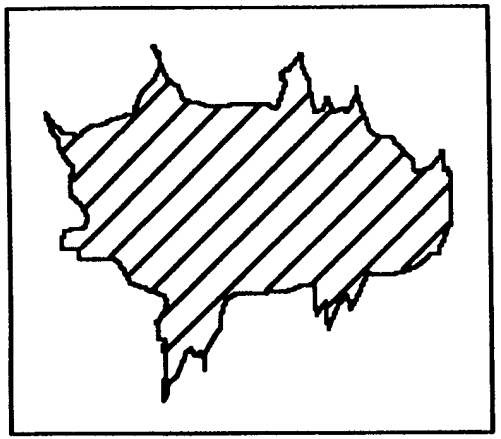
(f)

Fig 2

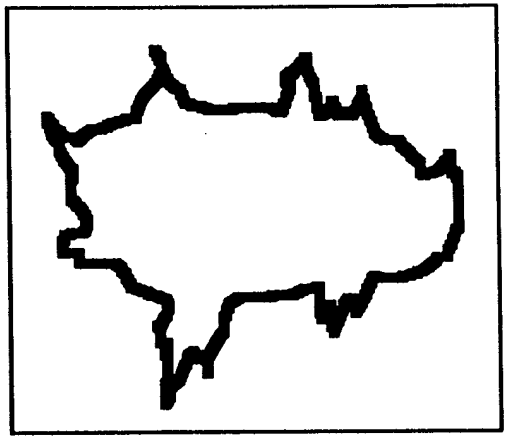


(g) (h)

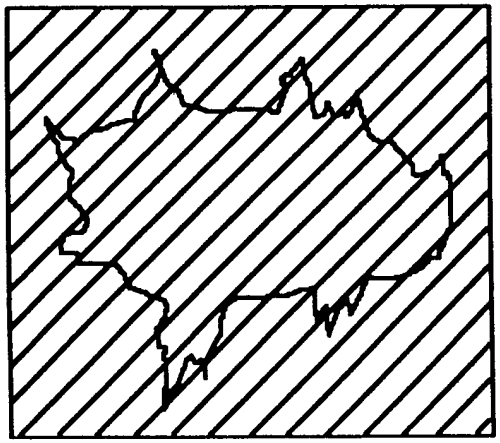
Fig 2



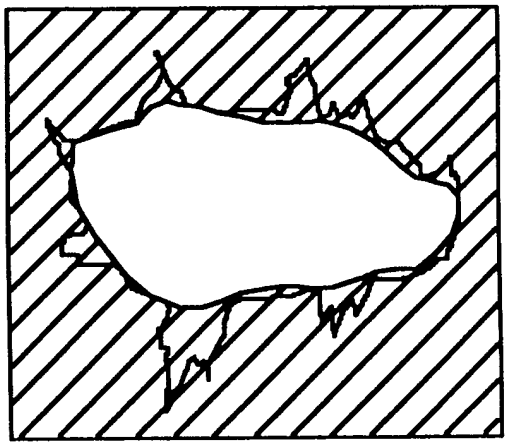
a)



b)



c)



d)

