CLUSTERED MICROCALCIFICATION DETECTION USING OPTIMIZED DIFFERENCE OF GAUSSIANS

THESIS

Edward M. Ochoa
Captain, USAF

AFIT/GEO/ENG/96D-13

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY
AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GEO/ENG/96D-13

CLUSTERED MICROCALCIFICATION DETECTION USING OPTIMIZED DIFFERENCE OF GAUSSIANS

THESIS
Edward M. Ochoa
Captain, USAF

AFIT/GEO/ENG/96D-13

Approved for public release; distribution unlimited

AFIT/GEO/ENG/96D-13

Clustered Microcalcification Detection Using Optimized Difference of Gaussians

THESIS

Presented to the Faculty of the Graduate School of Engineering

of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science in Electro-Optical Engineering

Edward M. Ochoa, BS

Captain, USAF

December, 1996

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U. S. Government.

Acknowledgements

First, and most importantly, I would like to thank Noriko, my wife, for giving me all her love, support, and encouragement during this time-consuming effort; especially since Nicholas, my son, arrived $1\frac{1}{2}$ months before this research was complete. Noriko, without your generosity, I would not have been able to devote most of my attention to this research. I want to welcome Nicholas to this world, and I'd like to thank him for his patience with his first-time father who has $1\frac{1}{2}$ months of catching up to do.

I also want to thank my mother, Julia Sedillo, my sisters, Lori Armenta and Theresa Esparza, and my grandmother, Gregoria T. Sedillo, for their unending support; especially when my father, Edward Y. Ochoa, passed away halfway through my AFIT experience. Dad, you will always be in our prayers.

I would also like to express my gratitude to Dr. Steve Rogers, my advisor. His uncanny ability to simplify difficult concepts and motivate the desire to learn more has had a tremendous impact on my AFIT experience and my professional outlook. My gratitude also goes to Maj. Jeff Hoffmeister, M.D., my sponsor; without his medical expertise, patience, and engineering insight, this research would not have been possible. I also thank Dr. Martin DeSimio and Dr. Byron Welsh, my committee members, for all their support and encouragement during this research. I also would like to personally thank Tom Rathbun for all his help with genetic algorithms; his advice made the optimization stage of my research possible. I also want to recognize the following other AFIT Breast Cancer Research Team members (past and present) for all their ground-breaking work and support: Bill Polakowski, Ron Dauk, Dru McCandless, Amy Magnus, Randy Broussard, Lem Myers, Jill Leighner, Brian Huether, Don Cournoy, and Kevin Lee. I would also like to thank Toby Reeves for his advice and encouragement throughout our "unique" AFIT experience. And last, but definitely not least, I want to thank Dan Zambon and Dave Doak for all their disk space and workstations; without your support this research would not have been completed before "*the heat-death of the universe*!" Thank you all!

Edward M. Ochoa

Table of Contents

List of Figures

List of Tables

Abstract

Women have a 1 in 8 chance of being diagnosed with breast cancer, and 1 in 30 will die of this disease over her lifetime. Although mammography is the current procedure of choice for breast cancer screening, radiologists fail to detect cancer in 10 to 30 percent of patients with breast cancer. Also, only 10 to 20 percent of patients reffered for biopsy based on mammographic findings prove to have cancer. Further, the malignancies missed by the radiologist are evident in two-thirds of the mammograms retrospectively. To increase sensitivity, a double reading has been suggested. However, the explosion in the number of screening mammograms makes this option unlikely. Alternatively, a computer-aided diagnosis (CADx) system may act as a "second reader" to assist the radiologist in detecting and diagnosing lesions.

Clustered microcalcifications are one of the earliest indicators of breast cancer, and are detected only by mammography; 30 to 50 percent of nonpalpable cancers are mammographically visible on the basis of microcalcifications alone. Furthermore, for early breast cancers, screening studies suggest that 70 to 90 percent were detected based on microcalcifications alone. This research proposes the following methodology for clustered microcalcification detection. First, preprocess the digitized film mammogram to reduce digitization noise. Second, spatially filter the image with a difference of Gaussians (DoG) kernel. To detect potential microcalcifications, segment the filtered image using global and local thresholding. Next, cluster and index these detections into regions of interest (ROIs). Identify ROIs on the digitized image (or hardcopy printout) for final radiologic diagnosis. Finally, to improve detection rates, globally optimize detection parameters using a genetic algorithm (GA), then locally optimize using the simplex method.

The database of 56 digitized (12 bit, 100 μm) full-breast (20×10 cm^2) film mammograms contained 63 biopsy-truthed clustered microcalcification ROIs over 28 cases. This technique demonstrated a true positive (TP) case detection rate of 96.4 percent (27/28), and TP ROI (54/63) and TP image (48/56) detection rates of 85.7 percent with 5.75 false positives (FPs) per full-breast image.

Clustered Microcalcification Detection Using Optimized Difference of Gaussians

I. Introduction

1.1 Background

Mammography, along with physical examination, is the current procedure of choice for breast cancer screening. Screening mammography has been responsible for an estimated 30 to 35 percent reduction in breast cancer mortality rates (58). However, in 1996 approximately 185,700 new breast cancer cases will be diagnosed and 44,300 women will die from this disease (47). Additionally, women have a 1 in 8 chance of being diagnosed, and 1 in 30 will die of this disease over her lifetime (60).

Although radiographic mammography is a well-studied and standardized methodology (35), for 10 to 30 percent of women diagnosed with breast cancer, their mammograms were interpreted as negative (21, 65). Further, the malignancies missed by radiologists were evident in two-thirds of the mammograms retrospectively (21, 65). Missed detections may be attributed to several factors: poor image quality, improper patient positioning, inaccurate interpretation, fibroglandular tissue obscuration (57), subtle nature of radiographic findings, eye fatigue, or oversight (21, 65).

To increase sensitivity, a double reading has been suggested (21, 65). However, the explosion in the number of screening mammograms makes this option unlikely (55). Alternatively, a computer-aided diagnosis (CAD or CADx) system may act as a "second reader" to assist the radiologist in detecting and diagnosing lesions (21, 65).

1.2 Computer-aided diagnosis

Researchers at the Air Force Institute of Technology (AFIT) have been developing modifications to automatic target recognition algorithms used in airborne reconnaissance sensors (19, 51). These modifications allow the technology previously used to identify

targets in sensor imagery to be used to identify suspect areas in mammography films, and to render a recommendation to the radiologist as to the pathology (malignancy) of suspect areas. Previous AFIT CADx studies have resulted in detection successes ranging from 77 to 93 percent (10, 33, 38, 48). The objective of the AFIT CADx system shown in Figure 1 is to improve radiological detection and interpretation of lesions in mammograms. Using features extracted from detected ROIs, artificial-intelligence techniques can then be used to determine the possibility of malignancy (21, 65).

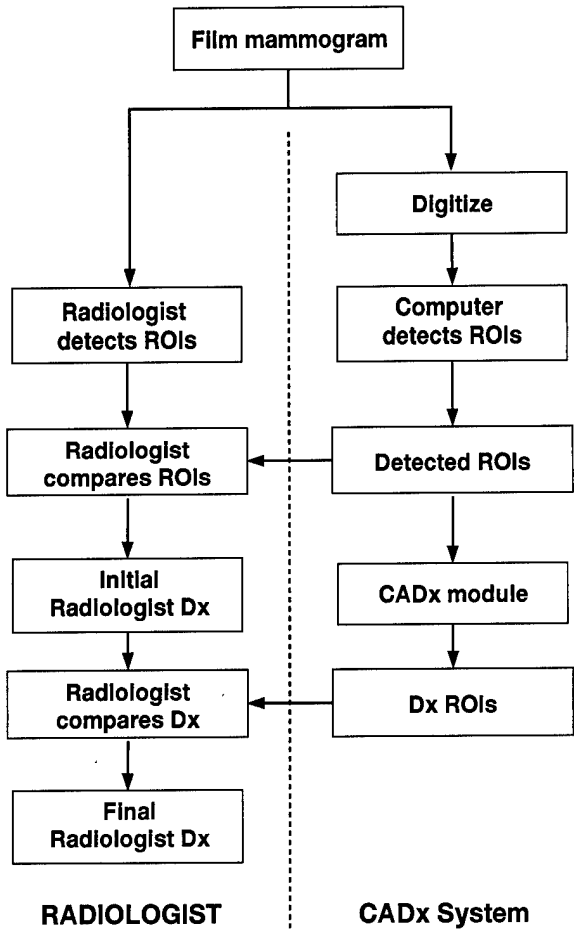


Figure 1. CADx System Methodology: After film digitization, the image is analyzed by a detection algorithm. Computer-detected ROIs are then identified for the radiologists to compare against their own detections. Next, the CADx module provides a computer diagnosis for the radiologist to consider. The radiologist makes the final diagnosis.

1.3 Problem Statement

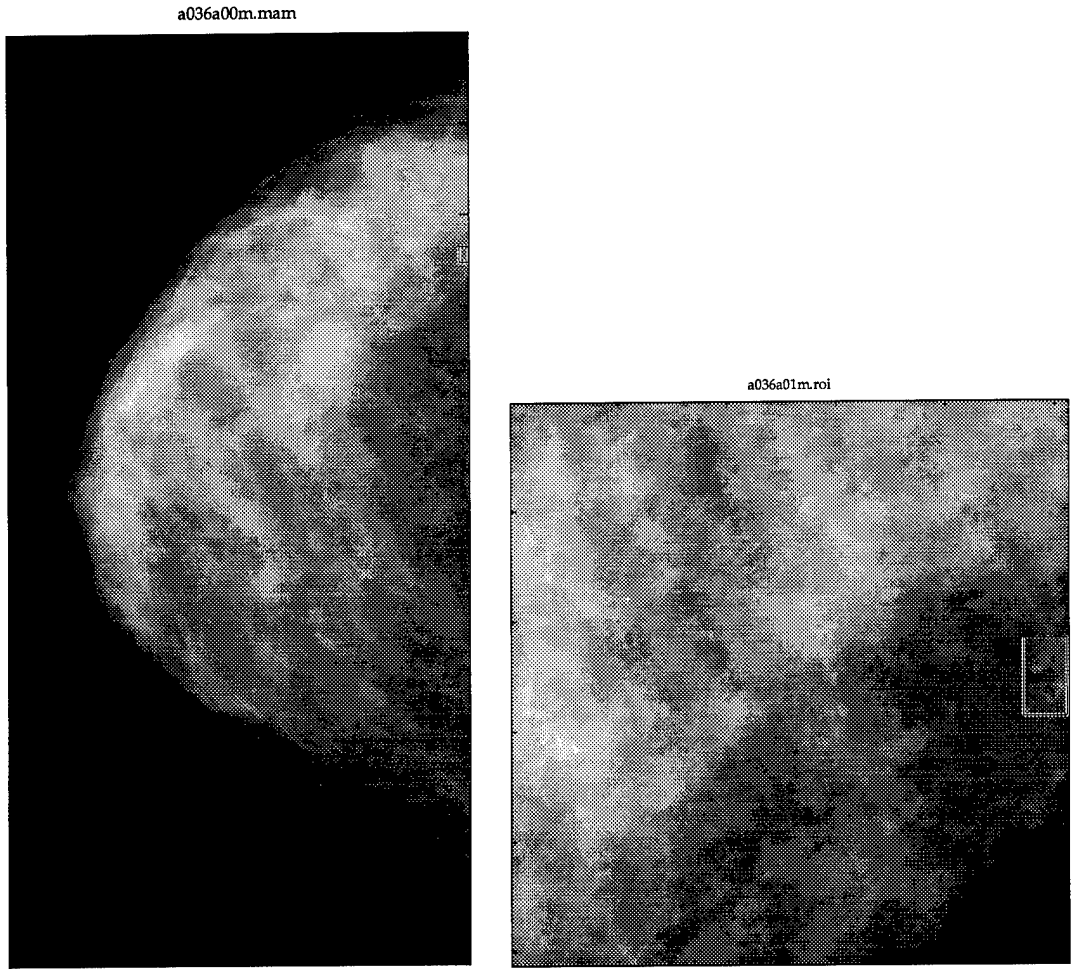
The objective of this thesis is to design an automated microcalcification detection system to be used as an aid in radiologic mammogram interpretation.

1.4 Scope

This research was developed to aid the radiologist in mammogram interpretation; the radiologist always makes the final diagnosis. Also, this thesis makes no attempt to detect or classify masses, or other mammographic densities.

The database of 56 digitized (12 bit, 100 μm) full-breast (20×10 cm^2) film mammograms contained 63 biopsy-truthed clustered microcalcification ROIs over 28 cases. All mammograms included a pathology report indicating location and diagnosis of biopsied regions. The data set included approximately 2 films per case with the following distribution: 1 case with 4 films, 2 cases with only 1 film, and the remaining 25 cases with 2 films.

Full-breast image optimization would have been extremely costly; approximately 30 to 60 minutes per digitized film on a Sun Ultra or Sun Sparc20. Thus, 52 256×256 and 9 512×512 subimages were extracted from these images to facilitate algorithmic development. The relationship between full-breast images, subimages, and ROIs is illustrated by example in Figure 2. The 61 subimages contained all 63 biopsy-truthed ROIs, and a physician (27) annotated abnormalities identified in corresponding pathology reports. These subimages considerably reduced algorithmic development times, and reduced computer memory requirements. Additional information on digitization and database management is located in Appendix A.



(a) (b)

Figure 2. Digitized (12 bit, 100 μm) full-breast film mammogram: (a) 20×10 cm^2 full-breast image: bounding box identifies biopsy-truthed region of interest (ROI); (b) 256×256 pixel subimage: the malignant ROI contained a cluster of microcalcifications.

1.5 Methodology

As shown in Figure 3, the following architecture is used to focus radiologist attention to computer-detected ROIs. First, the digitized film mammogram is preprocessed to reduce digitization noise. Second, the image is spatially filtered with an optimized difference of Gaussians (DoG) kernel. To detect potential microcalcifications, the filtered image is segmented using global and local thresholding. These detections are clustered and indexed into ROIs. Finally, these ROIs are identified for final radiologic diagnosis.

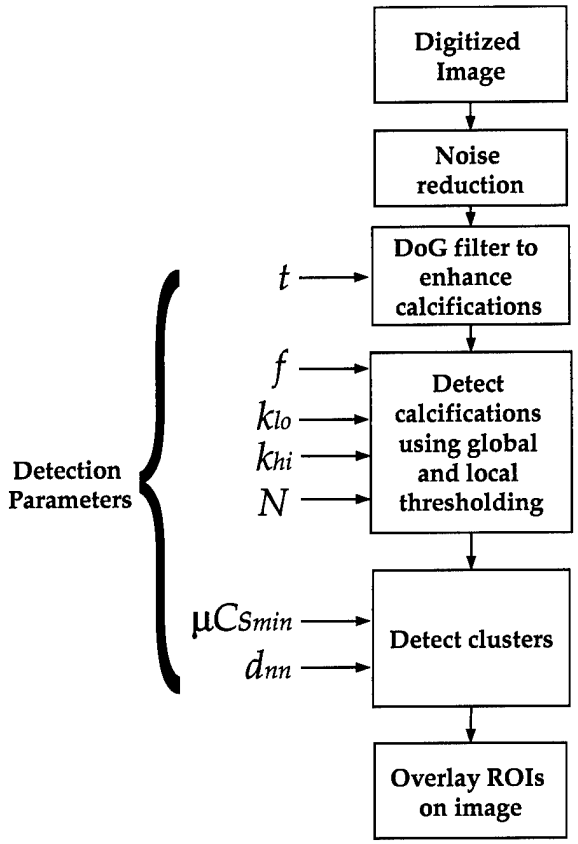


Figure 3. Overview of Cluster Detection Methodology: t controls the DoG filter used to enhance calcifications; f, k_{lo}, k_{hi}, and N control the detection stage of the system; μCs_{min} and d_{nn} control the clustering module.

1.6 Overview

Chapter II reviews breast cancer and related CADx research efforts. Chapter III provides an overview of detection methodology. Chapter IV reviews detection results and analyis. Conclusions are drawn in Chapter V. Appendix A contains digitization and database information used in this research. Appendix B contains Matlab and C code used in this research.

II. Background

This chapter reviews breast cancer and related CADx research. Specifically, this chapter reviews the following topics:

- Breast cancer: statistics, mammographic characterization, radiologist detection and diagnosis of breast lesions
- Related CADx research: computer detection of microcalcifications

2.1 Breast cancer

Breast cancer results from failure of normal regulation of cell differentiation and proliferation (34). According to Kopans (34), "Breast cancer kills through metastatic spread and compromise of the function of other organs." Since many cancers require time to develop the ability to spread successfully to other organs (34), early detection is the goal in screening mammography (60).

2.1.1 Statistics. Screening mammography has been responsible for an estimated 30 to 35 percent reduction in breast cancer mortality rates (58). However, in 1996 approximately 185,700 new breast cancer cases will be diagnosed and 44,300 women will die from this disease (47). Additionally, women have a 1 in 8 chance of being diagnosed, and 1 in 30 will die of this disease over her lifetime (60). Although mammography is the current procedure of choice for breast cancer screening, radiologists fail to detect cancer in 10 to 30 percent of patients with breast cancer (21, 65). Also, only 10 to 20 percent of patients reffered for biopsy based on mammographic findings prove to have cancer (55).

2.1.2 Mammographic characterization. To improve radiologist performance and reduce mortality rates, researchers are pursuing computerized analysis to aid in mammogram interpretation (55). Radiologic interpretation of mammograms includes detection and diagnosis of suspicious masses and clustered microcalcifications. However, once an abnormality is detected, one of the first challenges is to determine whether the abnormality is actually present (2, 16). Numerous overlapping structures in the breast form summation shadows that may mimic true abnormalities (8, 16).

2.1.2.1 Masses. A mass is a space-occupying lesion shown in two different projections; otherwise, the term "density" may be used until the three dimensional features are confirmed (2, 16). Densities that mimic masses include pectoral muscle, nodular breast tissue, and summation shadows (2). Breast masses may be characterized by shape, margins, density relative to surrounding tissue, associated findings, and location (2). According to Adler (2), "Masses are three-dimensional, distinct from the surrounding tissue, and most often asymmetric compared with the contralateral breast." Mammmographically seen margins are one of the major determinants of benign or malignant status because masses with indistinct margins have a higher risk for malignancy (2, 16). Additional features used for interpretation of masses are discussed in more detail in Adler's (2) and Dorsi's (16) articles.

2.1.2.2 Microcalcifications. Microcalcifications are calcium deposits typically on the order of 0.1 to 0.3 mm in diameter (8, 56) and are detected only by mammography (20). Recognition of their presence is the challenge for radiologists because 30 to 50 percent of nonpalpable cancers are visible on the basis of microcalcifications alone (20). Furthermore, for early breast cancers, screening studies suggest that 70 to 90 percent were detected based on microcalcifications alone (20). (The terms microcalcification and calcification will now be used interchangeably to simplify discussion of characteristics.)

As mentioned earlier, one of the first steps in radiologic analysis of calcifications is to determine whether they have been detected. Mimics or pseudocalcifications include deodorant, talcum powder, pickoff of film emulsion, dust particles, and scratches (20). After false detections are eliminated, calcifications are analyzed for potentially benign or malignant features.

Feig (20) lists the following examples of potentially benign calcifications:

- "Eggshell or rim calcifications" typically have a round calcific rim surrounding a hollow sphere.
- "Spherical or lucent-centered calcifications" have thicker walls than eggshell, have smooth surfaces, and are round or oval.

- "Vascular calcifications" typically appear in parallel producing a "railroad track" appearance. These may be difficult to distinguish from potentially malignant fine linear calcifications if one wall of the vessel is calcified, or if the segment is short or discontinuous.
- "Large rod-like calcifications" follow the course of the ducts, radiating from the retroareolar area and are frequently bilateral.
- "Coarse or popcorn-like calcifications" appear as a circumscribed mass with coarse, mostly round calcifications.
- "Milk of calcium calcifications" are layers of calcium in the dependent portion of cysts, and may be meniscus shaped when seen on the mediolateral (ML) view and less distinct on the craniocaudal (CC) view.
- "Skin calcifications" typically appear as 1 to 2 mm lucent centered or solid spheres.
- "Dystrophic calcifications" have a bizarre shape, typically result from surgery or radiation therapy, and may be difficult to distinguish from potentially malignant calcifications.

Shape, distribution, size, and contour are the major criteria used to distinguish malignant from benign calcifications (20). Feig (20) and Shaw de Paredes (56) list the following characteristics of potentially malignant calcifications (assuming one or more are present):

- "Linear shaped calcifications" are typically considered suspicious; however, some benign conditions may also produce similar calcifications (20).
- "Linearly distributed or segmentally distributed calcifications" are extremely suspicious since most malignant calcifications form within ducts, and over 90 percent of invasive cancers arise from ducts (20).
- "Markedly clustered and unilateral calcifications" are also extremely suspicious since malignant calcifications usually appear with this feature (20). The greater the number of calcifications within an area, the more suspicious (56). As Shaw de Paredes notes, "malignant calcifications tend to occur in tight clusters of 1 cm diameter or

less" (56). However, Shaw de Paredes cautions: "...the disease and the calcifications can be extensive, involving an entire quadrant or an entire breast" (56). Finally, solitary clusters are typically considered for biopsy if no similar calcifications exist elsewhere in the breast, or in the opposite breast (20).

2.1.3 Radiologist detection and diagnosis of breast lesions. Although radiographic mammography is a well-studied and standardized methodology (35), for 10 to 30 percent of women diagnosed with breast cancer, their mammograms were interpreted as negative (21, 65). Further, the malignancies missed by radiologists were evident in two-thirds of the mammograms, retrospectively (21, 65). Missed detections may be attributed to several factors: poor image quality, improper patient positioning, inaccurate interpretation, fibroglandular tissue obscuration (57), subtle nature of radiographic findings, eye fatigue, or oversight (21, 65).

To increase sensitivity, a double reading has been suggested (21, 65). However, the explosion in the number of screening mammograms makes this option unlikely (55). Alternatively, a computer-aided diagnosis (CADx) system may act as a "second reader" to assist the radiologist in detecting lesions and making diagnostic decisions (21, 65).

2.2 Related CADx research: clustered microcalcification detection

To improve radiologist performance, researchers are pursuing computerized analysis to aid in mammogram interpretation (55). As mentioned earlier, one research area of interest is detection and diagnosis of microcalcifications. Although researchers generally desire the same goal, several methodologies exist for each stage of the detection and/or diagnosis process.

2.2.1 Database. Several research groups are currently investigating detection of clustered microcalcifications (9, 44, 46, 64, 68). Most investigators use different databases; thus, it is not possible to compare results directly (44). Additionally, no universally accepted criteria exists for the number of bits required to represent available gray levels in films or minimum resolution required for detection of clustered microcalcifications. How-

ever, several researchers use a sample spacing of 100μm per pixel, with a gray-level resolution of 10 to 12 bits per pixel (9, 18, 44, 64, 68).

There is also a wide variation in the area of mammograms used for analysis. Image area is an important factor in increasing the false positive detection rate since the larger the test area, "...the greater the chance for producing false positives in general" (18). Typical mammogram image areas generally range from 8×10 cm^2 to approximately 10×10 cm^2 (9, 18, 44, 68); however, robust detection algorithms should not be limited by image area. As mentioned in Chapter I, we use approximately 20×10 cm^2 (2K×1K pixel) full-breast images.

In the literature, several CADx researchers (9, 44, 46, 68, 64) develop and test detection techniques on the same database (not to be confused with a common database). Nishikawa *et al.* (45) have shown "...the choice of clinical cases used to train and test a computer-aided diagnosis (CAD) scheme can affect the test results (i.e. error rate)." Nishikawa *et al.* (45) also argue, "Because of the strong dependence of measured performance on the testing database, it is difficult to estimate reliably the accuracy of a CAD scheme. Furthermore, it is questionable to compare different CAD schemes when different cases are used for testing."

Although a common database has been suggested, this would require standardization of pixel size and number of gray levels used (45). Also, if the common database were enlarged, all previously evaluated CADx techniques would need to be reevaluated if images from new digitized imaging or film-screen systems were added (45). Furthermore, Nishikawa *et al.* argue, "...because there is no quantitative or rigorous means of determining whether a particular common database is representative of the clinical population, it is difficult to predict the true clinical impact of a CAD scheme."

2.2.2 Cluster detection criteria. No universally accepted criteria exist for scoring detections of clustered microcalcifications (63, 64). Nishikawa *et al.* (44) rely on visual interpretation for determining true positive (TP) detections since "...it is not possible to determine the $x-y$ location of microcalcifications that are not visible in the mammogram." Nishikawa *et al.* do not define their criteria for false positive (FP) detections.

Strickland *et al.* (63, 64) rely on the cluster detection criteria proposed by Karssemeijer (32). "A cluster is considered detected if two or more microcalcifications are found in the region of film (truth circle) identified by the radiologist" (63, 64). Strickland *et al.* continue, "...a false positive (FP) is counted if two or more erroneous detections are made within an empty, closed region of 0.5 cm in width (63, 64). As Strickland *et al.* note, "One weakness of this metric is that at high sensitivity thresholds the number of false positives may actually drop due to merging of previously separated false clusters" (64). However, this usually occurs at FP rates too high to be of any clinical value (64).

Chan *et al.* (9) score a signal detected if it was within 0.5 mm of a true microcalcification. Further, a cluster is scored as a TP if its centroid is within a cluster radius of 5 mm from the centroid of a true cluster and at least two of its member microcalcifications are scored as a TP. Once a microcalcification or cluster is considered detected, it is eliminated from further matching.

2.2.3 Noise reduction. In screen-film mammography, the film is used for detection, storage, and display (15). Since full-area detectors suitable for digital mammography are still in the experimental stage, an alternative way of acquiring mammograms in digital form is to digitize a film mammogram (67). Although the main limitation in image quality should be the granularity of the film emulsion (67), noise is introduced from the process of digitization (15). This noise may later be detected as a pseudocalcification.

Although noise reduction may be performed prior to processing, Nishikawa *et al.* (44) and Yoshida *et al.* (68) apply a noise reduction technique during the detection process. To reduce the effects of noise, they propose use of morphological erosion to eliminate detections less than 3 pixels in area, since very small signals are likely to be due to random noise, and not by the presence of true calcifications.

2.2.4 Calcification enhancement. To enhance calcifications, several CAD microcalcification detection methods employ a difference image technique; subtraction of a signal suppressed image from a signal enhanced image (9, 18, 44). In practice, the spatial kernels used for enhancement and suppression are combined into a single linear filter, equivalent

to bandpass filtering (18). Several of these methods use a box-rim based kernel to enhance calcifications (9, 18, 44).

Strickland *et al.* (63, 64) and Yoshida *et al.* (68) use wavelets to enhance calcifications. Strickland et al. (63, 64) use sub-band decomposition to obtain individual wavelet matched filter responses. Yoshida *et al.* (68) use wavelet techniques to decompose, modify, then reconstruct wavelet-processed mammograms. Finally, other techniques which have been successfully applied use locally adaptive or region growing methods to enhance contrast in film mammograms (12, 13, 14, 42, 66).

2.2.5 Detection of signals. Once an image has been preprocessed to enhance differences between targets and background, the image is more easily segmented. Several researchers threshold their filtered images to detect potential calcifications (44, 64, 68). However, thresholding schemes typically vary according to the methodology used for enhancement.

Nishikawa *et al.* (44) and Yoshida *et al.* (68) use adaptive thresholding based on local statistics of gray-level values within a 51×51 mm area. Pixels are kept if their value exceeds 2 to 3.8 times the local standard deviation plus the local mean (44, 68). Nishikawa *et al.* (44) use a 64×64 mm region centered on each signal to reduce false positives. If the first moment of the averaged power spectrum exceeds 3 mm^{-1}, the signal is discarded. Yoshida *et al.* (68) use texture analysis to reduce false detections.

Since Strickland *et al.* (63, 64) generate multiple images during processing, they pool individually thresholded images to establish a test statistic at each pixel location. Each image is thresholded at some fixed percentile of the corresponding histogram (the authors do not report these corresponding percentiles). These images are then summed pixel-wise. Pixels with values of 3 or more are passed to the final detection map.

2.2.6 Detection of clusters. Recall, Feig (20) listed "markedly clustered and unilateral calcifications" as extremely suspicious since malignant calcifications usually appear with this feature. Chan *et al.* (9) and Nishikawa et al (44, 46) also argue true microcalcifi-

cations of clinical interest typically appear in clusters. Thus, these investigators use fairly similar methodologies.

Chan *et al.* (9) use a 1 cm diameter neighborhood criteria; if the number of signals found within the neighborhood is greater than an input minimum number, the signal is kept. Similarly, Nishikawa *et al.* (44, 46) cluster detected signals by passage of a 32×32 mm box over the entire image. If 3 or more signals exist within the box, the box is passed to the output image as a ROI.

2.2.7 Calcification detection results. Chan *et al.* (9) evaluate the success of their convolution neural network at detecting clustered microcalcifications over 52 100μm, 12-bit films (size/area not reported). To demonstrate the dependance of the performance of their CAD scheme on the database, they (subjectively) divide their films into obvious, average-subtle, and subtle sets. For the obvious set, they report a TP rate of 100 percent with less than 0.1 FPs per image. For the average-subtle set, they achieve a 93 percent TP rate with one FP per image. Finally, for the subtle set, they report a TP rate of 87 percent with 1.5 FPs per image.

Nishikawa *et al.* and Yoshida *et al.* use the same database, as Yoshida's article is an earlier investigation of a wavelet enhancement technique (44, 68). Yoshida *et al.* achieve a detection rate of 85 percent with 5 FPs per image over 39 100μm, 800×1000 pixel (8×10 cm^2), 10-bit images (68). Similarly, over this identical dataset, Nishikawa *et al.* report a TP cluster detection rate of 85 percent with 2.5 FPs per image (44, 46).

Strickland *et al.* (63, 64) use 40, 50 μm, 12-bit, 2048×2048 images. Because Strickland *et al.* used 50 μm/pixel images, this corresponds to a 10×10 cm^2 region on a film mammogram. One of their image pairs accounts for 27 percent of the total clusters in the database (104 total clusters). Thus, they report results with and without this pair. Over all 40 images, they achieve a TP detection rate of approximately 84 percent with approximately 3 false clusters per image. (Approximations are made from the cluster detection performance curve included in their article (64).) Without this pair, at approximately the same FP rate, they achieve a TP detection rate of approximately 90 percent.

Table 1. Clustered microcalcification ROI detection results; NR indicates not reported.

Investigators	Database (# images)	Image Area (cm^2)	Resolution (μm/pixel)	TP rate (%)	FP rate (#/image)
Chan *et al.* (overall)	52	NR	100	92	1
Yoshida *et al.*	39	8×10	100	85	5
Nishikawa *et al.*	39	8×10	100	85	2.5
Strickland *et al.*	40	10×10	50	84	3

2.3 Summary

As shown in Table 1, several candidate microcalcification detection techniques have been developed yielding promising results; however, no single method has been developed which is able to reliably detect all microcalcification clusters. An alternative may be to fuse the results from several complimentary techniques. Pattern recognition research at AFIT focuses on developing algorithms for use in detection and classification of objects in military imagery (17, 26, 37, 51, 52, 59). Several of these methodologies have also been applied to the area of breast cancer detection and diagnosis (10, 33, 38, 48). A new technique which uses DoG filtering for detection of clustered microcalcifications is presented in the next chapter.

III. Methodology

3.1 Introduction

This chapter presents a new technique for detecting and labeling clustered microcalcifications in digitized film mammograms.

3.2 Algorithm development

In the literature, several CADx researchers (9, 44, 46, 68, 64) develop and test detection techniques on the same database (not to be confused with a common database). This research adopts the same approach. Although partitioning the database into a training and testing set was considered, Nishikawa *et al.* (45) have shown "...the choice of clinical cases used to train and test a computer-aided diagnosis (CAD) scheme can affect the test results (i.e. error rate)." Nishikawa *et al.* (45) also argue, "Because of the strong dependence of measured performance on the testing database, it is difficult to estimate reliably the accuracy of a CAD scheme. Furthermore, it is questionable to compare different CAD schemes when different cases are used for testing."

Although a common CADx digitized mammogram testing database has been suggested, this would require standardization of pixel size and number of gray levels used (45). Also, if the common database were enlarged, all previously evaluated CADx techniques would need to be reevaluated if images from new digitized imaging or film-screen systems were added (45). Furthermore, Nishikawa *et al.* argue, "...because there is no quantitative or rigorous means of determining whether a particular common database is representative of the clinical population, it is difficult to predict the true clinical impact of a CAD scheme." Thus, AFIT researchers are currently facilitating a clinical study to determine the impact of this research as well as other AFIT CADx schemes (10, 33, 38, 48).

3.3 Database

In screen-film mammography, the film is used for detection, storage, and display. (15). Since full-area detectors suitable for digital mammography are still in the experimental stage, an alternative way of acquiring mammograms in digital form is to digitize

a film mammogram (67). This research was performed using digitized film mammograms and subimages annotated by a physician trained in radiology (27).

Specifically, the database of 56 digitized (12 bit, 100 μm) full-breast (20×10 cm^2) film mammograms contained 63 biopsy-truthed clustered microcalcification ROIs over 28 cases. All mammograms included a pathology report indicating location and diagnosis of biopsied regions. The data set included approximately 2 films per case with the following distribution: 1 case with 4 films, 2 cases with only 1 film, and the remaining 25 cases with 2 films.

Full-breast image optimization would have been extremely costly; approximately 30 to 60 minutes per digitized film on a Sun Ultra or Sun Sparc20. Thus, 52 256×256 and 9 512×512 subimages were extracted from these images to facilitate algorithmic development. The relationship between full-breast images, subimages, and ROIs is illustrated by example in Figure 2. The 61 subimages contained all 63 biopsy-truthed ROIs, and a physician (27) annotated abnormalities identified in corresponding pathology reports. These subimages considerably reduced algorithmic development times, and reduced computer memory requirements. Additional information on digitization and database management is located in Appendix A.

3.4 Cluster detection criteria

As Strickland *et al.* (64) note, no universally accepted criteria exist for the detection of clustered microcalcifications. Thus, for each of the subimages extracted from mammograms with corresponding pathology, a physician identifies the bounding box of each cluster (as shown in Figure 4a) and stores the top left and bottom right vertices in a truth file. This technique for creating "ROI truth" streamlines tracking of detection results.

For counting TPs, a cluster is considered detected if any cluster signal is found within a bounding box and the cluster's centroid is within 5 mm from the bounding box. A FP is counted if all detections within the cluster are not made within a bounding box. TP and FP detection rates are used to determine receiver operator characteristics (ROC) for performance measurements of radiologic imaging systems (31, 39, 40). Typically,

these systems are optimized such that the average TP rate is maximized while minimizing the average FP rate. Judy's (31) and Metz's (39, 40) articles provide a more complete discussion of radiologic ROC analysis.

3.5 Noise reduction module

Although the main limitation in image quality should be the granularity of the film emulsion (67), noise is introduced from the process of digitization (15). This noise may be detected later as a pseudocalcification. There are several noise reduction techniques available in the literature (3, 53, 61). In this research, median filtering is used because this technique has been shown to be extremely effective at removing noise (3). However, the type of median filter used differs from the typical since it is cross-shaped. The cross shape is formed by the set of pixels which include the center pixel and its four nearest neighbors. The cross shape preserves lines and corners better than the standard block-shaped median filter (3, 61). Additionally, the cross shape limits the possible substitution to the four nearest neighbors, reducing the potential for severe edge displacement (3). The subimage in Figure 4a was processed with this technique, and results are shown in Figure 4b.

3.6 Calcification enhancement module: DoG filter

After noise has been reduced, the image is DoG filtered. This section discusses DoG filtering by example. First, the physiological basis for and construction of a DoG filter is identified. Next, DoG spatial kernels are generated based on potential target sizes. These filters are then individually applied to the noise reduced subimage shown in Figure 4b.

3.6.1 Physiological basis. In chapter two of *Vision* (36), Marr reviews requirements for and efficient detection of intensity changes, identifies the physiological basis for the DoG, and (indirectly) provides an engineering rule of thumb for construction of DoG filters. To detect intensity changes efficiently, Marr argues one should search for a filter which is a differential operator and is tunable to act at any desired scale (36). Raff (49) provides further insight into Marr's argument. In short, two basic ideas underlie detection of intensity changes in images. First, intensity changes occur at different scales in images,

(a)

(b)

Figure 4. 256×256 subimage: (a) ROI bounding box contains biopsy-truthed malignant cluster of microcalcifications; (b) noise-reduced subimage

and second, sudden changes in intensities result in peaks of first derivatives and hence "zero crossings" in the second derivatives.

Marr argues the Laplacian of Gaussian (LoG) is the optimum filter which satisfies these conditions (36). There are two basic steps to LoG filtering. First, use a Gaussian to effectively wipe out all structure at scales smaller than the space constant (standard deviation) of the Gaussian. Second, use a Laplacian as a differential operator to detect intensity changes in the blurred image (36). As Marr notes, the benefit of using a Gaussian is that it is smooth and localized in both the space and frequency domains. Additionally, the benefits of using the Laplacian are that it is an orientation-independent operator, and it reduces the extra computational burdens associated with using second-order directional derivative operators (36). Next, Marr identifies the biological implications of the LoG filter.

Neurophysiological experiments provide evidence the human visual pathway includes a set of "channels" that are orientation and spatial frequency selective (36, 49, 61). Essentially, at each point in the visual field, there are four size-tuned filters or masks analyzing the image (36). The operation of these spatial receptive fields (in the form of retinal ganglion cells) can be approximated closely by a DoG (36, 49, 61). Marr argues LoG filters form the basis for these psychophysically determined channels and the LoG operator can be approximated closely by a DoG when the ratio of the DoG space constants is 1:1.6 (36).