Fig 3

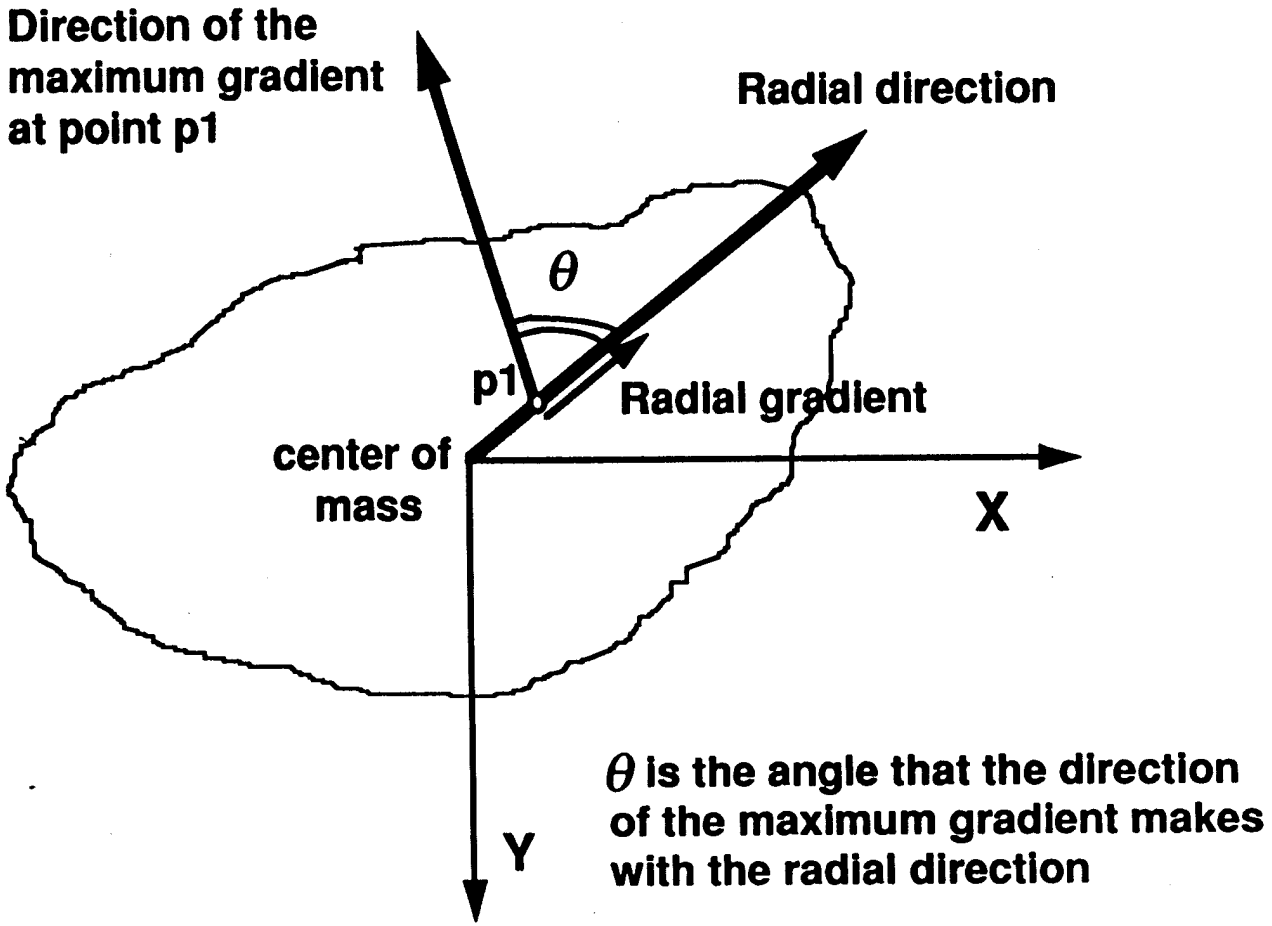
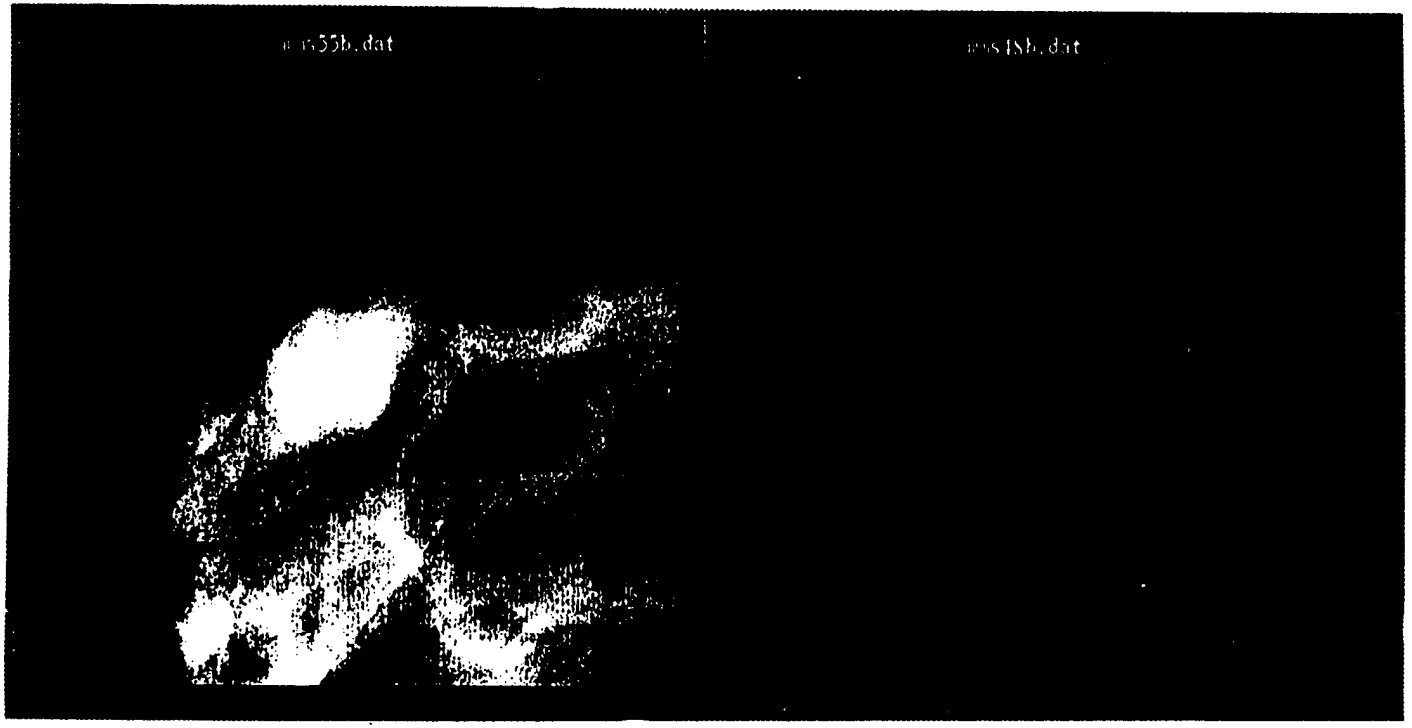
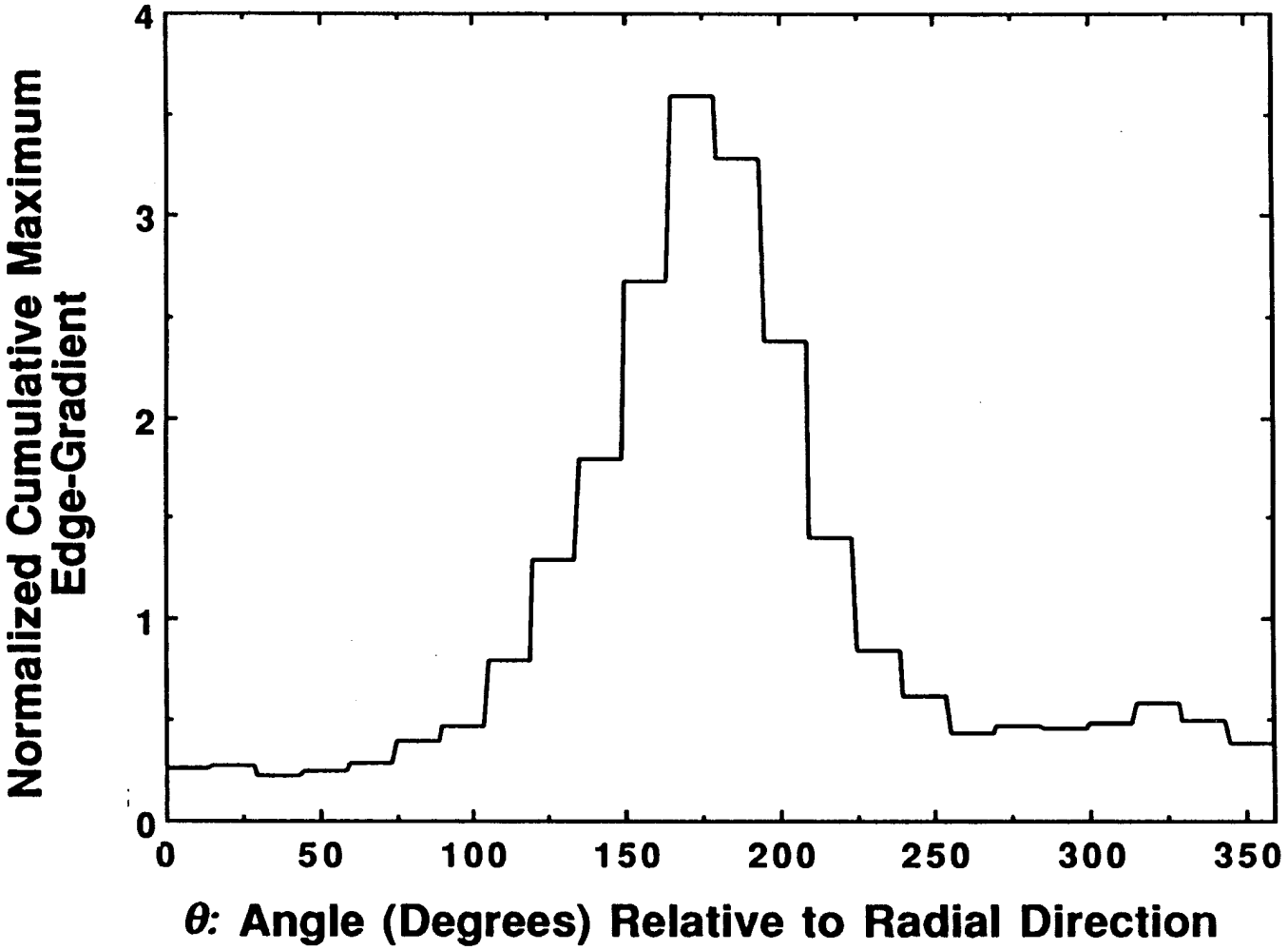