3.6.2 DoG parameters. The 2-D Gaussian smoothing operator is given as

$$G(x,y) = ce^{\frac{-(x^2+y^2)}{2\sigma^2}} \tag{1}$$

where c normalizes the sum of mask elements to unity, x and y are horizontal and vertical indices, and σ is the space constant (61). Using Equation 1, the difference of two Gaussians with different σ yields

$$DoG(x,y) = c_1 e^{\frac{-(x^2+y^2)}{2\sigma_1^2}} - c_2 e^{\frac{-(x^2+y^2)}{2\sigma_2^2}} \tag{2}$$

For a target of size (average width) t pixels, use

$$\sigma_2 = \frac{t}{2} \tag{3}$$

and, from Marr's ratio (36), use

$$\sigma_1 = \frac{\sigma_2}{1.6} \tag{4}$$

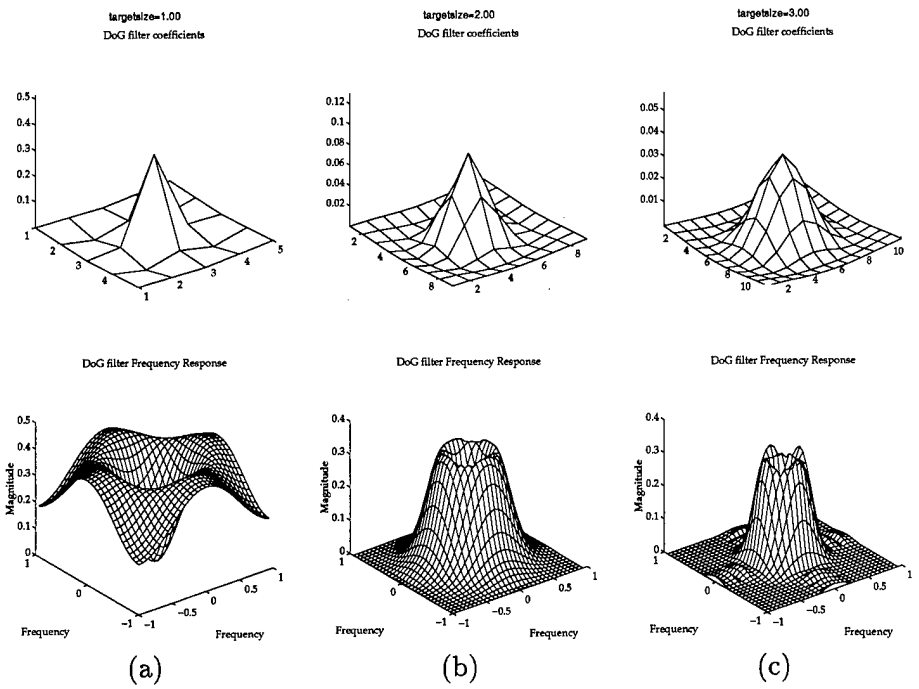
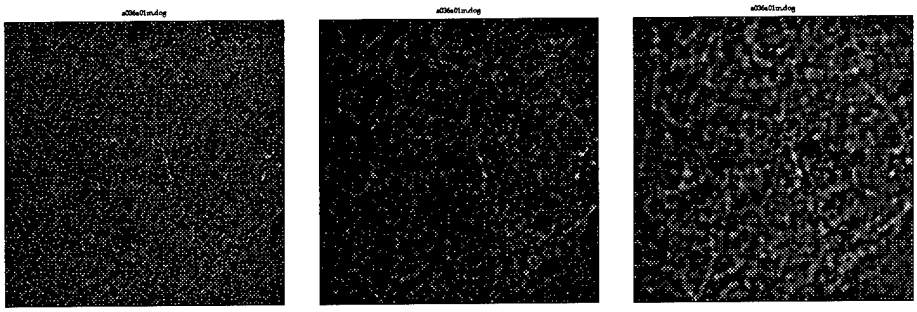


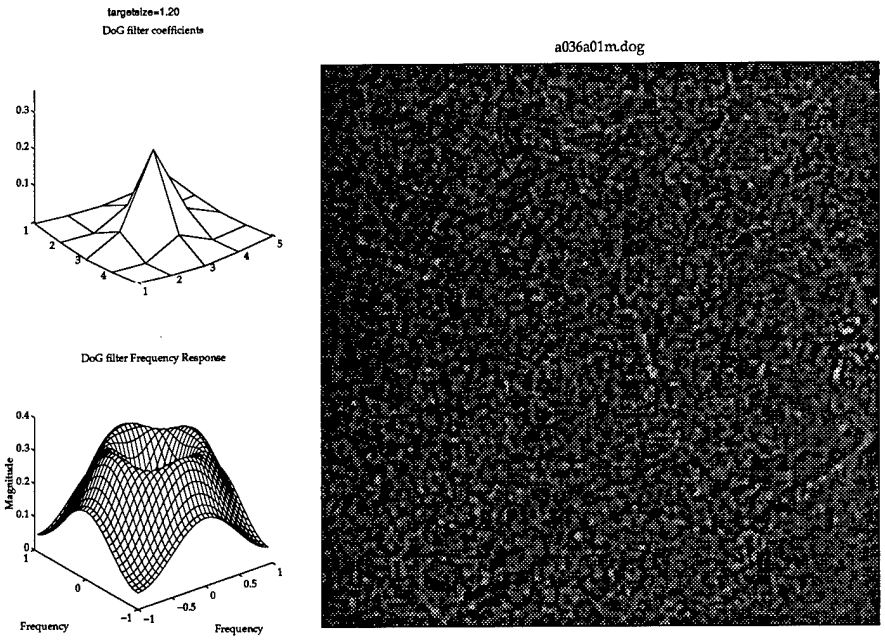
Figure 5. DoG kernels with their corresponding frequency response; (a) $t = 1.0$, (b) $t = 2.0$, and (c) $t = 3.0$. Units of t are pixels, where 1 pixel = 100 μm.

3.6.3 DoG-filtered examples. Since microcalcifications typically range from 100 to 300 μm in diameter (8, 56), potential target sizes for the 100 μm digitized mammograms correspond to 1 to 3 pixels. Using $t = \{1.0, 2.0, 3.0\}$, we find $\sigma_2 = \{0.5, 1.0, 1.5\}$ and $\sigma_1 = \{0.3125, 0.6250, 0.9375\}$, respectively. DoG filters were constructed using these paired parameters as shown in Figure 5. These kernels were individually used as spatial filters on the noise-reduced subimage shown in Figure 4b with corresponding results shown in Figure 6. Alternatively, an "optimized" DoG kernel is shown in Figure 7a with corre-



(a) (b) (c)

Figure 6. Results of applying three DoG kernels to a noise-reduced subimage; (a) $t = 1.0$, (b) $t = 2.0$, and (c) $t = 3.0$. Units of t are pixels, where 1 pixel = 100 μm. As targetsize increases, larger image structures are emphasized as expected.



(a) (b)

Figure 7. Results of DoG filtering noise reduced subimage using "optimized" targetsize; (a) DoG kernel, $t = 1.1995$ pixels, and (b) DoG filtered subimage

sponding filtered subimage results shown in Figure 7b. (Optimization of t is discussed in Section 3.10.)

3.7 Calcification detection module

Once an image has been preprocessed to enhance differences between targets and background, the image is more easily segmented. The DoG-filtered subimages (Figure 6 and Figure 7b) contain obvious differences in gray levels between potential microcalcifications and background; however, the optimum method for segmentation is not so clear. Although microcalcifications tend to be among the brightest objects in DoG-filtered subimages, they may exist within regions of high average gray levels and thus prove difficult to reliably segment. A method that generally addresses these concerns uses pair-wise pixel "anding" of the results of global histogram and locally adaptive thresholding.

If targets tend to exist within an image's higher gray levels, then the global threshold may be approximated by finding the level which segments a preselected percentage of the corresponding higher levels in the image histogram. Locally adaptive thresholding may be implemented by varying the high and low thresholds based on the local pixel value mean and standard deviation (43).

According to Gonzales, "the histogram of a digital image with gray levels in the range [0, L-1] is given as

$$p(r_k) = \frac{n_k}{n} \tag{5}$$

where r_k is the kth gray level, n_k is the number of pixels in the image with that gray level, n is the total number of pixels in the image, and $k = 0, 1, 2, ..., L-1$" (24). After the histogram has been computed, the gray level, g, used to segment (threshold) a preselected upper fraction of the histogram, f, is found using

$$f = 1 - \sum_{k=0}^{g} p(r_k) \tag{6}$$

where $0 \leq g \leq L-1$.

The locally adaptive thresholds, t_{lo} and t_{hi}, are found using

$$t_{lo} = k_{lo}\sigma(x,y) + m(x,y) \tag{7}$$

and

$$t_{hi} = k_{hi}\sigma(x,y) + m(x,y) \tag{8}$$

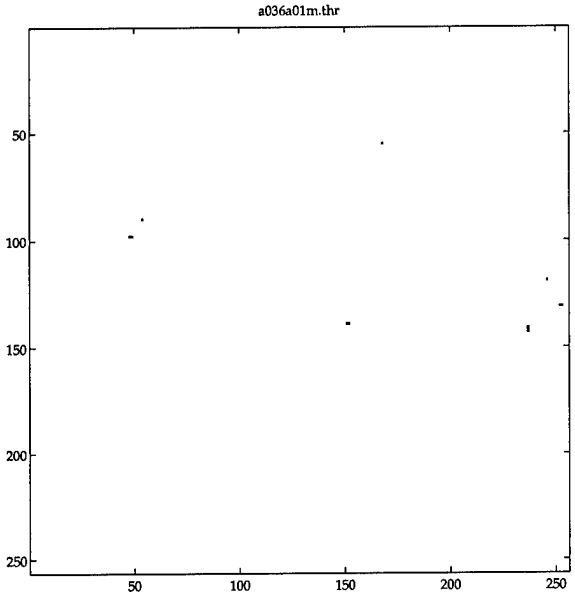
where k_{lo} and k_{hi} are used to preselect the multiple of $\sigma(x,y)$, the local standard deviation, and $m(x,y)$ is the local mean of the $N \times N$ neighborhood centered on the pixel at (x,y) (43). (Optimization of f, k_{lo}, k_{hi}, and N is discussed in Section 3.10.)

Once the globally and locally thresholded images are generated, they are combined into an initial calcification detection mask, M_i, using

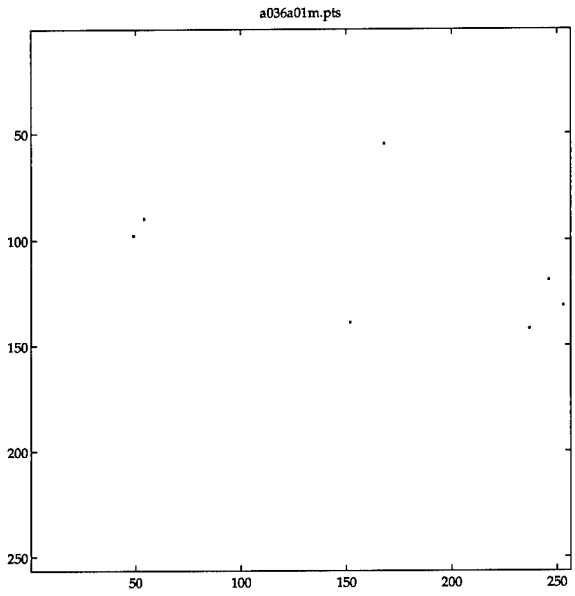
$$M_i(x,y) = T_g(x,y) \cap T_{lo}(x,y) \cap T_{hi}(x,y) \tag{9}$$

where $T_g(x,y)$ is the globally thresholded image, and $T_{lo}(x,y)$ and $T_{hi}(x,y)$ are the images locally thresholded using k_{lo} and k_{hi} respectively. For example, this method was performed on the DoG filtered subimage in Figure 7b with results shown in Figure 8a. To facilitate clustering, the centroids of individual detections are used to generate the final calcification detection mask, M_f, shown in Figure 8b.

When working with full-breast images, however, detections may occur on film labels which are obvoiusly not ROIs. Thus, to ensure detections are within the breast, a breast segmentation utility was developed to segment the breast from the image. This utility consists of five steps. First, a variance image is computed from a subsampled version of the full breast image. Second, the original image and the variance image are rescaled and added. Next, Floyd-Steinberg image dithering is used to reduce the grayscale image to two gray levels. Only the largest contiguous region is set to one assuming it always corresponds to the breast. All other regions are set to zero. Finally, the resulting binary image is upsampled to generate a full-image breast mask. This utility returns a binary breast mask with ones corresponding to the region of film containing the breast.



(a)



(b)

Figure 8. Initial and Final detection masks; (a) M_i: $f = 0.0009$, $k_{lo} = 4.1226$, $k_{hi} = 7.000$, and $N = 69$, and (b) M_f: Centroids of individual detections in M_i

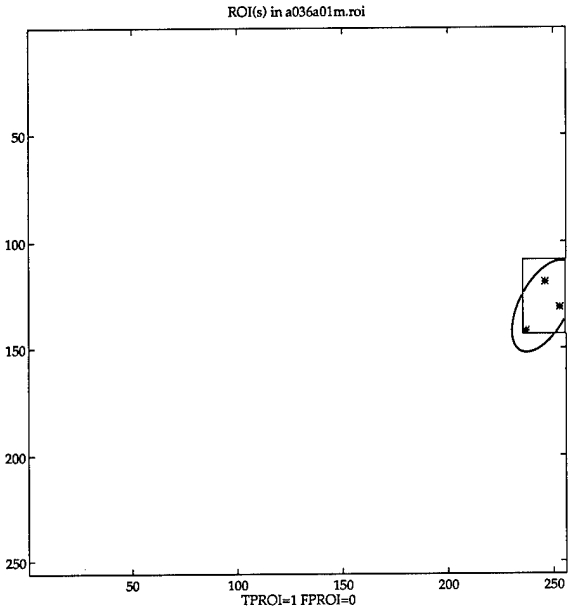
Thus, to ensure detections are within the breast, the centroids of individual locations are compared with the breast mask. If the centroids are found within the breast mask, the centroids are then passed to the cluster detection module.

3.8 Cluster-detection module

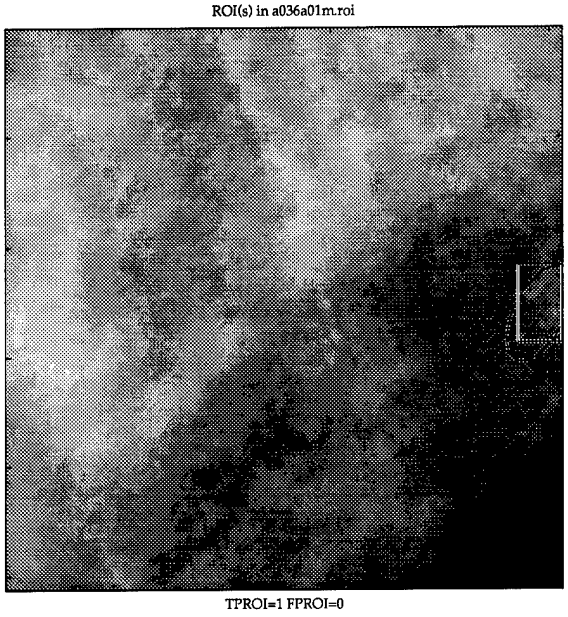
Recall, "...malignant calcifications tend to occur in tight clusters of 1 cm diameter or less " (56); however, calcifications can be extensive (56). Thus, the cluster detection module identifies clusters based on the clustering hypothesis proposed by Chan, *et al.* (9). Specifically, a suspicious ROI (cluster) is μCs_{min} or more detected signals linked by an Euclidean nearest neighbor distance, d_{nn}. (Optimization of μCs_{min} and d_{nn} is discussed in Section 3.10.)

3.9 Labeling computer-detected ROIs

Once detections have been indexed according to their corresponding ROIs, an ellipse may be used to identify computer-detected ROIs, as shown in Figure 9a. The shape of the ellipse may be generated based on the distribution of cluster signals. Figure 9b provides an example of the final output with corresponding truth for the subimage shown in Figure 4a. Figure 10 shows the corresponding full image result which may be provided in hardcopy form to the radiologist. Figure 11 illustrates overall detection methodology.



(a)



(b)

Figure 9. Cluster detection results for $t = 1.1995, f = 0.0009, k_{lo} = 4.1226, k_{hi} = 7.000, N = 69, \mu Cs_{min} = 3$, and $d_{nn} = 51.0031$; (a) computer-detected ROI with corresponding microcalcification detections, and (b) computer-detected ROI for radiologist consumption. Since the "ground truth" for the cluster was available, the corresponding bounding box is shown and detection results are reported.

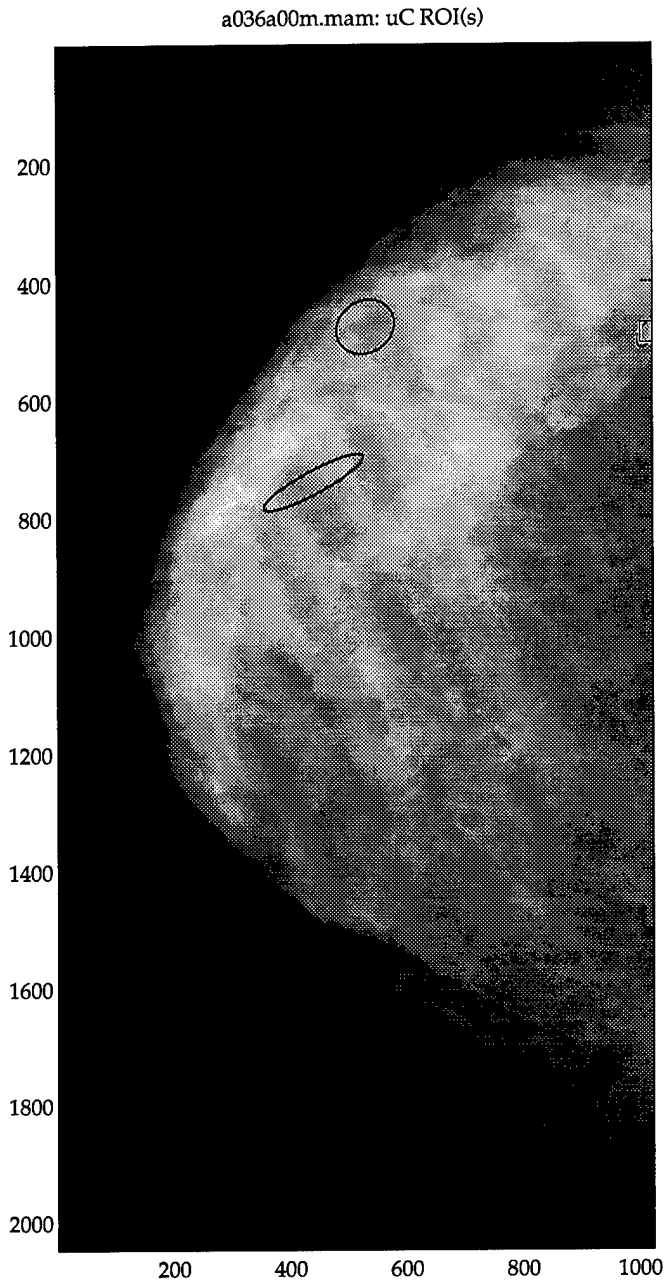


Figure 10. Full image (20×10 cm^2) cluster detection results for $t = 1.1995, f = 0.0009, k_{lo} = 4.1226, k_{hi} = 7.000, N = 69, \mu Cs_{min} = 3$, and $d_{nn} = 51.0031$; Since the "ground truth" for the cluster was available, the corresponding bounding box is shown. The upper ROI not corresponding to the truth box contained a true cluster of microcalcifications, however the region was not biopsied. The lowest ROI was due to scratches on the film.

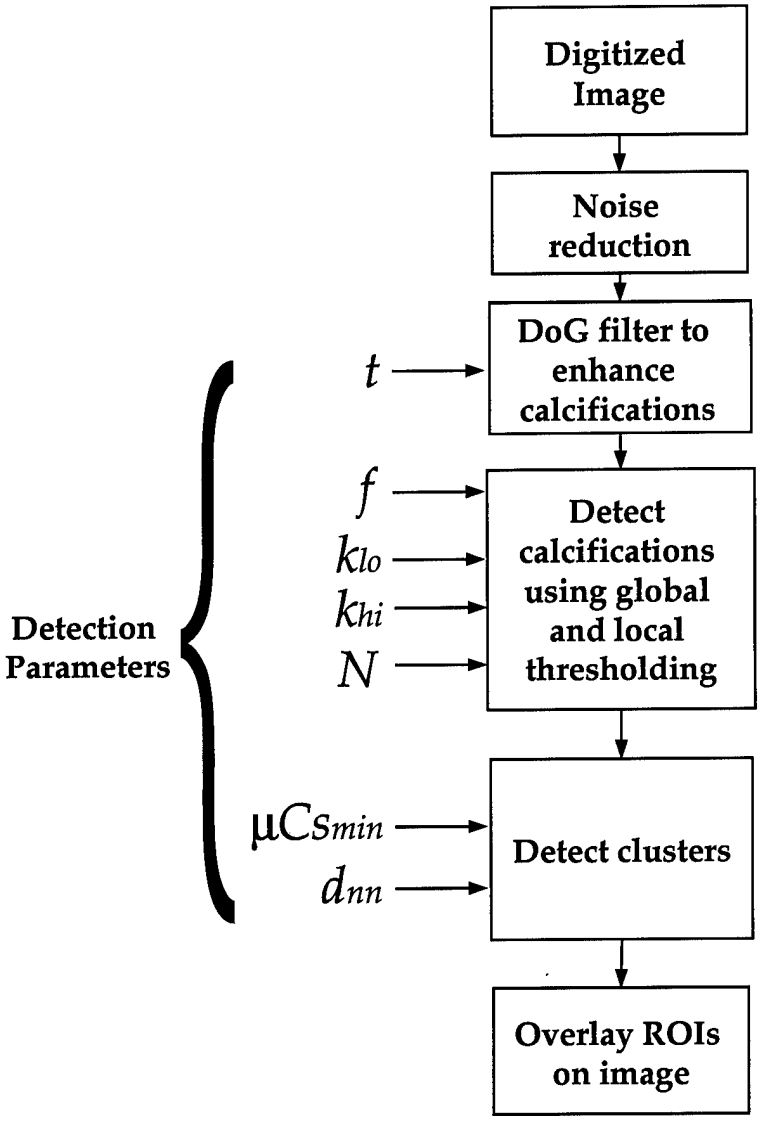


Figure 11. Cluster Detection Methodology: t controls the DoG filter used to enhance calcifications; f, k_{lo}, k_{hi}, and N control the detection stage of the system; μCs_{min} and d_{nn} control the clustering module.