Direction of the
maximum gradient
at point p1
Radial direction
θ
p1
Radial gradient
center of
mass
X
Y
θ is the angle that the direction
of the maximum gradient makes
with the radial direction

Fig 4



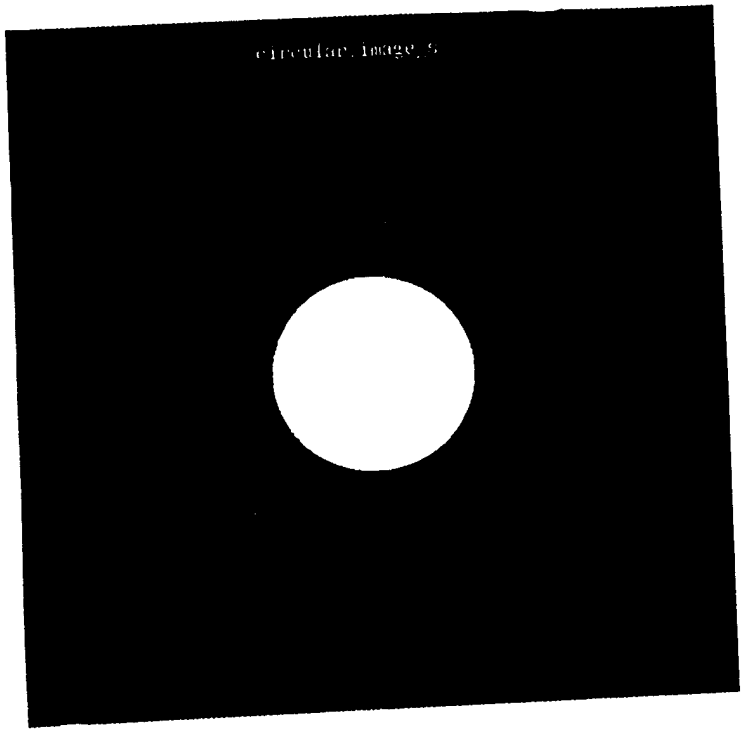
(a) (b)

Fig 5


Normalized Cumulative Maximum
Edge-Gradient
4
3
2
1
0
0
50
100
150
200
250
300
350
θ: Angle (Degrees) Relative to Radial Direction

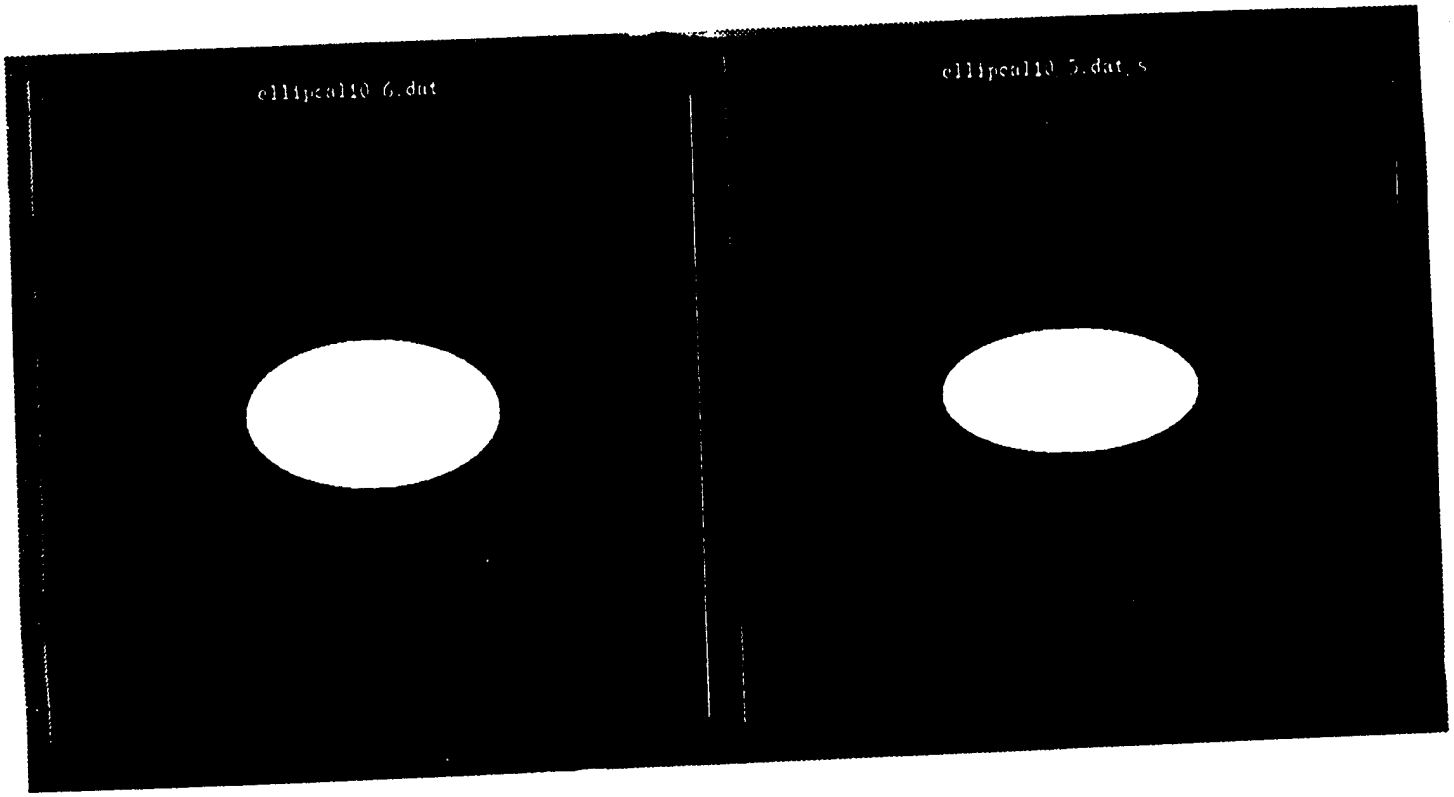
(c)

Fig 5



(a)

Fig 6

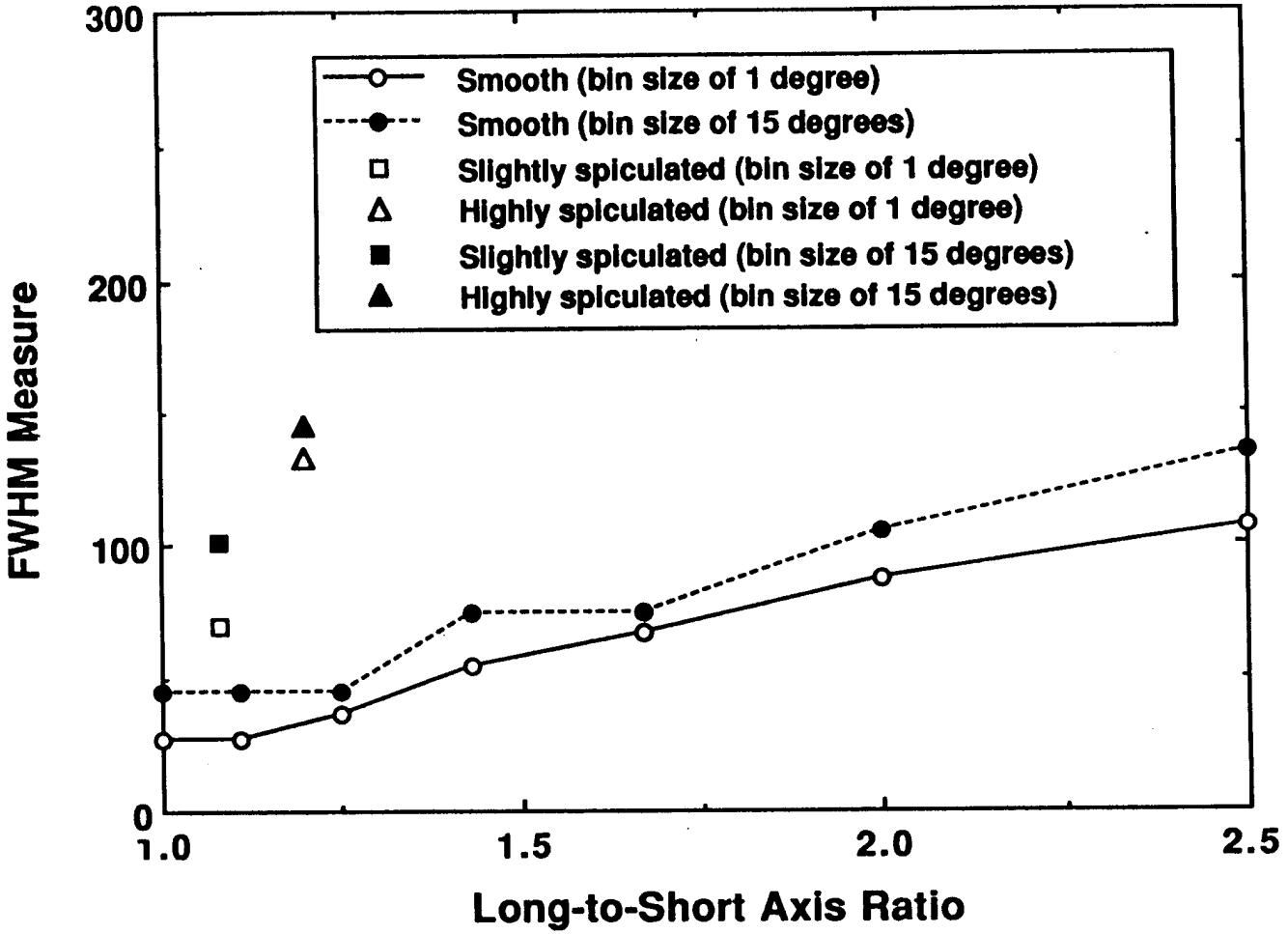


(b) (c)