3.10 Cluster detection optimization

As shown in Figure 11, the following seven parameters are used to control clustered microcalcification detection:

- t, microcalcification target size (average width) in pixels (1 pixel = 100 μm); i.e., $t = \{1.0, 1.15, 2.13, \ldots\}$, where $t > 0$
- f, upper fraction in intensity of the DoG-filtered image histogram; i.e., $f = \{0.001, 0.02, 0.05, \ldots\}$, where $0 < f < 1$
- k_{lo} and k_{hi}, multiples of $\sigma(x, y)$, the standard deviation in brightness computed from neighborhood centered on DoG filtered pixel at (x,y); i.e., $k_{lo} = \{3.6, 3.75, 4.17, \ldots\}$, and $k_{hi} = \{6.3, 7.0, 8.25, \ldots\}$ where $0 < k_{lo} < k_{hi}$.
- N, block size of neighborhood centered on DoG-filtered pixel at (x,y); i.e., $N = \{51, 53, 67, \ldots\}$, where $N > 0$, and $N \in Odd$
- μCs_{min}, minimum number of detected microcalcifications linked by d_{nn} to trigger a ROI detection; i.e., $\mu Cs_{min} = \{2, 3, 4, 5, \ldots\}$, where $\mu Cs_{min} \geq 2$
- d_{nn}, Euclidean nearest neighbor distance (in pixels) between detected microcalcifications; i.e., $d_{nn} = \{50.23, 51.5, 60.4, \ldots\}$, where $d_{nn} > 0$

Recall, t controls the DoG filter used to enhance calcifications. Additionally, f, k_{lo}, k_{hi}, and N control the detection stage of the system. Finally, d_{nn} and μCs_{min} control the clustering module. Although empirical results may be obtained by manually modifying these parameters, an alternative, more efficient method is to use an optimization algorithm.

3.10.1 Function optimization. A function optimization problem may be described as follows: given some finite domain, D, a particular setting of cluster detection parameters, $\mathbf{x} = \{t, f, k_{lo}, k_{hi}, N, \mu Cs_{min}, d_{nn}\}$ where $\mathbf{x} \in D$, and an objective function $f_{obj} : D \rightarrow \Re$, find the $\mathbf{x}$ in D that maximizes or minimizes f_{obj} (61). Optimization problems require an objective function which reflects the "goodness" of a solution (61).

$$f_{obj}(\mathbf{x}) = \begin{cases} -FP(\mathbf{x}) & , TP(\mathbf{x}) \geq TP_{min} \\ FP_{penalty} & , \text{otherwise} \end{cases} \tag{10}$$

Recall, radiologic imaging systems may be optimized to maximize the TP rate while minimizing the FP rate. This objective may be recast into the functional form shown in Equation 10, where *maximization* is the goal. (The genetic algorithm technique used in this research seeks maximization of an objective/fitness function.) For a particular setting of cluster detection parameters, if the minimum acceptable TP rate, TP_{min}, is exceeded, the objective function returns the negative of the FP rate. Otherwise, if the TP rate falls below TP_{min}, the objective function returns a constant value, $FP_{penalty}$; i.e., $FP_{penalty} = -10$ because $-FP(\mathbf{x})$ values observed in this research typically range from approximately $-.27$ to -1.0. Although Richardson *et al.* (50) argue "...infeasible solutions should provide information and not just be thrown away," they acknowledge "...many researchers believe penalty functions should be harsh" so optimization methods "...will avoid the forbidden spaces." Furthermore, since Richardson *et al.* (50) support their argument with only two contrived optimization problems, a harsh $FP_{penalty}$ is used.

Conventional approaches to optimization use calculus-based methods (61). However, these approaches are generally limited to objective functions that possess "nice" properties such as continuity, differentiability, etc. (29) Their main limitation is their local behavior; the search can easily end in a local maximum, and the global maximum can be missed (61). Thus, several techniques are used to improve the probability of finding the global maximum: start at several points in the search space, use dynamic programming, apply random searches, etc. (61). One relatively new and widely applicable approach is to use genetic algorithms augmented by one of these standard methods (28, 61).

3.10.2 Genetic algorithms. Genetic algorithms (GAs) have been successfully applied to many diverse and difficult optimization problems (29) such as jet engine turbine design (28), tracking of criminal suspects through face composites (6), and binary phase-only filter design for pattern recognition (7). GAs "...are at their best when exploring complex landscapes [solution spaces] to find regions of enhanced opportunity" (28). However, as Holland (inventor of the GA) (28) notes, "...if a partial solution can be improved further by making small changes in a few variables, it is best to augment the genetic algorithm with other, more standard methods." Thus, this research uses a Matlab imple-

mentation of a GA developed by Houck *et al.* (29) to find promising parameter settings, then augments the GA with the simplex method (62) included with Matlab.

GAs search the solution space to maximize a fitness (objective) function using simulated evolutionary operators; i.e., mutation, sexual recombination, etc. (28, 29, 61). GAs manipulate and maintain a population of solutions while searching for *better* solutions (29). The fittest individuals of any population tend to reproduce more and survive longer. This increases the average fitness in successive generations (29, 61). However, as Houck *et al.* (29) note, inferior individuals can survive and also reproduce. A more complete discussion of GAs and related topics can be found in Michalewicz's book (41). Houck (29) summarizes the GA as follows:

1. Supply [an initial] population P_0 of [n] individuals and respective [fitness] function values
2. $i \leftarrow 1$
3. $P_i' \leftarrow selection\_function(P_i - 1)$
4. $P_i \leftarrow reproduction\_function(P_i')$
5. $i \leftarrow i + 1$
6. Repeat step 3 until termination [termination criteria defined later]
7. Print out best solution found

GA use requires determination of several issues: objective function design, parameter set representation, population initialization, choice of selection function, choice of genetic operators (reproduction mechanisms) for simulated evolution, and identification of termination criteria (29, 61). Although one may argue the GA itself appears to require optimization of GA parameter settings, several researchers have closely examined this "problem" (4, 11, 22, 23, 30, 50, 54). As there is no standard agreement on optimum GA parameter settings, this research adopts the following general "rules of thumb" culled from several investigations in the literature:

- As Sonka *et al.* (61) note, "No optimization problem can guarantee finding a good solution to the problem if the objective function does not reflect the goodness of the solution. Therefore, the design of the objective function is a key factor in the performance of any optimization algorithm." Recall Equation 10: if an acceptable TP rate is reached, then minimization of the FP rate is desired.
- For parameter representation, Janikow and Michalewicz (30, 41) showed the real-valued GA is an order of magnitude more efficient in CPU time than the binary GA and provides higher precision with more consistent results across replications (29). Thus, a floating-point representation is used in this research.
- Because GAs can iteratively improve existing solutions (29), several GA users seed the initial population with some members known beforehand to be in an interesting part of the search space (4, 22). This technique is used because preliminary cluster detection parameter settings appear to be in an interesting region of the search space. Also, while the minimum number of members in a GA population is three (22), Schaffer *et al.* (54) argue good online performance can be expected with 20 to 30 members. Because Schaffer et al. support their suggestion using a large problem suite over 8400 genetic searches (10K evaluations/search) (54), and each objective function evaluation is computationally expensive (approximately 30 to 60 minutes on a Sun Ultra or Sun Sparc20) 20 members are used in this research.
- Several schemes exist for the (typically) probabilistic selection process used to identify candidates for reproduction: roulette wheel or proportional selection, scaling techniques, tournament, elitist models, and ranking methods (29, 41, 61). According to Grefenstette and Baker (25), although several selection algorithms have been studied, many "...have not enjoyed the same level of theoretical analysis as the proportional selection algorithm." The authors argue a promising alternative to proportional selection is ranking because "the resulting algorithm is less prone to premature convergence cause by individuals that are far above average." The basic idea of ranking is to select strings (solutions) for the mating pool based on the relative fitness between strings. This research uses the ranking method included in the Matlab GA, normalized geometric ranking (29).

- Two basic types of genetic operators are commonly used in GAs: crossover and mutation (28, 54, 61). According to Sonka *et al.* (61), "...the basic idea [of crossover] is to randomly mate the ... strings [solutions] " identified by the selection process by "...randomly choosing a position for the border of each pair of strings," and producing "...new strings by swapping all characters between the beginning of the string pairs and the border position." Sonka *et al.* (61) note the mutation operator's role is to "...randomly change one character of some string of the population from time to time." Mutation is used because "...some local configuration of characters in strings of the population can be totally lost" due to the selection and crossover operations (61). As there is no general agreement on specific crossover and mutation schemes, this research adopts the default crossover and mutation schemes included with the Matlab GA: arithmetic crossover and nonuniform mutation (29).
- Finally, the termination criteria used in this research adopts the commonly used technique of running each iteration until a specified number of generations are done (29, 54); here, ten generations are used due to the high computational cost mentioned earlier. Although termination due to loss of population diversity and/or lack of improvement is efficient when crossover is the primary source of variation in a population (4, 22, 29), homogeneous populations can be succeeded by populations with better (higher) fitness when using mutation (4). Additionally, since Goldberg (22) argues restarting populations that may have converged proves useful, this research runs several iterations of the GA until a consistent lack of improvement is recognized.

3.10.3 Simplex method. As mentioned earlier, once potentially optimum solutions are found by the GA, the most fit GA solution may be further optimized using the simplex method developed by Dantzig (28, 62). The Matlab implementation of the simplex algorithm requires a basic (initial) feasible solution, $\mathbf{x_0}$ (found by the GA), and a cost function (Equation 11) (5, 62). In this research, the simplex method is used to locate other feasible solutions (members of the "feasible set") while minimizing the cost function until the optimal solution is obtained (5). It should be noted, since the GA provides potentially optimum solutions not necessarily near or equal to the global optimum, the

simplex method is not guaranteed (nor expected) to locate the global optimum. Rather, the simplex method is used to further optimize the most fit solution found by the GA. A cursory overview of the simplex method follows.

$$f_{cost}(\mathbf{x}) = -f_{obj}(\mathbf{x}) = \begin{cases} FP(\mathbf{x}) & , TP(\mathbf{x}) \geq TP_{min} \\ -FP_{penalty} & , \text{otherwise} \end{cases} \tag{11}$$

First, several terms must be defined. According to Bronson (5), "A *mathematical program* is an optimization problem in which the objective and constraints are given as mathematical functions and functional relationships." If these relationships are linear in each of their arguments a mathematical program is *linear* (5). The simplex method was developed to solve linear programs (62). Although the cost function and its constraints (Equation 11) may not be strictly linear, the simplex method is effective at further optimizing the most fit solutions found by the GA.

The cost function (Equation 11) brings to the problem a family of parallel planes (62). The solution with zero cost is the plane that goes through the origin; other planes give all other possible costs (62). These planes sweep out the entire n-dimensional solution space, and the optimal solution occurs where the planes first touch the feasible set (62).

Thus, as Strang (62) notes, there are basically the two phases in the simplex method:

1. Locate a corner of the feasible set, where a *corner* is the meeting point of n different planes.
2. Within the feasible set, go from corner to corner along an edge guaranteed to decrease the cost. Terminate when a corner is reached from which all edges increase the cost.

Finally, while the total number of simplex steps needed to find the optimum solution is impossible to answer, Strang (62) reports the simplex method has been shown to operate in an average time which is polynomial (as opposed to the undesirable class of non-polynomial, or NP). Because each cost/objective function evaluation is computationally expensive (approximately 30 to 45 minutes on Sun Ultra or Sun Sparc20), this research uses 300 simplex steps. A more complete discussion of the simplex method may be found in Strang (62).

3.11 Summary

This chapter presented a new technique to detect clustered microcalcifications in digitized film mammograms. First, the image is preprocessed to reduce digitizaion noise. Next, a DoG filter is used to enhance microcalcifications. The filtered image is thresholded, and detections are passed to the cluster detection module. Finally, computer-detected ROIs are overlaid on the original image for radiological analysis. To improve detection rates, globally optimize detection parameters using a genetic algorithm (GA), then locally optimize using the simplex method.

IV. Results

This chapter presents the results of the clustered microcalcification detection technique.

4.1 Optimization

As mentioned earlier, full-breast image optimization of detection parameters would have been computationally expensive. Thus, subimages were used during the optimization stage. All subimages had at least one biopsy-truthed ROI. If at least one subimage ROI was detected, a TP of 1 was recorded. All FPs within each subimage were recorded as well. The goal of the optimization stage was to reduce the subimage FP rate while maintaining the minimum acceptable TP rate, $TP_{min} = 84.5$ percent.

4.1.1 TP_{min}. Since the optimization method described in Chapter III requires a minimum acceptable TP rate to minimize the FP rate, an optimized ROC analysis was performed using $TP_{min} = 84.5, 82.5, 80.0$, and 77.0 percent. Although higher subimage TP_{min} goals were easily attainable, they were accompanied by higher subimage FP rates as expected; approximately 0.8 to 2.6 FPs per subimage.

Recall, image area is an important factor in increasing the FP rate since the larger the test area, "...the greater the chance for producing false positives in general" (18). Thus, because most of the subimages were approximately 1/32 the area of a database image, subimage detection rates of 0.8 to 1.3 may lead to approximately 25.6 to 41.6 FPs per full-breast image. Continuing with the analysis, if a maximum of approximately 5 FPs per full-breast image is desired, approximately 0.1562 FPs per subimage may be considered the goal.

4.2 Subimage ROC results

To provide insight from a ROC perspective into how subimage TP and FP detection rates vary with the choice of TP_{min}, the GA and simplex methods were used to search the feasible set of parameter settings. Additionally, the case TP rate was tracked to

Table 2. TP_{min} vs. Optimized subimage parameter settings

$TPmin$	t	f	k_{lo}	k_{hi}	N	d_{nn}	μCs_{min}
84.5	1.1995	0.0009	4.1226	7.0000	69	51.3019	3
82.5	1.1995	0.0009	4.1226	7.0000	69	51.3420	3
80.0	1.2286	0.0009	4.2289	7.0000	69	51.3019	3
77.0	1.2276	0.0007	4.2306	7.0002	69	51.3031	3

Table 3. Optimized subimage ROC results

$TPmin$	Case TP	Subimage TP	Subimage FP
84.5	96.43	85.25	0.2295
82.5	96.43	85.25	0.2295
80.0	92.86	81.97	0.1967
77.0	92.86	81.97	0.1803

demonstrate optimization effects. Optimized subimage parameter settings found during this search are summarized in Table 2, with corresponding TP and FP rates in Table 3. Recall, t controls the DoG filter used to enhance calcifications. Additionally, f, k_{lo}, k_{hi}, and N control the detection stage of the system. Finally, d_{nn} and μCs_{min} control the clustering module.

As TP_{min} was lowered, the subimage FP rates fell as expected, however the corresponding drop in subimage TP rates was accompanied by a drop in the case TP rate from 96.43 to 92.86 percent. Also, as mentioned earlier, although higher subimage TP_{min} goals were attainable, they were accompanied by higher subimage FP rates as expected. Thus, optimized parameter settings found using $TP_{min} = 84.5$ percent were chosen for full digitized mammogram image evaluation. These settings corresponded to a TP case detection rate of 96.43 percent and a subimage detection rate of 85.25 percent with 0.2295 FPs per subimage.

4.3 Database results

4.3.1 Physician Evaluation of Detection Scheme. Ultimately, full image results were evaluated as shown in Table 4 using the following optimized parameter settings: $t = 1.1995, f = 0.0009, k_{lo} = 4.1226, k_{hi} = 7.000, N = 69$, and $\mu Cs_{min} = 3$. To ensure full

image detection results were accurate, a physician trained in radiology (27) individually verified each ROI detection on the original film as a TP or FP. (Recall, the physician only truthed subimages, not full images.) Again, if at least one image ROI was detected, a TP of 1 was recorded. Also, since the detection algorithm was evaluated on full-breast images, several additional non-biopsied microcalcification ROIs were detected. Thus, an additional category of "Xtra ROIs" was used to track additional physician verified ROIs. Recall, Figure 10 shows an example of an "Xtra ROI." All FPs within each subimage were recorded as well.

To provide insight into the sources of FPs, the physician (27) used the following two additional categories to describe noteworthy FPs: Scratch and BB induced ROIs. As mentioned earlier, scratches and pick-off of film emulsion are a known source of pseudocalcifications in film mammograms (20). The term "scratch induced ROI" is used for ROIs due to scratches and/or film emulsion pick-off. Recall, Figure 10 shows an example of a scratch induced ROI. BBs are typically used as a marker on mammograms to identify palpable lesions in the breast. It should be noted that no TPs were due to BBs. Also, as shown in Table 4, only 9 of 56 images had double-digit FP rates. The high FP rate on these images may be attributed to parenchymal edge effects; the functioning structure of these breasts (ducts, lobules, etc.) introduced a large number of thin, sharp densities which appeared calcification-like to the detection scheme.

4.3.2 Analysis: Detection Scheme Results. The detection technique demonstrated the following results (Table 4) over 56 full-breast digitized film mammograms:

- 96.4% (27/28) TP case detection rate
- 85.7% (54/63) TP ROI detection rate
- 85.7% (48/56) TP image detection rate
- 5.75 FPs per full-breast image (20×10 cm, 12 bit, 100 μm)
- 0.68 scratch induced FPs per image
- 3 BB induced false ROIs
- 2.1964 "Xtra ROIs" per image

Table 4. Results using $t = 1.1995, f = 0.0009, k_{lo} = 4.1226, k_{hi} = 7.000, N = 69$, and $\mu Cs_{min} = 3$: a TP_{case} rate of 96.43%, and TP_{ROI} and TP_{image} rates of 85.7% with 5.75 FPs per full-breast image. Note: Only 9 images had double digit FPs mostly due to parenchymal edge effects.

image	TP	FP	Xtra ROIs	Scratch ROIs	BB ROIs	ROIs detected	ROIs in image
a002a00m.mam	1	2	6	2	0	1	1
a002b00m.mam	1	1	8	0	0	2	2
a002c00m.mam	1	2	4	0	0	1	1
a002d00m.mam	1	6	6	0	0	1	1
a003a00m.mam	0	7	2	1	0	0	1
a003c00m.mam	1	8	0	3	0	1	1
a004b00m.mam	1	3	3	0	1	1	1
a004d00m.mam	1	6	0	2	0	1	1
a005b00m.mam	1	7	0	0	0	1	1
a005d00m.mam	1	7	0	0	0	1	1
a006b00b.mam	1	10	1	0	0	1	1
a006d00b.mam	1	14	0	0	0	1	1
a009a00m.mam	1	10	1	0	0	1	1
a009c00m.mam	0	12	2	0	0	1	1
a010b00b.mam	1	4	0	0	0	1	1
a010d00b.mam	1	2	1	0	0	1	1
a011b00m.mam	1	1	2	0	0	1	1
a011d00m.mam	0	3	2	1	0	0	1
a014b00m.mam	1	1	8	0	0	1	1
a014d00m.mam	1	8	1	0	0	1	1
a015a00m.mam	1	1	3	0	0	1	1
a015c00m.mam	0	8	2	2	0	0	1
a016d00b.mam	1	13	7	1	0	1	1
a024b00b.mam	1	5	5	0	0	2	2
a024d00b.mam	1	8	1	1	0	2	2
a026d00m.mam	1	16	1	1	0	1	1
a029b00b.mam	1	3	0	0	0	1	1
a029d00b.mam	1	2	1	0	0	1	1
a031a00b.mam	1	3	2	0	0	1	1
a031c00b.mam	1	1	3	0	0	1	1
a033b00m.mam	1	5	5	1	1	1	1
a033d00m.mam	1	7	3	0	0	1	1
a036a00m.mam	1	1	1	1	0	1	1
a036c00m.mam	1	5	0	1	0	1	1
a043b00b.mam	1	9	2	2	0	1	1
a043d00b.mam	1	16	3	1	0	1	1
a049b00m.mam	1	12	0	2	1	0	1
a049d00m.mam	1	6	0	0	0	1	1
a052b00b.mam	1	1	0	1	0	1	1
a052d00b.mam	1	4	0	1	0	1	1
a053b00m.mam	0	12	2	0	0	0	1
a053d00m.mam	0	6	0	2	0	0	1
a058b00m.mam	1	5	2	0	0	1	1
a058d00m.mam	1	2	1	0	0	1	1
a060b00m.mam	1	2	10	0	0	2	2
a060d00m.mam	1	3	14	0	0	2	2
a062a00m.mam	0	4	0	0	0	0	1
a062c00m.mam	1	3	1	2	0	1	1
a067a00b.mam	1	1	1	0	0	1	1
a067c00b.mam	1	6	2	2	0	1	1
a068b00b.mam	1	8	2	3	0	1	1
a068d00b.mam	1	7	0	1	0	1	1
a069a00m.mam	1	6	1	3	0	1	1
a069c00m.mam	1	8	1	0	0	1	1
a073b00b.mam	0	3	0	0	0	0	2
a073d00b.mam	1	6	0	1	0	2	2

Table 5. Clustered microcalcification ROI detection results; NR indicates not reported.