Fig 6

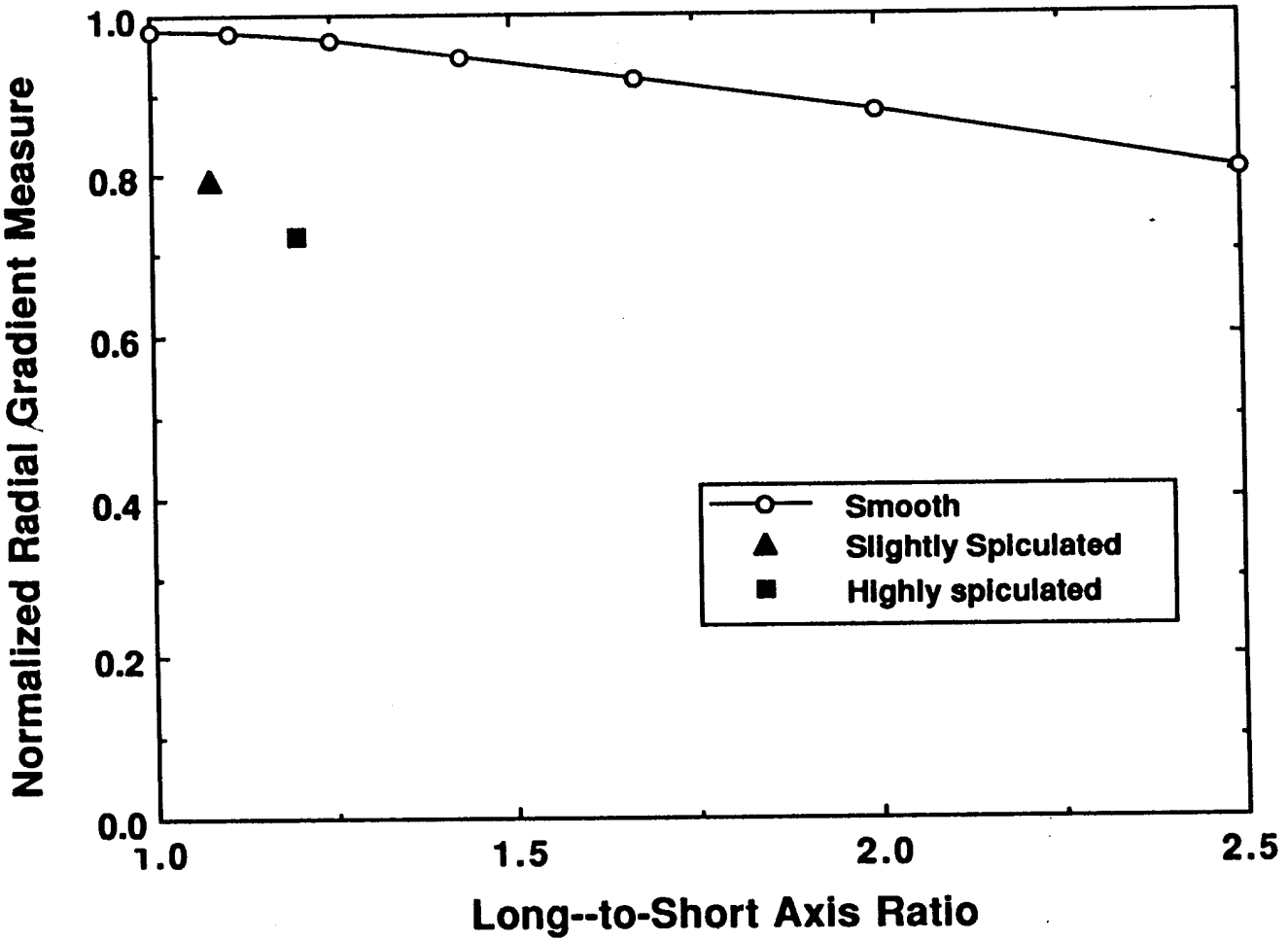
(d) (e)

Fig 6


300
200
100
0
FWHM Measure
1.0
1.5
2.0
2.5
Long-to-Short Axis Ratio
Smooth (bin size of 1 degree)
Smooth (bin size of 15 degrees)
Slightly spiculated (bin size of 1 degree)
Highly spiculated (bin size of 1 degree)
Slightly spiculated (bin size of 15 degrees)
Highly spiculated (bin size of 15 degrees)

(9)

Fig 7


Normalized Radial Gradient Measure
1.0
0.8
0.6
0.4
0.2
0.0
1.0
1.5
2.0
2.5
Long--to-Short Axis Ratio
Smooth
Slightly Spiculated
Highly spiculated

(b)