Investigators	Number of images	Image Area (cm^2)	Resolution (μm/pixel)	TP (%)	FPs (#/image)	Area adjusted FPs (#/image)
Chan *et al.* (overall)	52	NR	100	92	1	NR
Yoshida *et al.*	39	8×10	100	85	5	12.5
Nishikawa *et al.*	39	8×10	100	85	2.5	6.25
Strickland *et al.*	40	10×10	50	84	3	6
Ochoa *et al.*	56	20×10	100	85.7	5.75	5.75

As mentioned earlier, a direct comparison of detection results against other CADx detection schemes (9, 44, 46, 64, 68) cannot be done since CADx researchers use different databases. However, as shown in Table 5, the 85.7 percent ROI TP detection rate found using this detection scheme is comparable with reported ROI TP detection rates of 84 to 93 percent (9, 44, 46, 64, 68). Although the performance of a CAD scheme is difficult to reliably estimate (45), the number of images used to determine TP and FP rates was larger than those used by Chan *et al.* (9), Yoshida *et al.* (68), Nishikawa *et al.* (44, 46), and Strickland *et al.* (64) as shown in Table 5.

A direct comparison of FP rates is made even more difficult due to the lack of standardized images areas, and image area has shown to be an important factor in FP detection rates (18). As shown in Table 5, several researchers use images half the area of those used in this research. Additionally, as shown in Table 3, if subimages are used, this research could claim FP rates as low as 0.2295 if image area is ignored. Nevertheless, the detection scheme's FP rate of 5.75 per full-breast image is comparable with reported rates of 1 to 5 FPs per (sub)image (9, 44, 46, 64, 68) as shown in Table 5. (Because Strickland *et al.* used 50 μm/pixel images, this corresponds to a 10×10 cm^2 region on a film mammogram, and the region of film covered in this research corresponds to a 20×10 cm^2 area.) Moreover, if image area is used to adjust FP rates for comparison purposes, then the Ochoa *et al.* detection scheme's FP rate is the lowest of the results shown in Table 5.

Additionally, although other investigators do not provide the following level of detail when reporting detection results, the following malignant and benign detection rates were identified:

- Malignant detection rates
  - 94.1% (16/17) TP case detection rate
  - 81.6% (31/38) TP ROI detection rate
  - 80.0% (28/35) TP full-breast image detection rate
  - 5.6 FPs per full-breast image (20×10 cm^2, 12 bit, 100 μm)
- Benign detection rates
  - 100% (11/11) TP case detection rate
  - 92.0% (23/25) TP ROI detection rate
  - 95.2% (20/21) TP full-breast image detection rate
  - 6.0 FPs per full-breast image (20×10 cm^2, 12 bit, 100 μm)

Because other investigators do not report this level of detail when describing their detection schemes, a direct comparison of these detection rates is not an option. However, the fairly large number of images used to estimate malignant or benign TP and FP rates is comparable with the size of the entire database used by Yoshida *et al.* (68), Nishikawa *et al.* (44, 46), and Strickland *et al.* (64) as shown in Table 5.

4.4 Summary

The database of 56 digitized (12 bit, 100 μm) full-breast (20×10 cm^2) film mammograms contained 63 biopsy-truthed ROIs over 28 cases. This technique demonstrated a TP case detection rate of 96.43 percent (27/28) and TP ROI (54/63) and TP image (48/56) detection rates of 85.7 percent with 5.75 FPs per full-breast image.

V. Conclusion

5.1 Summary

This research proposes the following methodology for clustered microcalcification detection. First, preprocess the digitized film mammogram to reduce digitization noise. Second, spatially filter the image with a difference of Gaussians (DoG) kernel. To detect potential microcalcifications, segment the filtered image using global and local thresholding. Next, cluster and index these detections into regions of interest (ROIs). Identify ROIs on the digitized image (or hardcopy printout) for final radiologic diagnosis. Finally, to improve detection rates, globally optimize detection parameters using a genetic algorithm (GA), then locally optimize using the simplex method.

As mentioned earlier, a direct comparison of detection results against other CADx detection schemes (9, 44, 46, 64, 68) cannot be done since CADx researchers use different databases. However, as shown in Table 5, the 85.7 percent ROI TP detection rate found using this detection scheme is comparable with reported ROI TP detection rates of 84 to 93 percent (9, 44, 46, 64, 68). Although the performance of a CAD scheme is difficult to reliably estimate (45), the number of images used to determine TP and FP rates was larger than those used by Chan *et al.* (9), Yoshida *et al.* (68), Nishikawa *et al.* (44, 46), and Strickland *et al.* (64) as shown in Table 5.

A direct comparison of FP rates is made even more difficult due to the lack of standardized images areas, and image area has shown to be an important factor in FP detection rates (18). As shown in Table 5, several researchers use images half the area of those used in this research. Additionally, as shown in Table 3, if subimages are used, this research could claim FP rates as low as 0.2295 if image area is ignored. Nevertheless, the detection scheme's FP rate of 5.75 per full-breast image is comparable with reported rates of 1 to 5 FPs per (sub)image (9, 44, 46, 64, 68) as shown in Table 5. Moreover, if image area is used to adjust FP rates for comparison purposes, then the Ochoa *et al.* detection scheme's FP rate is the lowest of the results shown in Table 5.

5.2 *Contributions*

The following list summarizes significant contributions made during the course of this research:

- A complete, simple to use, end-to-end microcalcification detection system was developed which demonstrated a 96.4% case TP rate, and an 85.7% TP rate with a 5.75 FPs rate over 56 biopsy-truthed full-breast images (20×10 cm^2, 12 bit, 100 μm).
- This system is the first reported clustered microcalcification detection technique developed using optimization methods rather than mostly *ad-hoc* modification of detection parameters.
- If the average targetsize is known, the DoG methodology presented here may easily be used in other target recognition schemes because the algorithm is independent of image area and has been shown to be insensitive to large variations in image quality; i.e., subtle targets in noisy backgrounds.
- The relatively straightforward optimization approach presented here may also be used in other target recognition research to efficiently determine optimimum detection parameter settings.
- Three of AFIT's most recent CADx schemes (10, 38, 48) were integrated into the first AFIT CADx system for batch-detecting radiographic ROIs in digitized film mammograms.
- During the course of this work, an image histogram equalization utility was developed and included in Abrahamson's (1) Pulse-Coupled Neural Network (PCNN) magnetic resonance imaging (MRI) research. The equalization step was included in a preprocessing stage to allow a PCNN to accurately segment brain MRIs.

5.3 *Conclusions and Recommendations*

The objective of this thesis was to design an automated microcalcification detection system which may be used by radiologists as an aid in mammogram interpretation. This objective was met as indicicated by a TP case detection rate of 96.43 percent (27/28) and

TP ROI (54/63) and TP image (48/56) detection rates of 85.7 percent with 5.75 FPs per full-breast image (20×10 cm^2, 12 bit, 100 μm).

To evaluate the clinical impact of this system, a clinical study should be performed on this research, as well as the other CADx techniques (10, 38, 48) included in the foremen-tioned batch-capable system. Also, to reduce the FP rate, future AFIT CADx researchers should investigate the utility of fusing the results of this research with other clustered microcalcification detection methods. Alternatively, augmenting this system with a neu-ral network using features derived from detected ROIs may also help reduce the FP rate. Finally, the optimized DoG filtering approach developed during the course of this research has strong potential for detecting small targets in other cluttered imagery, such as detecting SCUD missile launchers in "noisy" long-range infrared images.

Appendix A. Database Information

A.1 Overview

In screen-film mammography, the film is used for detection, storage, and display. (15). Since full-area detectors suitable for digital mammography are still in the experimental stage, an alternative way of acquiring mammograms in digital form is to digitize a film mammogram (67). This research was performed using digitized film mammograms and subimages annotated by a physician trained in radiology.

Specifically, the database of 56 digitized (12 bit, 100 μm) full-breast (20×10 cm^2) film mammograms contained 63 biopsy-truthed clustered microcalcification ROIs over 28 cases. All mammograms included a pathology report indicating location and diagnosis of biopsied regions. The data set included approximately 2 films per case with the following distribution: 1 case with 4 films, 2 cases with only 1 film, and the remaining 25 cases with 2 films. Full-breast image optimization would have been extremely costly; approximately 30 to 60 minutes per digitized film on a Sun Ultra or Sun Sparc20. Thus, 52 256×256 and 9 512×512 subimages were extracted from these images to facilitate algorithmic development. The 61 subimages contained all 63 biopsy-truthed ROIs, and a physician annotated abnormalities identified in corresponding pathology reports. These subimages considerably reduced algorithmic development times, and reduced computer memory requirements. Additional information on digitization and database management is located in the Appendix.

The following sections were written by Capt. William E. Polakowski, Capt. Edward M. Ochoa, and Jill Leighner.

A.2 Digitization of film mammograms

Film Mammogram Digitization Instructions

I. Power up the Big Mac (Mac IIfx) in the X-ray filing room. (if necessary)
II. To calibrate the digitizer.
  A. Note that the digitizer only needs to be calibrated once per digitization session.
  B. From the large folder in the center of the screen with three icons on it, click on Film Digitizer.
  C. Note that to enter data into Mac text entry windows, you must hit the TAB key instead of return.
  D. Login as AFIT then press TAB.
  E. Type in the password then press RETURN.
  F. From the button bar on top of the screen, select "Special then Calibrate".
  G. Obtain any 6, 14" x 17" films from underneath the digitizer.
  H. Load the 6 films into the digitizer, which is located to the right of the Big Mac.
  I. To begin the calibration procedure, select "OK".
III. Obtain and use the 16 hour bar code. (Do this while calibrating.)
  A. Login to the small Mac (Mac IIsi) located to the left of the Big Mac. There should be a large folder in the center of the screen with two icons on it.
  B. Select Paris.
  C. Below the box stating Patients Name, type "AFIT, TEST". Then select Find Patient.
  D. From the top of the Paris window click on icon called "Exams for patient". (An exam is one digitization session.)
  E. Under the list of exams for AFIT, TEST select any exam. (It does not matter which exam is selected during one digitization session.)
  F. On the top of the screen, on the button bar, select "Exams then Duplicate".
    1. (If necessary), in the center right portion of the screen, in the Requesting Ward/Clinic box, type "MAMMOGRAPHY".
  G. At the bottom of the screen, click the "New Exam" button.
  H. Remove the barcode slip from the machine to the left of the small Mac.
  I. Select "File then Quit". Remember the barcode slip is good for 16 hours.
IV. To digitize the films.
  A. Scan the barcode and listen for the beep from the Big Mac.
  B. Prepare mammography films you intend to digitize.
    1. Make sure that the films are as clean as allowable. (If OK, wipe off any grease pencil or smudge marks, and if possible remove stickers - they tend to make the digitizer refuse to scan the film.)
    2. Orient the films as the radiologist views them.
    3. Order the films by stacking them (sticker side down) in the following bottom to top order : RCC, LCC, RMLO, LMLO (if all are available).
    4. Flip the stack (top toward you) so the RCC view is now on the top of the stack and rotate the top of the films away from you.
    5. Remember, the digitizer pulls the films from the top of the stack to the bottom, so the final order of digitization will be RCC, LCC, RMLO, LMLO.
  C. Place and align the films in the digitizer, put the long side of the films flush against the right side of the digitizer's auto feeder.
  D. To begin the digitization process, on the Big Mac, in the Film Digitizer window, click "Digitize" then click "OK".
  E. After the films have completed digitization, click "OK".
  F. To end digitization, go to the top of the screen, and from the button bar, select "File then Quit" from the pull down menu, then select "Quit" when the Login window reappears.
  G. Return/Store films that have been digitized.

A.3 Digitized film storage instructions

Digitized Mammogram Storage Instructions

I. To connect portable hard drive for digitized film downloading.
   A. Follow directions carefully.
   B. Locate the drive adapter cable for the portable Mac drive.
   C. To connect the portable Mac drive to the Big Mac.
      1. From the main window, select "Special then Shutdown". THE SYSTEM MUST BE SHUT DOWN BEFORE CONNECTING THE PORTABLE HARD DRIVE.
      2. To attach the drive.
         a. Remove the digitizer's black connector in back of the Big Mac.
         b. Attach the portable drive's power cord. Make sure the portable drive's power switch is OFF. (The switch in the back of the drive should be to the empty circle position.)
         c. Attach the portable drive to the Big Mac's socket the digitizer was connected to. Check to make sure that the keyboard was not disconnected during this process.
      3. To activate the drive.
         a. Note that the next few steps may need to be repeated until the Big Mac recognizes the portable drive.
         b. Turn on the drive (switch in the back of the drive). The light in front will turn off.
         c. Press the Big Mac startup key. If the Big Mac fails to turn on check that the keyboard was not disconnected while attaching the drive or push the reset switch in the rear, bottom right corner of the back of the Big Mac.
         d. As the Big Mac powers up, the drive will turn on, then off.
         e. From the large folder in the middle of the screen select "File then Go to Finder" and enter the password.
         f. To use each disk.
            1. Insert disk (Be sure to keep track of the ones you have already used.)
            2. Open the numbered icon (click on it).
            3. Clear each disk so it can hold the digitized mammograms. Select from the top menu "Special then Erase Disk then Initialize"

II. To get image information and download images to disks.
   A. From the large folder in the main window, double click on FWD060 and double click on Lite Box.
   B. Login as AFIT, then press TAB.
   C. Type the password, then press RETURN.
   D. To obtain the image name and size information.
      1. Go to the Patient Name box and type AFIT, TEST (no return key is required here, it will automatically find the AFIT list for you).
      2. Select your exam (the last exam at the bottom of the list). It should be an exam with an "I" in the ONLINE column. The "I" means that the process is still active.
      3. To display all of the exam images, on top of the "All exams" window , click on "Images then Display".
      4. Select OK when prompted.
      5. Click on the first image.
      6. To obtain specific image information.
         a. From the top button bar select "Image then Image About".
         b. Record filenames (only the last 8 characters) and image sizes, the size is in the middle of the list and the filename is near the bottom.
         c. Repeat for all remaining images.
   E. To download images to portable disks.
      1. Close AFIT, TEST window.
      2. Select "File then Open" (the same exam should already be selected).
      3. Select "Images then Verify" to change "I" to "Y".

4. Select "Images then Store Exam".
5. Make sure the numbered disk is already selected.
6. Select "Store". When the disk is full, remove the disk and insert a new one (make sure that you are using clear disks). Note that each disk holds approximately 12-15 mammograms.

F. To exit Lite Box select "File then Quit" from the top button bar, DO NOT SAVE CHANGES.

III. To reset the Big Mac and digitizer.

A. To remove the remaining 150MB disks from the drive, select "Special then Eject Disk".
B. From the main window select "Special then Shutdown". THE SYSTEM MUST BE SHUT DOWN BEFORE RECONNECTING THE DIGITIZER.
C. Carefully reconnect the digitizer. (Reverse drive connect steps.)
D. Press the restart button and then the startup key.

IV. Store the drive adapter cable.

A.4 File transfer instructions

Digitized Mammogram Image Transfer Instructions

I. To transfer digitized mammogram files to AFIT.
   A. Go to room 2011.
   B. Locate the adapter cable.
   C. Shut down the host system.
   D. Connect the portable hard drive.
   E. Insert disk.
   F. Open the numbered drive window.
   G. Select local exams and then open the numbered folder.
   H. Now rename the files, removing all but the last 8 characters.
   I. Open Vulcan and select Applications folder.
   J. Open TCP/Connect II Folder and select TCP/Connect II.
   K. Select "Services" on the button bar, then "FTP".
   L. Click on "more options" and unselect "Use Web-Style FTP client"
   M. Set the host name to pinna and set username.
   N. Click Image data type.
   O. Under "Options", select "Binary" and unselect "Prompt for filenames".
   P. Set the directory for storage (for example /home/pinna1/bdata/Uncut).
   Q. Click on Desktop, select the files and click on "Copy".

Breast Cancer Group Filenaming Guide

Breakdown of a003a05m.dog: a 003 a 0 5 m .dog		
a	Source	a: Wright-Patterson AFB Medical Center, WPAFB, OH g: Grandview Hospital, Dayton, OH s: Southview Hospital, Dayton, OH f: Franciscan Health Center, Dayton Campus, Dayton, OH
003	Case Number	Each case typically contains 1-4 digitized mammograms, and corresponds to a single date
a	Orientation	a: RCC b: LCC c: RMLO d: LMLO e: RML f: LML
0	Date Number	Dates are represented in increasing order, where the earliest is represented by 0 and the latest is the highest integer value.
5	Subimage Index	If not 0, then this number is used to index subimages extracted from each digitized mammogram.
m	Diagnosis	m: Malignant, Cancer b: Benign u: Unknown, no biopsy n: Not available yet
.dog	File Type	dog - Difference of Gaussians filtered, grayscale, BIN hit - Hit and miss filtered, grayscale, BIN mam - Original mammogram, grayscale, BIN nul - Region without lesions, 256×256 pixels, grayscale, BIN pcn - PCNN processed SUBIMAGE roi - SUBIMAGE, 256×256 pixels, grayscale, BIN ro2 - SUBIMAGE, 512×512 pixels, grayscale, BIN tru - Radiologist and/or pathology verified truth, ASCII txt - Comments (corresponding date, etc.), ASCII unc - Uncut (full digitized mammogram), grayscale, BIN wav - Wavelet processed, grayscale, BIN

Appendix B. Matlab and C Code

The file *eucdet.m* is the parent Matlab function used to automatically detect clustered microcalcifications. The CMEX file, *findcluster.c* was written by Capt. Amy Magnus and modified by Capt. Edward Ochoa to compile on Solaris.

```
::::::::::::::
GetFileNames.m
::::::::::::::
function [FullFileNames,JustFileNames] = GetFileNames(ListOfFiles,DirOfFiles);
% GETFILENAMES Used to load ascii filenames into matlab.
% function FullFileNames = GetFileNames(ListOfFiles,DirOfFiles);
% ASSUMES: All filenames have the same number of characters
% EX: FullFileNames = GetFileNames('filenames.txt','/home/DataBaseDir')
% EX: FullFileNames = GetFileNames('filenames.txt',0) % if in cwd
%
% by: Capt. Edward M. Ochoa, GEO-96D

FileID = fopen(ListOfFiles,'r');
if FileID == -1,
   error(sprintf('ERROR: %s did not open correctly (GetFullFileNames.m)',ListOfFiles));
end;

FileList=fread(FileID);
FileList=setstr(FileList');

for filenum=1:sum(isspace(FileList))
   if isstr(DirOfFiles)
      if ~isdir(DirOfFiles),
         error(sprintf('ERROR: %s is not a directory (GetFullFileNames.m)',DirOfFiles));
      end;

      [FileN,FileList]=strtok(FileList);
      FullFileNames(filenum,:)=sprintf('%s/%s',DirOfFiles,FileN);
      JustFileNames(filenum,:)=sprintf('%s',FileN);

   else
      [FullFileNames(filenum,:),FileList]=strtok(FileList);

   end

   FileNameID = fopen(FullFileNames(filenum,:),'r');
   if FileNameID == -1,
     error(sprintf('ERROR: %s not found\nCheck DirOfFiles (GetFullFileNames.m)',FullFileNames(filenum,:)));
   end;
   fclose(FileNameID);

end

fclose(FileID);

::::::::::::::
GetMammo.m
::::::::::::::
function mammo=GetMammo(fullfilename,NUMROWS,DISPLAYTF);
% GETMAMMO Used to load a breast cancer database image.
% mammo=GetMammo(fullfilename,NUMROWS,DISPLAYTF);
%  LOADS FULL MAMMO, REGARDLESS OF SIZE
% by: Capt Amy Magnus, Capt. Edward M. Ochoa, GEO-96D

FileNotFound = 0;
fid = fopen(fullfilename, 'r');
if fid == -1
```

```
   FileNotFound = 1;
end;

% IF FILE IS FOUND, LOAD DIGITIZED MAMMOGRAM

if ~FileNotFound
    [mammo,cnt] = fread(fid,inf,'ushort');
    numcols = cnt/NUMROWS;
    mammo = reshape(mammo,NUMROWS,numcols);

    clear cnt
    fclose(fid);
else
    error(sprintf('%s not found',fullfilename));
end;

if DISPLAYTF
   mamodisp=mammo(1:4:NUMROWS,1:4:numcols);
   colormap(gray(4095))
   image(mamodisp)
   axis image
   title(sprintf('%s',fullfilename));
   axis off
end
```

```
:::::::::::::::
PutMammo.m
:::::::::::::::
function PutMammo(M,fullfilename)
% PUTMAMMO Used to write out an image using the breast cancer database format.
% PutMammo(M,fullfilename)
% This m-file writes the matrix M to
% the file 'fullfilename'. The onus
% is on the user to remeber M's dimensions.
% by: Capt Amy Magnus, Capt. Edward M. Ochoa, GEO-96D

fid=fopen(fullfilename, 'w+');
fwrite(fid,M,'ushort');
fclose(fid);
```

```
:::::::::::::::
ebuildpt.m
:::::::::::::::
function detect = ebuildpt(xon,yon,numrows,numcols);
% EBUILDPT Given a set of pixel coordinates, construct an mask image.
% detect = ebuildpt(xon,yon,numrows,numcols);
% by: Capt. Edward M. Ochoa, GEO-96D

detect=zeros(numrows,numcols);

for i=1:length(xon)

   detect(xon(i),yon(i))=1;

end
```

```
:::::::::::::::
ecenters.m
:::::::::::::::
function [centers]=ecenters(inlist,map);
% ECENTERS Compute centroids of clusters.
% [centers]=ecenters(inlist,map);
% by: Capt. Edward M. Ochoa, GEO-96D

centers=zeros(size(map));
```

```
for i=1:size(map,1);
   xy=inlist(map(i,1):map(i,2),:);
   if size(xy,1)==1
      centers(i,:)=xy;
   else
      centers(i,:)=round(mean(xy));
   end
end

:::::::::::::::
echeckpts.m
:::::::::::::::
function [goodpts]=echeckpts(origimg,ptcoords);
% ECHECKPTS Used to verify detections were made within breast
% [goodpts]=echeckpts(origimg,ptcoords);
% by: Capt. Edward M. Ochoa, GEO-96D

if ptcoords(1,1)~=-1
   mask=esegmam(origimg);
   numpts=size(ptcoords,1);
   keep=ones(numpts,1);
   for i=1:numpts
      if ~mask(ptcoords(i,1),ptcoords(i,2))
        keep(i)=0;
      end
   end
   goodpts=ptcoords(keep,:);
   if isempty(goodpts)
      goodpts=[-1 -1];
   end
else
   goodpts=ptcoords;
end

:::::::::::::::
ecrnois.m
:::::::::::::::
function [outimg]=ecrnois(img);
% ECRNOIS Cross-shaped hybrid median filter an image.
% [outimg]=ecrnois(img);
% by: Capt. Edward M. Ochoa, GEO-96D

winsize=3;

[numrows,numcols]=size(img);
bufr = (winsize-1)/2;

% MY "MIRROR" PADDING TO ELIMINATE EDGE EFFECTS IN CODE

img=emirrpad(img,winsize);

% INITIALIZE OUTPUT IMAGE;

outimg = zeros(numrows,numcols);


% ELIMINATE NOISE

grpv=zeros(5,1);

temp=0;
j=0;
i=0;
minval=0;
```

```
minind=0;

for m=1:numrows
   mb=m+bufr;
   for n=1:numcols
      nb = n+bufr;


      % GET + PIXELS
      grpv(1)=img(mb,nb-1);
      grpv(2)=img(mb,nb);
      grpv(3)=img(mb,nb+1);
      grpv(4)=img(mb-1,nb);
      grpv(5)=img(mb+1,nb);

      for i=1:3
         minval=grpv(i);
         minind=i;
         for j=i+1:5
            if grpv(j)<minval
               minval=grpv(j);
               minind=j;
            end
         end
         if (minval<grpv(i))
            temp=grpv(i);
            grpv(i)=minval;
            grpv(minind)=temp;
         end
      end

      outimg(m,n)=grpv(3);

   end

end


::::::::::::::
edetclst.m
::::::::::::::
function [inlist,map]=edetclst(list,radius,numpts);
% EDETCLST Used to detect clusters of pts
% [inlist,map]=edetclst(list,radius,numpts);
% by: Capt. Edward M. Ochoa, GEO-96D

pairs = epairs(list,radius);
inlist=[];
map=[];

if ~isempty(pairs)

   [ind,indmap]=erecolr(pairs);

   startr=1;
   for i=1:size(indmap,1)
      inds=ind(indmap(i,1):indmap(i,2));
      clstr=list(inds,:);
      numcpts=size(clstr,1);
      if numcpts>=numpts
         inlist=[inlist;clstr];
         endr=startr+numcpts-1;
         map=[map;startr endr];
         startr=endr+1;
      end
```

```
   end

end

if isempty(inlist)
   inlist=[-1 -1];
   map=inlist;
end

:::::::::::::::
edogk.m
:::::::::::::::
function [dogkernel,G1,G2]=edogk(targetsize);
% EDOGK Used to construct the DoG kernel given targetsize.
% [dogkernel,G1,G2]=edogk(targetsize);
% by: Capt. Edward M. Ochoa, GEO-96D

if (targetsize>33) | (targetsize<=0)
   disp('Sorry, limited to 0 < targetsize < 33 pixels');
   return
end

sig = [.391 .625 1 1.6 2.56 4.096 6.536 10.486 16.777];
Nsize = [3 5 9 11 15 21 29 43 55];

s2 = targetsize/2;
s1=s2/1.6;

% Find optimum N (kernel size)

i=1;
while sig(i)<s2
   i=i+1;
end

N=Nsize(i);

G1 = fspecial('gaussian',N,s1);
G2 = fspecial('gaussian',N,s2);
dogkernel = G1 - G2;

:::::::::::::::
edrawROI.m
:::::::::::::::
function edrawROI(inlist,map,numrows,numcols,drawPtsTF)
% EDRAWROI Used to highlight ROIs
% edrawROI(inlist,map,numrows,numcols,drawPtsTF)
% by: Capt. Edward M. Ochoa, GEO-96D

for i=1:size(map,1)
  index=map(i,1):map(i,2);
  M=[inlist(index,2) inlist(index,1)];
  if i==1
    if det(cov(M))
       ellipse(mean(M),inv(cov(M)),3.5);
    else
       axis('ij');
       plot(M(:,1),M(:,2),'r+');
    end
    axis('ij');
    hold on;
    if drawPtsTF & (det(cov(M)))
       plot(M(:,1),M(:,2),'b+');
    end
    axis([1 numcols 1 numrows]);
```

```
    axis('image');
  else
    if drawPtsTF & (det(cov(M)))
       plot(M(:,1),M(:,2),'b+');
    end
    if det(cov(M))
       ellipse(mean(M),inv(cov(M)),3.5);
    else
       axis('ij');
       plot(M(:,1),M(:,2),'r+');
    end
  end
end


:::::::::::::::
efilt.m
:::::::::::::::
function [outimg] = efilt(img,numbits,kernel);
% EFILT Handy utility for convolving (filtering) an image with a kernel.
% [outimg] = efilt(img,numbits,kernel);
% by: Capt. Edward M. Ochoa, GEO-96D

[numrows,numcols]=size(img);
N=size(kernel,1);
bufr = (N-1)/2;

% MY "MIRROR" PADDING TO ELIMINATE EDGE EFFECTS IN CODE
img=emirrpad(img,N);

% RESCALE THE IMAGE TO INTENSITY VALUES [0,1];
img=erescale(img);

% CONVOLVE THE KERNEL WITH THE IMAGE
outimg=conv2(img,kernel,'valid');

% RESET THE OUTPUT IMAGE TO THE COLORDEPTH DESIGNATED BY NUMBITS
outimg=round((2^numbits-1)*erescale(outimg));



:::::::::::::::
egettruth.m
:::::::::::::::
function truth=egettruth(SOURCEDIR,IMGNAME,class);
% EGETTRUTH Utility for loading pathology truth.
% truth=egettruth(SOURCEDIR,IMGNAME,class);
% by: Capt. Edward M. Ochoa, GEO-96D

TRUTHFULLFILENAME=[SOURCEDIR IMGNAME '.tru'];
eval(sprintf('load %s;',TRUTHFULLFILENAME))
eval(sprintf('truth=%s;',IMGNAME));
targetindex=find(truth(:,5)==class);
if ~isempty(targetindex)
   truth=truth(targetindex,:);
else
   truth=-1;
end

:::::::::::::::
eglth2.m
:::::::::::::::
function glthr=eglth2(fimg,NUMBITS,globp,segwinsize,k,ktop);
% EGLTH Implements global and local thresholding for segmentation/detection.
% glthr=eglth2(fimg,NUMBITS,globp,segwinsize,k,ktop);
```

```
% by: Capt. Edward M. Ochoa, GEO-96D

gl=ehistgr(fimg,NUMBITS,globp);
glthr=elthresh2(fimg,NUMBITS,segwinsize,k,gl,ktop);
glthr=bwmorph(glthr,'bridge'); % BRIDGE 1 PIXEL-WIDE GAPS

::::::::::::::
ehist.m
::::::::::::::
function hgram = ehist(img,numbits);
% EHIST Utility for computing an image histogram.
% hgram = ehist(img,numbits);
% by: Capt. Edward M. Ochoa, GEO-96D

ind=0;
img=round(img(:));
limg=size(img,1);
hgram=zeros( realpow(2,numbits) , 1 );
for i=1:limg
   ind=img(i)+1;
   hgram(ind)=hgram(ind)+1;
end

::::::::::::::
ehistgr.m
::::::::::::::
function [gr,imhst]=ehistgr(img,NUMBITS,percentile)
% EHISTGR Used to segment an image using a percentile of an image histogram.
% gr=ehistgr(img,NUMBITS,percentile)
% by: Capt. Edward M. Ochoa, GEO-96D

imhst=ehist(img,NUMBITS);
numcolrs=length(imhst);
chst=cumsum(imhst);
cdf=chst/chst(numcolrs);
gr=min(find(cdf>=percentile));

::::::::::::::
ellipse.m
::::::::::::::
function ellipse(m,Q,k);
% ELLIPSE Used to plot an ellipse.
% ellipse(m,Q,k);
%
% Plots the ellipse described by
%
% {x: (x-m)'Q(x-m) = rsq}
%
% where
% m = mean of data points. (vector)
% Q = inverse of covariance of data points. (matrix)
% k = number of std devs to include in ellipse. (3 works well)
%
% USAGE: ellipse(mean([xpts ypts]),inv(cov([xpts ypts])),k)
% by: Maj Ruck

m = m(:);
pts = 256;
circle_cmplx = exp(j*2*pi/(pts-1)*[0:pts-1]);
circle = [real(circle_cmplx); imag(circle_cmplx)];
ellip = sqrt(k)*Q^(-0.5)*circle + m*ones(1,pts);

ecolor1='k';
lwidth1=2;
style1='-';
```

```
ecolor2='r';
lwidth2=1;
style2='-';

h1=plot(ellip(1,:),ellip(2,:));
set(h1,'LineWidth',lwidth1);
set(h1,'Color',ecolor1);
set(h1,'LineStyle',style1)
hold on

h2=plot(ellip(1,:),ellip(2,:));
set(h2,'LineWidth',lwidth2);
set(h2,'Color',ecolor2);
set(h2,'LineStyle',style2)

::::::::::::::
elthresh2.m
::::::::::::::
function [outimg]=elthresh2(img,numbits,winsize,LTH,MINTHRESH,LTHTOP);
% ELTHRESH Implements local and global thresholding
% [outimg]=elthresh2(img,numbits,winsize,LTH,MINTHRESH,LTHTOP);
% by: Capt. Edward M. Ochoa, GEO-96D

MAXGRAY=2^numbits-1;
% BEGIN THRESHOLD TEST

[numrows,numcols]=size(img);
bufr = (winsize-1)/2;

% MY "MIRROR" PADDING TO ELIMINATE EDGE EFFECTS IN CODE

img=emirrpad(img,winsize);


% INITIALIZE OUTPUT IMAGE;

outimg = zeros(numrows,numcols);

% COMPUTE FIRST ROW WINDOW'S TOTALS

tot=0;
Oldtot=0;
totOfsquares=0;
OldtotOfsquares=0;
winsize2=realpow(winsize,2);

for m=1:winsize
   for n=1:winsize
      tot = tot + img(m,n);
      totOfsquares = totOfsquares + img(m,n)^2;
   end
end

% INITIALIZE FIRST ROW WINDOW'S RESULTS
Oldtot=tot;
OldtotOfsquares = totOfsquares;

% COMPUTE LOCAL MEAN AND STANDARD DEVIATION
mn = tot/winsize2;
sdev = realpow( (totOfsquares - tot^2/winsize2)/(winsize2-1), 1/2 );

% COMPUTE FIRST THRESHOLD TEST
seed=img(bufr+1,bufr+1);
if (seed>(mn + LTH*sdev)) & (seed<(mn + LTHTOP*sdev)) & (seed> MINTHRESH)
   outimg(1,1)=1;
```

```
end

% COMPUTE THRESHOLD TEST FOR REST OF IMG

for m=1:numrows
   mb=m+bufr;
   tot=Oldtot;
   totOfsquares=OldtotOfsquares;
   for n=2:numcols
      nb = n+bufr;
      for index=-bufr:bufr
         oldpix = img(mb+index,nb-bufr-1);
         newpix = img(mb+index,nb+bufr);
         tot = tot -  oldpix + newpix;
         totOfsquares = totOfsquares - oldpix^2 +  newpix^2;
      end

      % FOR COLUMNS 2:NUMCOLS, COMPUTE THRESHOLD TEST

      mn = tot/winsize2;
      sdev = realpow( (totOfsquares - tot^2/winsize2)/(winsize2-1), 1/2 );

      seed=img(mb,nb);
      if (seed>(mn + LTH*sdev)) & (seed<(mn + LTHTOP*sdev)) & (seed> MINTHRESH)
         outimg(m,n)=1;
      end

   end

   % SINCE AT LAST COL, THEN FOR NEXT ROW, COLUMN 1,
   % COMPUTE THRESHOLD TEST

   if m~=numrows
      nb = bufr + 1;
      tot=Oldtot;
      totOfsquares=OldtotOfsquares;
      for index=-bufr:bufr
         oldpix = img(mb-bufr,nb+index);
         newpix = img(mb+bufr+1,nb+index);
         tot = tot -  oldpix + newpix;
         totOfsquares = totOfsquares - oldpix^2 +  newpix^2;
      end

      Oldtot = tot;
      OldtotOfsquares = totOfsquares;

      mn = tot/winsize2;
      sdev = realpow( (totOfsquares - tot^2/winsize2)/(winsize2-1), 1/2 );

      seed=img(mb+1,bufr+1);
      if (seed>(mn + LTH*sdev)) & (seed<(mn + LTHTOP*sdev)) & (seed> MINTHRESH)
         outimg(m+1,1)=1;
      end

   end

end

:::::::::::::::
emailmsg.m
:::::::::::::::
function emailmsg(emailaddress,message);
% EMAILMSG Used to send email message to user.
% emailmsg(emailaddress,message);
% by: Capt. Edward M. Ochoa, GEO-96D
```

```
UCMD1=sprintf('echo "%s" > /tmp/emailmsg.txt',message);
UCMD2=sprintf('echo "~s %s" >> /tmp/emailmsg.txt',message);
UCMD3=sprintf('echo "~." >> /tmp/emailmsg.txt');
UCMD4=sprintf('mail %s < /tmp/emailmsg.txt',emailaddress);
UCMD5=sprintf('!rm /tmp/emailmsg.txt');

UCMDS=['! ' UCMD1 ';' UCMD2 ';' UCMD3 ';' UCMD4 ';'];

eval(UCMDS)
eval(UCMD5)

:::::::::::::::
emakeodd.m
:::::::::::::::
function oddnum = emakeodd(num);
% EMAKEODD Used to convert input to closest odd integer (ex: 60 -> 61)
% oddnum = emakeodd(num);
% by: Capt. Edward M. Ochoa, GEO-96D

rnum=round(num);
ndiff=num-rnum;

isoddTF=rem(rnum,2);

if ~isoddTF
   orignum=rnum+num;
   upnum=rnum+1;
   downnum=rnum-1;
   if abs(upnum-orignum)<=abs(downnum-orignum)
      oddnum=upnum;
   else
      oddnum=downnum;
   end
else
   oddnum=rnum;
end

:::::::::::::::
emirrpad.m
:::::::::::::::
function img=emirrpad(img,winsize)
% EMIRRPAD Used to mirror pad image.
% img=emirrpad(img,winsize)
% by: Capt. Edward M. Ochoa, GEO-96D

[numrows,numcols]=size(img);
bufr = (winsize-1)/2;


% MY "MIRROR" PADDING TO ELIMINATE EDGE EFFECTS IN CODE

tbuf = flipud(img(2:1+bufr,:));
bbuf = flipud(img(numrows-bufr:numrows-1,:));
corner = ones(bufr,bufr);
lbuf = [corner;fliplr(img(:,2:1+bufr));corner];
rbuf = [corner;fliplr(img(:,numcols-bufr:numcols-1));corner];

img = [lbuf [tbuf;img;bbuf] rbuf];

tlc = flipud(fliplr(img(1+bufr:winsize,1+bufr:winsize)));
trc = flipud(fliplr(img(1+bufr:winsize,numcols:numcols+bufr)));
blc = flipud(fliplr(img(numrows:numrows+bufr,1+bufr:winsize)));
brc = flipud(fliplr(img(numrows:numrows+bufr,numcols:numcols+bufr)));
```

```
img(1:1+bufr,1:1+bufr)=tlc;
img(1:1+bufr,numcols+bufr:numcols+winsize-1)=trc;
img(numrows+bufr:numrows+winsize-1,1:1+bufr)=blc;
img(numrows+bufr:numrows+winsize-1,numcols+bufr:numcols+winsize-1)=brc;

:::::::::::::::
epairs.m
:::::::::::::::
function pairs = epairs(list,radius);
% pairs = epairs(list,radius);
% by: Capt. Edward M. Ochoa, GEO-96D

numpts=size(list,1);
pairs=[];
for i=1:numpts
   pt1=list(i,:);
   for j=i+1:numpts
      pt2=list(j,:)';
      if dist(pt1,pt2)<=radius
         pairs = [pairs;i j];
      end
   end
end

:::::::::::::::
erecolr.m
:::::::::::::::
function [list,map]=erecolr(crash);
% ERECOLR Used to identify the number of unique colors given crash list
% [list,map]=erecolr(crash);
% by: Capt. Edward M. Ochoa, GEO-96D

list=[];
map=[];

blist=1;

tmp=crash;
pairs=[];
i=0;
keepcheckingTF=1;
pairs=tmp(1,:);
while keepcheckingTF
   for j=1:size(pairs,2)
      rows=[];
      if ~isempty(tmp)
         [rows,cols]=find(pairs(j)==tmp);
      end
      if ~isempty(rows)
         rows=eunique(rows);
         newp=tmp(rows,:);
         newp=eunique(newp);
         pairs=eunique([pairs newp]);
         tmp(rows,:)=[];
      end
   end
   if isempty(tmp)
      keepcheckingTF=0;

      list=[list; pairs(:)];
      elist=blist+size(pairs(:),1)-1;
      map=[map;blist elist];
      blist=elist+1;
   elseif ~any(evsame(pairs,tmp(:)))
```

```
      list=[list; pairs(:)];
      elist=blist+size(pairs(:),1)-1;
      map=[map;blist elist];
      blist=elist+1;

      pairs=tmp(1,:);
   end
end

:::::::::::::::
erescale.m
:::::::::::::::
function imgout = erescale(img);
% ERESCALE Used to rescale any image to the [0,1] range.
% imgout = erescale(img);
% by: Capt. Edward M. Ochoa, GEO-96D

lo = min(min(img));
hi = max(max(img));

imgout=(img-lo)/(hi-lo);

:::::::::::::::
erocem.m
:::::::::::::::
function [truepos, falsepos, goodroi]=erocem(truth,inlist,map);
% EROCEM Utility for identifying whether a CLUSTER was detected.
% [truepos, falsepos, goodroi]=erocem(truth,inlist,map);
% by: Capt. Edward M. Ochoa, GEO-96D

% A cluster is considered a TP if any of the following conditions exist
%  1: ALL cluster pixels were within a truth box.
% 2: At least MINDETS were found within a corresponding
% truth box and the centroid of the cluster is within
%               the truth box perimeter.
% 3: At least MINDETS were found within a corresponding
% truth box and the centroid of the cluster is within MINDIST
% pixels from the truth box perimeter.
%
% Note: (1)-(3) may be combined into the following statement:
% At least MINDETS were found within a corresponding truth box
% and the cluster's centroid is within MINDIST pixels from the truth box

MINDIST=50; % 5mm for 100um digitized images
MINDETS=1;

if truth(1,1)~=-1

   truepos=0;
   good=ones(size(map,1),1);
   goodroi=ones(size(map,1),1);

   centroids=ecenters(inlist,map);
   for i=1:size(map,1)
      good(i)=0;
      index=map(i,1):map(i,2);
      centroid=centroids(i,:)';

      % FOR EACH CLUSTER, COUNT THE NUMBER OF DETS WITHIN A TRUTH BOX
      for k=index
         for j=1:size(truth,1)
            rlo=truth(j,2);
            rhi=truth(j,4);
            row=inlist(k,1);
            rin=(rlo<=row)&(row<=rhi);
```

```
            clo=truth(j,1);
            chi=truth(j,3);
            col=inlist(k,2);
            cin=(clo<=col)&(col<=chi);
            inboxTF = (rin & cin);
            if inboxTF
               good(i)=good(i)+1;
            end
         end
      end

      numgood=good(i);
      good(i)=good(i)>MINDETS-1;

      % IF MINDETS WITHIN A TRUTH BOX IS SATISFIED AND ALL DETS ARE NOT
      % WITHIN A TRUTH BOX, CHECK IF CENTROID WITHIN TRUTH BOX
      if good(i) & (numgood~=size(index,2))
         crow=centroid(1);
         ccol=centroid(2);
         for j=1:size(truth,1)
            rlo=truth(j,2);
            rhi=truth(j,4);
            rin2=(rlo<=crow)&(crow<=rhi);
            clo=truth(j,1);
            chi=truth(j,3);
            cin2=(clo<=ccol)&(ccol<=chi);
            centroidinboxTF = (rin2 & cin2);
         end

         if centroidinboxTF
            good(i)=1;
         % IF MINDETS WITHIN A TRUTH BOX IS SATISFIED AND ALL DETS ARE NOT
         % WITHIN A TRUTH BOX, AND CENTROID NOT IN BOX, CHECK MINDIST
         % OF CLUSTER CENTROID
         else
            good(i)=0;
            for j=1:size(truth,1)
               rlo=truth(j,2);
               rhi=truth(j,4);
               clo=truth(j,1);
               chi=truth(j,3);
               boxpts = etrubox(rlo,rhi,clo,chi);  % PERIMETER PIXELS
               centroidv=centroid*ones(1,size(boxpts,1));
               % dists=dist(boxpts,centroidv);
               dists=edist(boxpts,centroidv);
               mindistTF=any(any(dists<MINDIST));
               if mindistTF
                  good(i)=good(i) | 1;
               end
            end
         end
      end

      if ~good(i)
         goodroi(i,1)=0;
      end
   end
   truepos=sum(goodroi==1)>0;
   falsepos=sum(goodroi==0);

else
   truepos=0;
   falsepos=size(map,1);
   goodroi=-1*ones(size(map,1));
end
```

```
:::::::::::::::
esegmam.m
:::::::::::::::
function mammask = esegmam(mam);
% ESEGMAM Used to segment breast mass from full mammogram.
% mammask = esegmam(mam);
% by: Capt. Edward M. Ochoa, GEO-96D

[numrows,numcols]=size(mam);

% SUBSAMPLE TO SIMPLIFY SEGMENTATION
mams=mam(1:16:numrows,1:16:numcols);

% HYBRID MEDIAN FILTER TWICE TO MERGE REGIONS
mams=ecrnois(mams);
mams=ecrnois(mams);

% COMPUTE AND RESCALE VARIANCE IMAGE (TO ENHANCE BREAST EDGE)
mamv=erescale(log(1+evarop(mams,12,9)));

% ADD VARIANCE IMAGE TO ORIGINAL (SIMPLIFIES SEGMENTATION)
mamsv=erescale(mamv+erescale(mams));