Fig 7

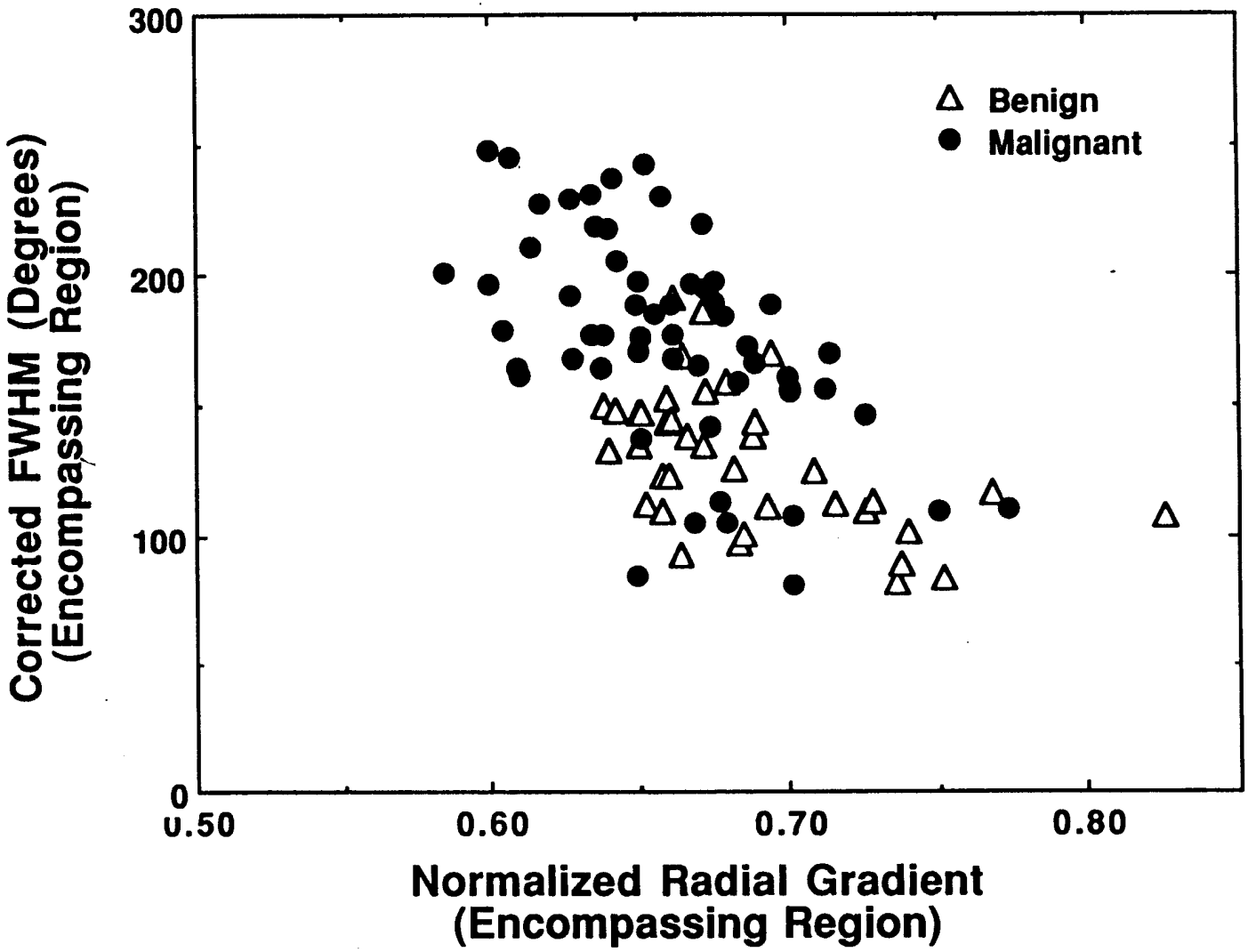
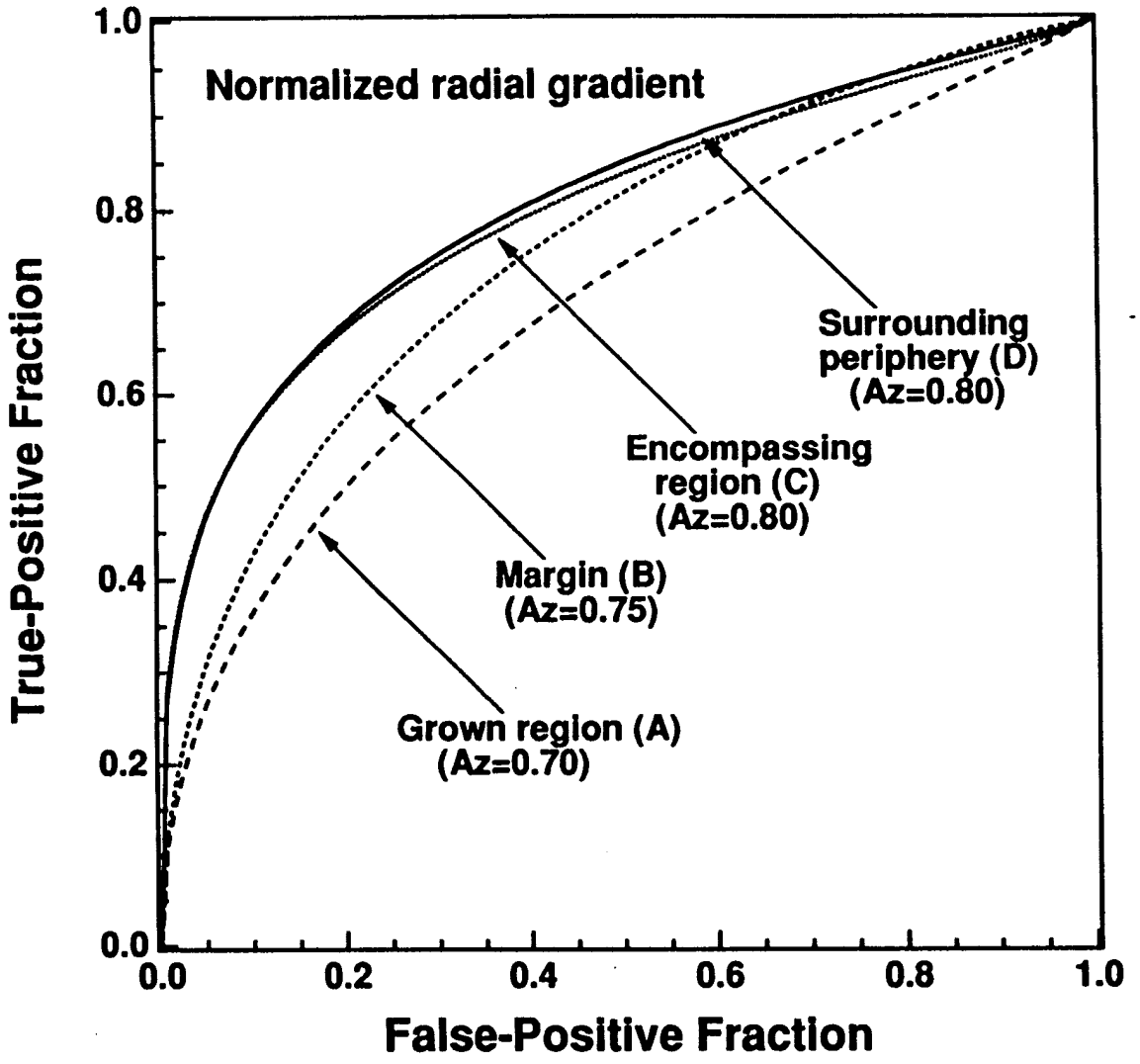
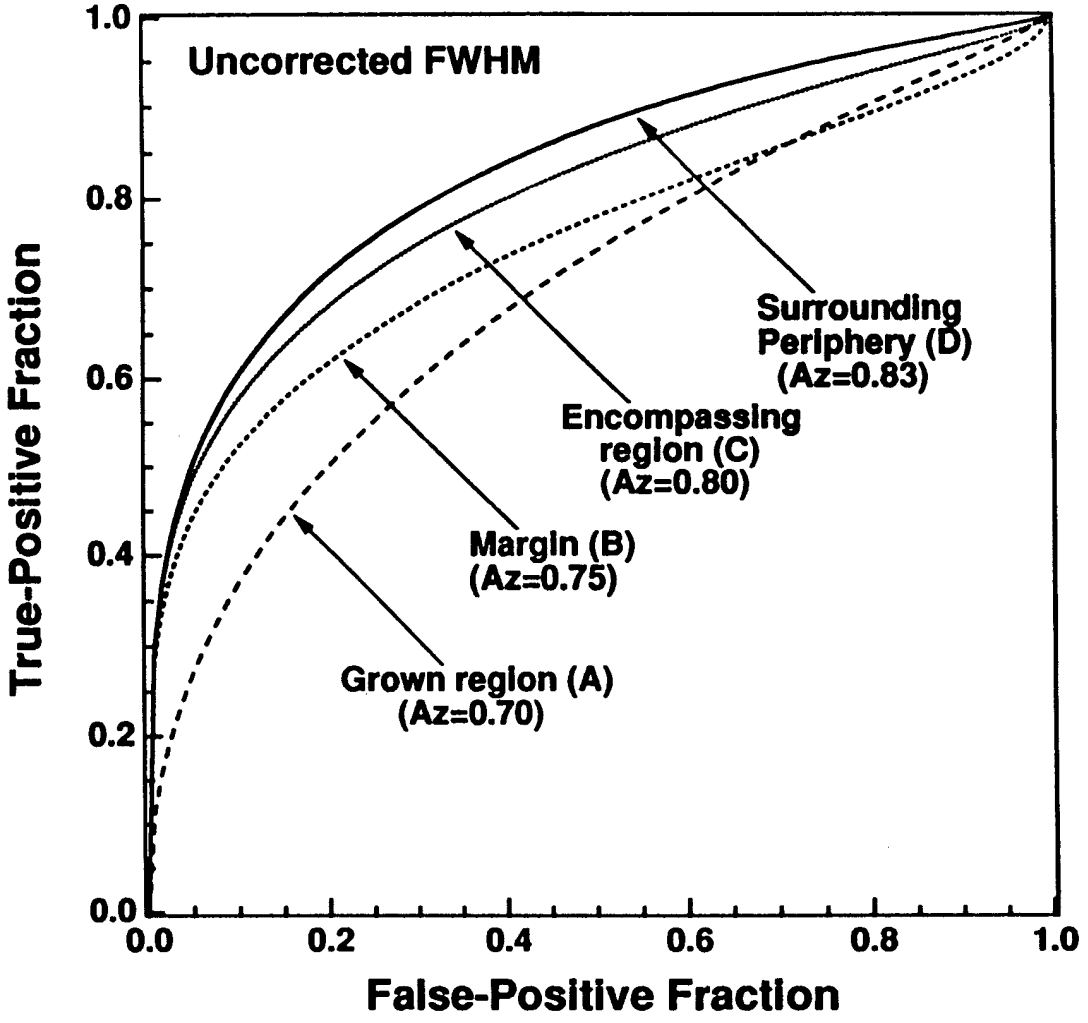

300
200
100
0
0.50
0.60
0.70
0.80
Corrected FWHM (Degrees)
(Encompassing Region)
Normalized Radial Gradient
(Encompassing Region)
Benign
Malignant

Fig 8


Normalized radial gradient
True-Positive Fraction
False-Positive Fraction
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Surrounding periphery (D) (Az=0.80)
Encompassing region (C) (Az=0.80)
Margin (B) (Az=0.75)
Grown region (A) (Az=0.70)

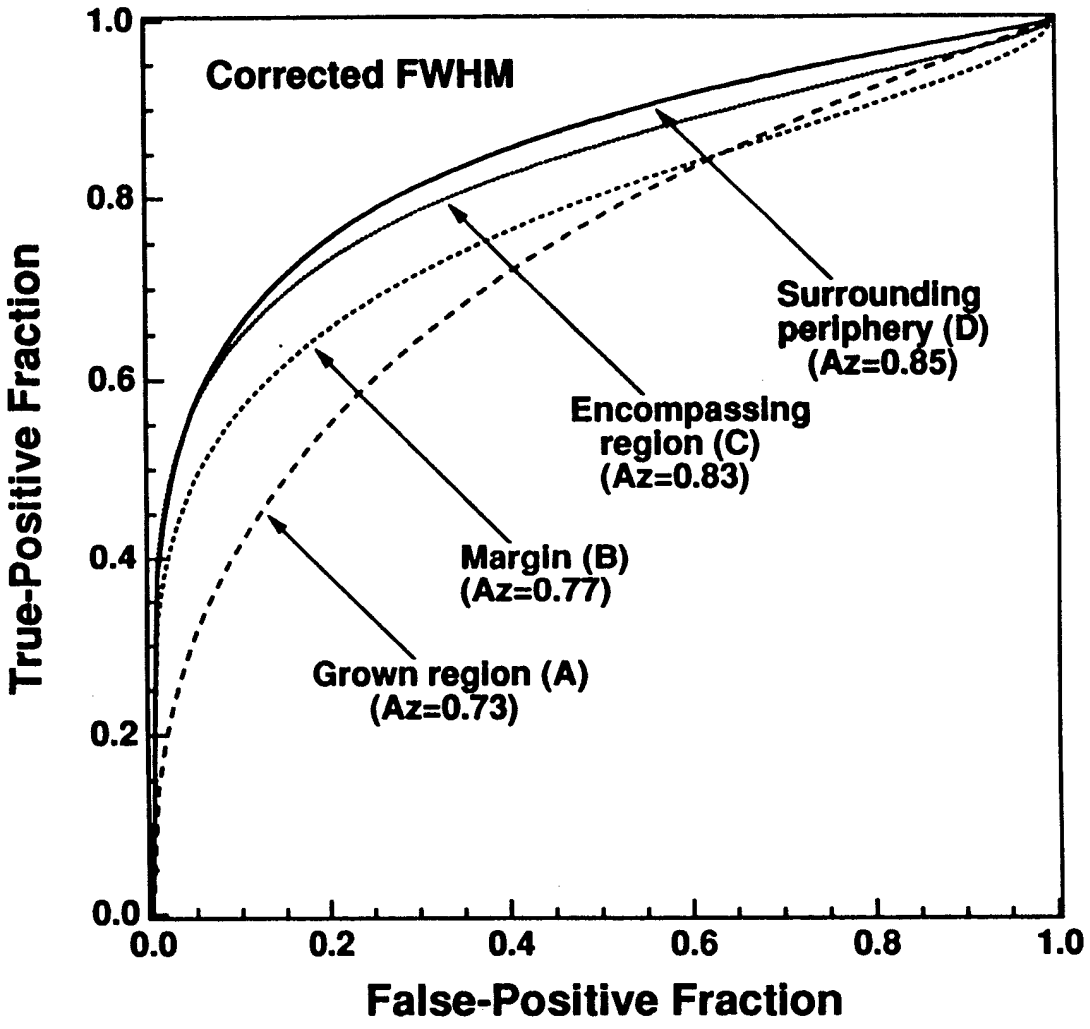
(a)