% USE FLOYD-STEINBERG DITHERING TO CONVERT TO BW
mammask=dither(mamsv);

% APPLY MORPHOLOGICAL OPERATORS TO SEGMENT BREAST
mammask=bwmorph(mammask,'majority',4);
mammask=bwmorph(mammask,'open',2);
mammask=bwmorph(mammask,'erode',3);

% REMOVE LABEL IF NECESSARY
mammask=findcluster(mammask);
maxlabel=max(mammask(:));
maxc=1;
for i=1:maxlabel-1
   if sum(mammask(:)==i)<sum(mammask(:)==(i+1))
      maxc=i+1;
   end
end

mammask=mammask==maxc;
[numr,numc]=size(mams);

% ORIENT IF NECESSARY (TO SIMPLIFY FILL)
lhs=sum(mams(:,1));
rhs=sum(mams(:,numc));
if lhs>rhs
   lmamTF=1;
else
   lmamTF=0;
end
if ~lmamTF
   mammask=fliplr(mammask);
end

% HORIZONTAL FILL
colsum=sum(mammask);
ind=min(find(max(colsum)==colsum));
fillcol=mammask(:,ind);
for i=1:ind
   mammask(:,i)=mammask(:,i) | fillcol;
end
```

```
% VERTICAL FILL
rowsum=sum(mammask');
begrow=min(find(rowsum));
endrow=max(find(rowsum));
topr=min(find(rowsum==max(rowsum(find(rowsum~=0)))));
botr=max(find(rowsum==max(rowsum(find(rowsum~=0)))));

sumr=max(rowsum);
for i=begrow:topr
   if (rowsum(i)<sumr) & (rowsum(i)~=0)
      sumr=rowsum(i);
      toprind=i;
   end
end
toprfill=mammask(toprind,:);
for i=begrow:topr
   mammask(i,:)=mammask(i,:) | toprfill;
end

sumr=max(rowsum);
for i=botr:endrow
   if (rowsum(i)<sumr) & (rowsum(i)~=0)
      sumr=rowsum(i);
      botrind=i;
   end
end
botrfill=mammask(botrind,:);
for i=botr:endrow
   mammask(i,:)=mammask(i,:) | botrfill;
end

% CONVERT BACK TO FULL SIZE MASK
if ~lmamTF
   mammask=fliplr(mammask);
end

mammask=imresize(mammask,[numrows numcols],'nearest');
mammask=round(mammask);

::::::::::::::
esepname.m
::::::::::::::
function [name,ext,fname,nrows]=esepname(FULLNAMES,isnameTF);
% ESEPNAME Handy string manipulator used to break down filenames, and ID num rows
% [name,ext,fname,nrows]=esepname(FULLNAMES,isnameTF);
% by: Capt. Edward M. Ochoa, GEO-96D

fname=[];
name=[];
ext=[];

if ~isnameTF
   revfullname=fliplr(FULLNAMES);

   for i=1:size(FULLNAMES,1)
      [n,rmder]=strtok(revfullname(i,:),'/');
      fname=[fname;n];
   end

   fname=fliplr(fname);
else
   fname=FULLNAMES;
end

for i=1:size(fname,1)
```

```
   [n,rmder]=strtok(fname(i,:),'.');
   name=[name;n];
   ext=[ext;rmder(2:size(rmder,2))];
end

for i=1:size(ext,1)
   if 3==sum(ext(i,:)=='mam')
      nrows(i,1)=2048;
   elseif 3==sum(ext(i,:)=='roi')
      nrows(i,1)=256;
   elseif 3==sum(ext(i,:)=='ro2')
      nrows(i,1)=512;
   end
end
```

```
:::::::::::::::
eshrink.m
:::::::::::::::
function [pts,ptcoords] = eshrink(detmask);
% ESHRINK Used to reduce detections to single points. Facilitates clustering.
% [pts,ptcoords] = eshrink(detmask);
% by: Capt. Edward M. Ochoa, GEO-96D


if sum(sum(detmask))==0
   pts=zeros(size(detmask));
else
   [numrows,numcols]=size(detmask);
   detmask=findcluster(detmask);

   % Get histogram on 'values' image
   h=ehist(detmask,8);
   obj_colors=find(h~=0);
   nc=length(obj_colors)-1;
   ptcoords=ones(nc,2);

   if (nc~=0)
      for i=1:nc
         [x,y]=find(detmask==i);
         if size(x,1)==1
            xy=[x y];
         else
            xy=round(mean([x y]));
         end
         ptcoords(i,:)=xy;
      end
   end

   pts=ebuildpt(ptcoords(:,1),ptcoords(:,2),numrows,numcols);

end
```

```
:::::::::::::::
etrubox.m
:::::::::::::::
function boxpts = etrubox(rlo,rhi,clo,chi);
% ETRUBOX Utility for converting truth indicies to perimeter pts
% boxpts = etrubox(truth);
% by: Capt. Edward M. Ochoa, GEO-96D

topcol=[clo:chi];
toprow=rlo*ones(size(topcol));

rgtrow=rlo+1:rhi;
rgtcol=chi*ones(size(rgtrow));
```

```
botcol=chi-1:-1:clo;
botrow=rhi*ones(size(botcol));


lftrow=[rhi-1:-1:rlo+1];
lftcol=clo*ones(size(lftrow));

boxpts=[toprow' topcol';...
rgtrow' rgtcol';...
botrow' botcol';...
lftrow' lftcol'];

:::::::::::::::
eucdet.m
:::::::::::::::
function ...
[inlist,map,numrows,numcols,roiTF,TPROI,FPROI,truth]=...
   eucdet( PARAMS,...
IMGNAMEEXT,...
DoStatsTF,...
DisplayTF);
% EUCDET Detect Clustered Microcalcifications: Capt. Edward M. Ochoa, GEO-96D
%
% [inlist,map,numrows,numcols,roiTF,TPROI,FPROI,truth]=...
%   eucdet( PARAMS,... <-- DETECTION PARAMETERS (described below)
% IMGNAMEEXT,...
% DoStatsTF,... <-- Compute ROC stats? (yes=1, no=0)
% DisplayTF); <-- Display processing results? (yes=1, no=0)
%
% DESCRIPTION OF DETECTION PARAMS = [p1 p2 p3 p4 p5 p6 p7] AND DEFAULT SETTINGS
% p1 - DOGT, targetsize (1.1995)
%  p2 - K, multiple of local std for local thresholding (4.1226)
%  p3 - GLOBP, global histogram percentile cutoff (0.9991)
%  p4 - DETWINSZ, odd window size used in local thresholding (69)
%  p5 - CLSTRRAD, max dist between neighbor microcalcs (51.3019)
%  p6 - NUMCALCS, min number of calcs to trigger a cluster detection (3)
%  p7 - KTOP, max multiple of local std for local thresholding (7)
%
% EXAMPLE - HOW TO USE DEFAULT SETTINGS:
%   [inlist,map,numrows,numcols]=eucdet([],'a036a01m.roi',1,1);
%   [inlist,map,numrows,numcols]=eucdet([],'a036a00m.mam',1,1);
%
% EXAMPLE - HOW TO DISPLAY DETECTION RESULTS:
%   edrawROI(inlist,map,numrows,numcols,1);
%
% by: Capt. Edward M. Ochoa, GEO-96D

% ---------- ALGORITHM SETTINGS ----------------------------------------
% DEFAULT PARAMETER SETTINGS
if isempty(PARAMS)
   PARAMS=[1.1995 4.1226 0.9991 69  51.3019 3 7];
end

% IMG INFO
[IMGNAME,EXT,NAME,IMGNUMROWS]=esepname(IMGNAMEEXT,0);
ProcMamTF=(sum((1*EXT==ones(size(EXT,1),1)*'mam')')==3)';

% DIR, EXT, AND IMG DEPTH INFO
if ProcMamTF
   SOURCEDIR='';
   MAMTRUDIR='/home/hawkeye3/96d/eochoa/Thesis/MC_code/MAMTRU/';
else
   SOURCEDIR='/home/pinna1/bdata/wpafbh/ROIs/';
```

```
end

NUMBITS=12;

% BLOCK PROCESSING SIZE (REDUCES MEMORY REQUIREMENTS AND SPEEDS UP PROCESSING)
BLKSZ=256;
BLKSZR=BLKSZ;
BLKSZC=BLKSZ;

% DoG TARGETSIZE and filter KERNEL
DOGT=PARAMS(1);
[KERNEL,dogk1,dogk2] =  edogk(DOGT);

% DETECTION (SEGMENTATION) PARAMETERS
K=PARAMS(2);
KTOP=PARAMS(7);
GLOBP=PARAMS(3);
DETWINSZ=emakeodd(PARAMS(4));

% CLUSTERING PARAMETERS
CLSTRRAD=PARAMS(5);
NUMCALCS=round(PARAMS(6));
% ---------- ALGORITHM BEGINS ------------------------------------------
% LOAD IMG
img=GetMammo([SOURCEDIR IMGNAME '.' EXT ],IMGNUMROWS,0);
% img=GetMammo(IMGNAMEEXT,IMGNUMROWS,0); % if IMGNAMEEXT always includes fullpath
[numrows,numcols]=size(img);

if ProcMamTF & (sum(img(:,numcols))>sum(img(:,numcols)))
   FlipTF=1;
   img=fliplr(img);
else
   FlipTF=0;
end

if DisplayTF | ProcMamTF
   origimg=img;
else
   origimg=-1;
end

if DisplayTF
   NUMGRAY=2^NUMBITS;
   figure
   imshow((NUMGRAY-1)*erescale(img),gray(NUMGRAY))
   title(sprintf('%s',NAME))
   orient tall
end
% ----------------------------------------------------------------------
% CLEAN NOISE
img=blkproc(img,[BLKSZR BLKSZC],'ecrnois');

if DisplayTF
   figure
   imshow((NUMGRAY-1)*erescale(img),gray(NUMGRAY))
   title(sprintf('%s.ncl',IMGNAME))
   orient tall
end
% ----------------------------------------------------------------------
% DoG FILTER
P1=NUMBITS;
P2=KERNEL;
img=blkproc(img,[BLKSZR BLKSZC],'efilt',P1,P2);

if DisplayTF
```

```
   figure
   ksize=size(KERNEL,1);
   subplot(2,1,1),mesh(KERNEL),axis square
   axis ij
   axis([1 ksize 1 ksize min([dogk1(:)' dogk2(:)']) max([dogk1(:)' dogk2(:)'])])
   title('DoG filter coefficients')
   subplot(2,1,2),freqz2(KERNEL,32,32);, axis square
   title('DoG filter Frequency Response')
   textsc(.5,.97,sprintf('targetsize=%1.2f',DOGT))
   orient tall

   figure
   imshow(img,gray(NUMGRAY))
   title(sprintf('%s.dog',IMGNAME))
   orient tall
end
% ------------------------------------------------------------------------
% USE GLOBAL/LOCAL THRESHOLDING TO SEGMENT/DETECT POTENTIAL TARGETS
P1=NUMBITS;
P2=GLOBP;
P3=DETWINSZ;
P4=K;
P5=KTOP;
img=blkproc(img,[BLKSZR BLKSZC],'eglth2',P1,P2,P3,P4,P5);

if DisplayTF
   figure
   imagesc(img), colormap(1-gray(2)),axis image
   title(sprintf('%s.thr',IMGNAME))
   orient tall
end
% ------------------------------------------------------------------------
% SHRINK DETECTIONS TO SINGLE PIXELS/POINTS
img=blkproc(img,[BLKSZR BLKSZC],'eshrink');

if DisplayTF
   figure
   imagesc(img), colormap(1-gray(2)),axis image
   title(sprintf('%s.pts',IMGNAME))
   orient tall
end
% ------------------------------------------------------------------------
% ELIMINATE FULL MAMMO EDGE DETECTIONS (BASED ON KERNEL SIZE)
if ProcMamTF
   N=size(KERNEL,1);
   bufr=(N-1)/2;
   img(1:bufr,:)=zeros(bufr,numcols);
   img(numrows-bufr+1:numrows,:)=zeros(bufr,numcols);
   img(:,1:bufr)=zeros(numrows,bufr);
   img(:,numcols-bufr+1:numcols)=zeros(numrows,bufr);
end

if FlipTF
   img=fliplr(img);
end

% ------------------------------------------------------------------------
% ID DETECTION/POINT LOCATIONS
[xon,yon]=find(img==1);

if isempty(xon),
   ptcoords=[-1 -1];
else
   ptcoords=[xon yon];
end
```

```
% ------------------------------------------------------------------------
% CHECK IF POINTS WITHIN BREAST
if ProcMamTF
   [ptcoords]=echeckpts(origimg,ptcoords);
end
% ------------------------------------------------------------------------
% USE CLUSTERING CRITERIA TO DETECT ROIS AND INDEX PTS WITHIN ROIS
[inlist,map]=edetclst(ptcoords,CLSTRRAD,NUMCALCS);
% ------------------------------------------------------------------------
if DoStatsTF
   % COMPUTE ROI DETECTION STATISTICS
   class=2;
   if ~ProcMamTF
      truth=egettruth(SOURCEDIR,IMGNAME,class);
   else
      truth=egettruth(MAMTRUDIR,IMGNAME,class);
   end

   if inlist(1,1)~=-1
      if truth~=-1
         [TPROI, FPROI, roiTF]=erocem(truth,inlist,map);
      else
         TPROI=0;
         FPROI=size(map,1);
         roiTF=zeros(size(map,1),1);
      end
   else
      TPROI=0;
      FPROI=0;
      roiTF=-ones(size(map,1),1);
   end
else
   truth=-1;
   TPROI=-1;
   FPROI=-1;
   roiTF=-1;
end
% ------------------------------------------------------------------------
% IF DESIRED, DISPLAY RESULTS
if DisplayTF
   figure
   if inlist(1,1)~=-1
      edrawROI(inlist,map,numrows,numcols,1);
      title(sprintf('ROI(s) in %s',NAME))
      if TPROI~=-1
         xlabel(sprintf('TPROI=%d FPROI=%d',TPROI,FPROI))
      end
   else
      title(sprintf('No ROI(s) in %s',NAME))
      if TPROI~=-1
         xlabel(sprintf('TPROI=%d FPROI=%d',TPROI,FPROI))
      end
   end
   if DoStatsTF
      axis ij
      axis([1 numrows 1 numcols])
      showroitruth(IMGNAME,2)
   end
   orient tall

   figure
   imshow((NUMGRAY-1)*erescale(origimg),gray(NUMGRAY))
   hold on
   if inlist(1,1)~=-1
      edrawROI(inlist,map,numrows,numcols,0);
```

```
      title(sprintf('ROI(s) in %s',NAME))
      if TPROI~=-1
         xlabel(sprintf('TPROI=%d FPROI=%d',TPROI,FPROI))
      end
   else
      title(sprintf('No ROI(s) in %s',NAME))
      if TPROI~=-1
         xlabel(sprintf('TPROI=%d FPROI=%d',TPROI,FPROI))
      end
   end
   if DoStatsTF
      showroitruth(IMGNAME,2)
   end
   orient tall
end
% ----------- ALGORITHM DONE ------------------------------------------


::::::::::::::
eunique.m
::::::::::::::
function uniquels = eunique(v);
% uniqels = eunique(v);
% by: Capt. Edward M. Ochoa, GEO-96D

tmp=v(:);
[nrows,ncols]=size(v);

uniquels=[];
i=0;
keepcheckingTF=1;
while keepcheckingTF
   i=i+1;
   ind=find(v(i)==tmp);
   totun=size(ind,1);
   tmp(ind)=[];
   if totun(1) | (totun==nrows)
      uniquels=[uniquels v(i)];
   end
   if isempty(tmp)
      keepcheckingTF=0;
   end
end

uniquels=sort(uniquels);

::::::::::::::
evarop.m
::::::::::::::
function [outimg]=evarop(img,numbits,winsize);
% EVAROP Computes the standard deviation image given neighborhood size (odd)
% [outimg]=evarop(img,numbits,winsize);
% by: Capt. Edward M. Ochoa, GEO-96D

MAXGRAY=realpow(2,numbits)-1;

% BEGIN VAR ALGORITHM

[numrows,numcols]=size(img);
bufr = (winsize-1)/2;

% MY "MIRROR" PADDING TO ELIMINATE EDGE EFFECTS IN CODE


img=emirrpad(img,winsize);
```

```
% INITIALIZE OUTPUT IMAGE;

outimg = zeros(numrows,numcols);

% COMPUTE FIRST ROW WINDOW'S TOTALS

tot=0;
Oldtot=0;
totOfsquares=0;
OldtotOfsquares=0;
winsize2=realpow(winsize,2);

for m=1:winsize
   for n=1:winsize
      tot = tot + img(m,n);
      totOfsquares = totOfsquares + realpow(img(m,n),2);
   end
end

% INITIALIZE FIRST ROW WINDOW'S RESULTS
Oldtot=tot;
OldtotOfsquares = totOfsquares;

% COMPUTE LOCAL STANDARD DEVIATION
sdev = realpow( (totOfsquares - tot^2/winsize2)/(winsize2-1), 1/2 );

% COMPUTE FIRST LOCAL STANDARD DEVIATION
outimg(1,1)=sdev;

% COMPUTE LOCAL STANDARD DEVIATION FOR REST OF IMG

for m=1:numrows
   mb=m+bufr;
   tot=Oldtot;
   totOfsquares=OldtotOfsquares;
   for n=2:numcols
      nb = n+bufr;
      for index=-bufr:bufr
         oldpix = img(mb+index,nb-bufr-1);
         newpix = img(mb+index,nb+bufr);
         tot = tot -  oldpix + newpix;
         totOfsquares = totOfsquares - realpow(oldpix,2) +  realpow(newpix,2);
      end

      % FOR COLUMNS 2:NUMCOLS, COMPUTE LOCAL STANDARD DEVIATION

      sdev = realpow( (totOfsquares - tot^2/winsize2)/(winsize2-1), 1/2 );

      outimg(m,n)=sdev;

   end

   % SINCE AT LAST COL, THEN FOR NEXT ROW, COLUMN 1,
   % COMPUTE LOCAL STANDARD DEVIATION

   if m~=numrows
      nb = bufr + 1;
      tot=Oldtot;
      totOfsquares=OldtotOfsquares;
      for index=-bufr:bufr
         oldpix = img(mb-bufr,nb+index);
         newpix = img(mb+bufr+1,nb+index);
         tot = tot -  oldpix + newpix;
```

```
        totOfsquares = totOfsquares - realpow(oldpix,2) +  realpow(newpix,2);
      end

      Oldtot = tot;
      OldtotOfsquares = totOfsquares;

      sdev = realpow( (totOfsquares - tot^2/winsize2)/(winsize2-1), 1/2 );

      outimg(m+1,1)=sdev;
   end

end

outimg=MAXGRAY.*erescale(outimg);
```

```
::::::::::::::
evsame.m
::::::::::::::
function vsame=evsame(v,value);
% EVSAME Used to identify elements common to both vectors
% vsame=evsame(v,value);
% by: Capt. Edward M. Ochoa

vsame=zeros(size(v));

for i=1:length(v)
   for j=1:length(value)
     vsame(i)=(v(i)==value(j)) | vsame(i);
   end
end
```

```
::::::::::::::
myprint.m
::::::::::::::
function myprint(figstart,figend,hardcopyTF,createfileTF)
% MYPRINT A handy utility to print figures
% myprint(figstart,figend,hardcopyTF,createfileTF)
%
% by: Capt. Edward M. Ochoa, GEO-96D

for k=figstart:figend
   if (k<10), fignoeps=['fig0' int2str(k) '.eps'];
      else fignoeps=['fig' int2str(k) '.eps']; end
   eval(['figure(' int2str(k) ')'])
   eval(['print -deps ' fignoeps])
   if hardcopyTF, eval(['!qmslpr ' fignoeps]); end
   if ~createfileTF, eval(['!rm ' fignoeps]); end
end
```

```
::::::::::::::
showmam.m
::::::::::::::
function img=showmam(IMGNAMEEXT,displayTF);
% SHOWMAM Used to load/display mammo image
% img=showmam(IMGNAMEEXT,displayTF);
% by: Capt. Edward M. Ochoa, GEO-96D

% IMG INFO
[IMGNAME,EXT,NAME,IMGNUMROWS]=esepname(IMGNAMEEXT,1);
ProcMamTF=(sum((1*EXT==ones(size(EXT,1),1)*'mam')')==3)';

% DIR, EXT, AND IMG DEPTH INFO
if ProcMamTF
   SOURCEDIR='/home/pinna1/bdata/wpafbh/';
else
```

```
   SOURCEDIR='/home/pinna1/bdata/wpafbh/ROIs/';
end

NUMBITS=12;

% LOAD IMG
img=GetMammo([SOURCEDIR IMGNAME '.' EXT ],IMGNUMROWS,0);

if displayTF
   % DISPLAY IMG
   MAXGRAY=realpow(2,NUMBITS)-1;
   imshow((MAXGRAY-1)*erescale(img),gray(MAXGRAY))
   title(sprintf('%s',NAME))
   orient tall
end
```

```
:::::::::::::::
showroitruth.m
:::::::::::::::
function showroitruth(ROINAME,CLASS);
% SHOWROITRUTH Used to display pathology truth box given class
% showroitruth(ROINAME,CLASS);
% by Capt. Edward M. Ochoa, GEO-96D

%SOURCEDIR='/home/hawkeye3/96d/eochoa/Thesis/MC_code/MAMTRU/';
SOURCEDIR='/home/pinna1/bdata/wpafbh/ROIs/';
TRUTHFULLFILENAME=[SOURCEDIR ROINAME '.tru'];
eval(sprintf('load %s;',TRUTHFULLFILENAME))
eval(sprintf('truth=%s;',ROINAME));
showtruth(truth,CLASS)
```

```
:::::::::::::::
showtruth.m
:::::::::::::::
function notruthTF=showtruth(truth,class)
% SHOWTRUTH Used to display pathology truth box given class of interest
% notruthTF=showtruth(truth,class)
% by: Capt. Edward M. Ochoa, GEO-96D

if truth(1,1)~=-1

   hold on

   targetindex=find(truth(:,5)==class);

   for index=1:length(targetindex);
      target=truth(targetindex(index),1:4);
      xstart=target(2);
      ystart=target(1);
      xend  =target(4);
      yend  =target(3);

      xplot=[xstart xstart xend xend xstart] - .5;
      yplot=[ystart yend yend ystart ystart] - .5;
      h1=plot(yplot,xplot,'k');
      set(h1,'LineWidth',2);
      if index==1
         hold on
      end
      plot(yplot,xplot,'b')

   end

end
```

```
::::::::::::::
textsc.m
::::::::::::::
function H = textsc(x,y,txt);
% TEXTSC places text in screen coordinates and places
% a title at the top of the figure.
%
% H = TEXTSC(X,Y,TXT) places the text string, TXT
% at the normalized coordinates X and Y.  H is the
% handle to the text object.
%
% H = TEXTSC(TXT,'title') places a title at the top
% of the figure window.  This is useful when you
% want to place a single title above multiple
% subplots.
%
% TEXTSC creates an invisible AXES which occupies
% the entire FIGURE.  The units of the AXES are
% normalized (range from 0 to 1).  TEXTSC checks
% all the children of the current FIGURE to determine
% if an AXES exist which meets these criteria already
% exist.  If one does, then it places the text relative
% to that AXES.
%
% NOTE:  Requires MATLAB 4.2 or above.


% Written by John L. Galenski III
% All Rights Reserved  January 21, 1994
% LDM031695jlg

% Basic error checking
if nargin < 2
  error('TEXTSC requires at least 2 inputs')
end

% Check to see if AXES already exist
ch = get(gcf,'Children');
ax = findobj(gcf,'Type','axes','Tag','TEXTSC');
if isempty(ax)
  ax = axes('Units','Normal','Position',[0 0 1 1], ...
            'Visible','Off','Tag','TEXTSC');
else
  axes(ax)
end

% Place the text
if nargin == 2 & isstr(x) & strcmp(lower(y),'title')  % Subplot title
  txt = x;
  x = .5;
  tmp = text('Units','normal','String','tmp','Position',[0 0 0]);
  ext = get(tmp,'Extent');
  delete(tmp)
  H = ext(4);
  y = 1 - .60*H;
end
h = text(x,y,txt,'VerticalAlignment','Middle', ...
         'HorizontalAlignment','Center');

% Make the original AXES current
if ~isempty(ch)
  set(gcf,'CurrentAxes',ch(1))
end

% Check for output
```

```
if nargout == 1
  H = h;
end

:::::::::::::::
findcluster.c
MEX file written by Capt. Amy Magnus.
To compile on Solaris, modified by Capt. Edward M. Ochoa
:::::::::::::::
#include  <stdio.h>
#include  <stdlib.h>
#include  <math.h>
#include "mex.h"

typedef struct {
    int rows;
    int cols;
    double *ptr;
               } matlab_MATRIX;

static
#ifdef __STDC__
void color_cluster(
int x,
int y,
long *color,
matlab_MATRIX Mask
)
#else
color_cluster(x,y,color,Mask)
  int x, y;
  long *color;
  matlab_MATRIX Mask;
#endif
{
  int i,j, xi,yj;

  Mask.ptr[x*Mask.rows +y] = (double) *color;

  for (i=-1; i<2; i++){
   xi = x + i;
   if ((xi >= 0)&&(xi < Mask.cols)){
      for (j=-1; j<2; j++){
       yj = y + j;
       if ((yj >= 0)&&(yj < Mask.rows)){

         if ( Mask.ptr[xi*Mask.rows +yj] == -1 ){
  color_cluster(xi,yj,color,Mask);
         }

       }
      }
   }
  }
  return;
}

static
#ifdef __STDC__
void find_clusters(
matlab_MATRIX Mask,
matlab_MATRIX IN,
double *threshold
)
#else
```

```
find_clusters(Mask,IN,threshold)
  matlab_MATRIX  IN,Mask;
  double *threshold;
#endif
{
  long *color,zero;
  int x, y;

  for (x=0; x<Mask.cols; x++){
    for (y=0; y<Mask.rows; y++){
      Mask.ptr[x*Mask.rows +y] = IN.ptr[x*IN.rows +y] >= *threshold ? -1 : 0;
    }
  }

/*initialize color*/

  zero = 0;
  color = &zero;

/* find the first pixel of cluster
   increment color
   call recursive "color_cluster" */

  for (x=0; x<Mask.cols; x++){
    for (y=0; y<Mask.rows; y++){
      if (Mask.ptr[x*Mask.rows +y] == -1){
*color += 1;
color_cluster(x,y,color,Mask);
      }
    }
  }
  return;
}

#ifdef __STDC__
void mexFunction(
        int             nlhs,
        Matrix  *plhs[],
        int             nrhs,
        Matrix  *prhs[]
        )
#else
mexFunction(nlhs, plhs, nrhs, prhs)
int nlhs, nrhs;
Matrix *plhs[], *prhs[];
#endif
{
  matlab_MATRIX  IN, OUT;
  double *threshold;
  double default_threshold;

if ((nrhs < 1) || (nrhs > 2)){
mexErrMsgTxt("find_cluster requires one or two input arguments.");
} else if (nlhs > 1) {
mexErrMsgTxt("find_cluster has one output argument.");
}

IN.rows = mxGetM(prhs[0]);
IN.cols = mxGetN(prhs[0]);
IN.ptr = mxGetPr(prhs[0]);

if (nrhs == 1){
default_threshold = 1;
threshold = &default_threshold;
} else {
```

```
threshold = mxGetPr(prhs[1]);
}
OUT.rows = IN.rows;
OUT.cols = IN.cols;
plhs[0] = mxCreateFull(OUT.rows,OUT.cols, REAL);
OUT.ptr = mxGetPr(plhs[0]);

find_clusters(OUT,IN,threshold);
return;
}
```