Fig 9


Uncorrected FWHM
True-Positive Fraction
False-Positive Fraction
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Surrounding Periphery (D) (Az=0.83)
Encompassing region (C) (Az=0.80)
Margin (B) (Az=0.75)
Grown region (A) (Az=0.70)

(b)

Fig 9


Corrected FWHM
Surrounding periphery (D) (Az=0.85)
Encompassing region (C) (Az=0.83)
Margin (B) (Az=0.77)
Grown region (A) (Az=0.73)
True-Positive Fraction
False-Positive Fraction
0.0
0.2
0.4
0.6
0.8
1.0

(C)

Fig 9

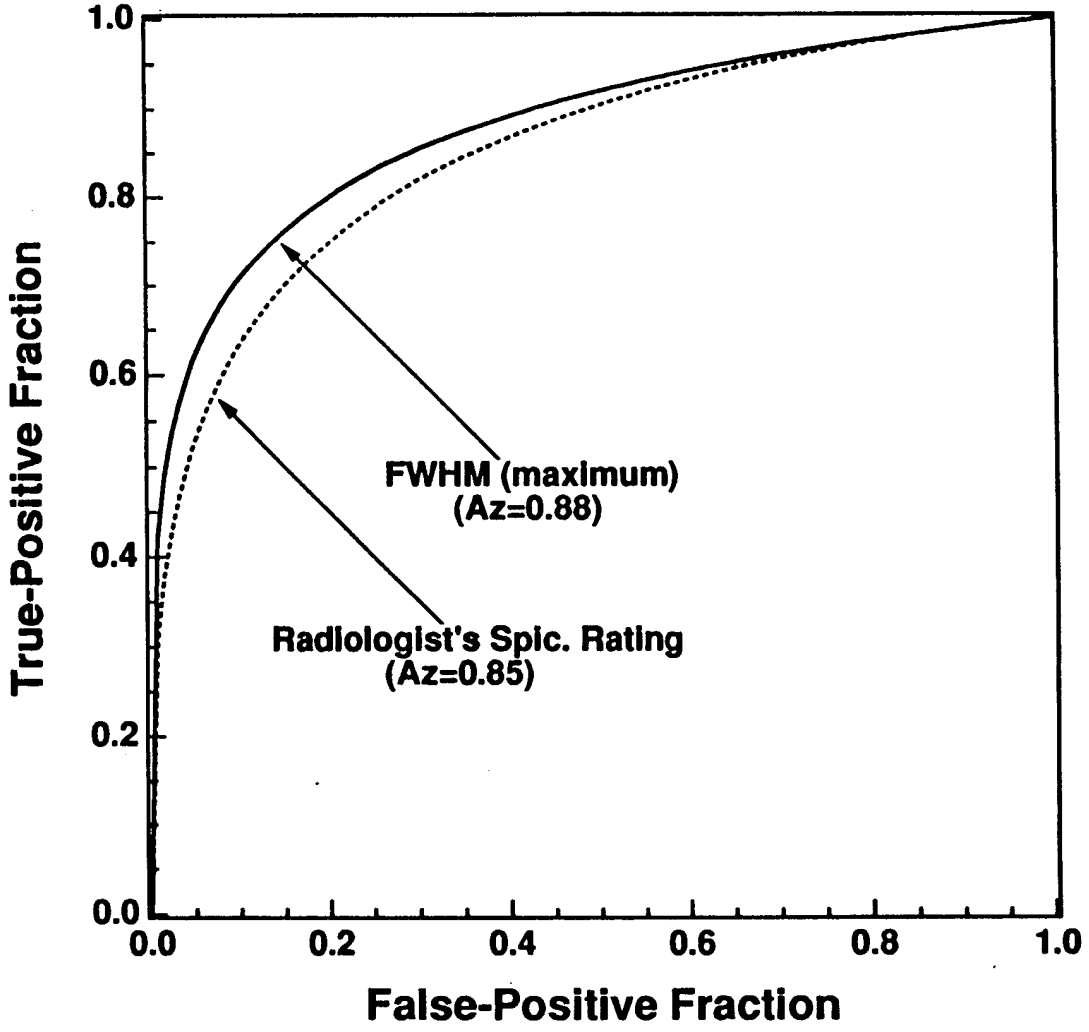

1.0
0.8
0.6
0.4
0.2
0.0
True-Positive Fraction
FWHM (maximum)
(Az=0.88)
Radiologist's Spic. Rating
(Az=0.85)
0.0
0.2
0.4
0.6
0.8
1.0
False-Positive Fraction

Fig 10