Bibliography

1. Abrahamson, Shane Lee. *Pulse Coupled Neural Networks for the Segmentation of Magnetic Resonance Brain Images.* MS thesis, AFIT/GCS/ENG/96, Graduate School of Engineering, Air Force Institute of Technology (AETC), Wright-Patterson AFB OH, 1996.
2. Adler, Dorit D. "Mammographic Evaluation of Masses," *RSNA Categorical Course in Breast Imaging*, 107–116 (1995).
3. Bovik, Alan Conrad, et al. "The effect of median filtering on edge estimation and detection," *IEEE Trans. Pattern Anal. Machine Intell.*, 181–194 (1987).
4. Bramlette, Mark F. "Initialization, Mutation, and Selection Methods in Genetic Algorithms for Function Optimization." *Fourth Int. Conf. Genetic Algorithms.* 100–107. 1991.
5. Bronson, Richard. *Theory and Problems of Operations Research.* Schaum's Outline, McGraw-Hill Book Company, 1982.
6. Caldwell, Craig and Victor S. Johnston. "Tracking a Criminal Suspect through "Face-Space" with a Genetic Algorithm." *Fourth Int. Conf. Genetic Algorithms.* 416–421. 1991.
7. Calloway, David L. "Using a Genetic Algorithm to Design Binary Phase-Only Filters for Pattern Recognition." *Fourth Int. Conf. Genetic Algorithms.* 422–429. 1991.
8. Carlton, Richard R. and Arlene M. Adler. *Principles of Radiographic Imaging* (2nd Edition). Delmar, 1996.
9. Chan, Heang-Ping, et al. "Computer-aided detection of mammographic micro-calcifications: Pattern recognition with an artificial neural network," *Med. Phys.*, *22*(10):1555–1567 (1995).
10. Dauk, Ronald C. *Computer-Aided Detection of Microcalcifications Using Texture Analysis.* MS thesis, AFIT/GEO/ENG/95-D01, Graduate School of Engineering, Air Force Institute of Technology (AETC), Wright-Patterson AFB OH, 1995.
11. De Jong, Kenneth. "Genetic Algorithms: A 10 Year Perspective." *First Int. Conf. Genetic Algorithms and Their Applications.* 169–177. 1985.
12. Dhawan, A. P., et al. "Enhancement of mammographic features by optimal adaptive neighborhood image processing," *IEEE Trans. Med. Imaging*, *5*(1):8–15 (March 1986).
13. Dhawan, A. P. and R. Gordon. "Reply to comments on enhancement of mammographic feature by optimal adaptive neighborhood image processing," *IEEE Trans. Med. Imaging*, *MI-6*(1):82–83 (March 1987).
14. Dhawan, A. P. and Eric Le Royer. "Mammographic feature enhancement by computerized image processing," *Computer Methods and Programs in Biomedicine*, *27*(1):25–35 (1988).
15. D'Orsi, Carl J. "Digital Mammography," *RSNA Categorical Course in Breast Imaging*, 80 (1995).

16. D'Orsi, Carl J. "Use of the American College of Radiology Breast Imaging and Data System," *RSNA Categorical Course in Breast Imaging*, 77–80 (1995).

17. Eisenbies, Christopher L. *Classification of Ultra High Range Resolution Radar Using Decision Boundary Analysis*. MS thesis, AFIT/GE/ENG/94D-07, Graduate School of Engineering, Air Force Institute of Technology (AETC), Wright-Patterson AFB OH, 1994.

18. Ema, Takehiro, et al. "Feature analysis and computer-aided diagnosis in mammography: Reduction of false-positive clustered microcalcifications using local edge-gradient analysis," *Med. Phys.*, *22*(2):161–169 (February 1995).

19. Ernisse, Brian Everett. *Automatic Target Cuer/Recognizer System for Tactical FLIR Images*. MS thesis, AFIT/GE/ENG/96-D, Graduate School of Engineering, Air Force Institute of Technology (AETC), Wright-Patterson AFB OH, 1996.

20. Feig, Stephen A. "Mammographic Evaluation of Calcifications," *RSNA Categorical Course in Breast Imaging*, 93–105 (1995).

21. Giger, Maryellen L. "Computer-Aided Diagnosis," *RSNA Categorical Course in Physics*, 283–298 (1993).

22. Goldberg, David E. "Sizing Populations for Serial and Parallel Genetic Algorithms." *Third Int. Conf. Genetic Algorithms*. 70–79. 1989.

23. Goldberg, David E. "Zen and the Art of Genetic Algorithms." *Third Int. Conf. Genetic Algorithms*. 80–85. 1989.

24. Gonzales, R. C. and R. E. Woods. *Digital Image Processing*. Addison-Wesley, 1992.

25. Grefenstette, John J. and James E. Baker. "How Genetic Algorithms Work: A Critical Look at Implicit Parallelism." *Third Int. Conf. Genetic Algorithms*. 20–27. 1989.

26. Harrup, Georgia K. *ROC Analysis of IR Segmentation Techniques*. MS thesis, AFIT/GE/ENG/94D-15, Graduate School of Engineering, Air Force Institute of Technology (AETC), Wright-Patterson AFB OH, 1994.

27. Hoffmeister, Jeffrey W., "Personal Interviews." Aerospace Physician, Armstrong Laboratory, Wright-Patterson AFB, OH. May 1995 - December 1996.

28. Holland, John H. "Genetic Algorithms," *Scientific American*, 66–72 (July 1992).

29. Houck, Chris, et al. *A Genetic Algorithm for Function Optimization: A Matlab Implementation*. Technical Report, NCSU-IE TR 95-09, 1995. The GAOT may be downloaded via WWW at http://www.eos.ncsu.edu/eos/info/ie589k_info/GAOT.

30. Janikow, Cezary Z. and Zbigniew Michalewicz. "An Experimental Comparison of Binary and Floating Point Representations in Genetic Algorithms." *Fourth Int. Conf. Genetic Algorithms*. 31–36. 1991.

31. Judy, Philip F., et al. "Measuring the Observer Performance of Digital Systems." *Computed Digital Radiography in Clinical Practice* edited by R. E. Green and J. W. Oestmann, 59–69, New York: Thieme Medical Publishers, 1992.

32. Karrsemeijer, N. "Adaptive noise equalization and image analysis in mammography." *13th Int. Conf. Inform. Processing Med. Imag.*. 472–486. 1992.

33. Kocur, Catherine M., et al. "Neural Network Wavelet Feature Selection for Breast Cancer Diagnosis," *IEEE Engineering in Medicine and Biology*, 95–102,108 (May/June 1996).

34. Kopans, Daniel B. "The Etiology of Breast Cancer," *RSNA Categorical Course in Breast Imaging*, 29–37 (1995).

35. Lewis, Brian J. *Cecil Textbook of Medicine*. W.B. Saunders Co., 1988.

36. Marr, David. *Vision*. San Francisco: W. H. Freeman and Co., 1982.

37. Martin, Curtis E. *Non-Parametric Bayes Error Estimation For UHRR Target Identification*. MS thesis, AFIT/GE/ENG/93D-26, Graduate School of Engineering, Air Force Institute of Technology (AETC), Wright-Patterson AFB OH, 1993.

38. McCandless, Donald A. *Detection of Clustered Microcalcifications Using Wavelets*. MS thesis, AFIT/GE/ENG/95D-17, Graduate School of Engineering, Air Force Institute of Technology (AETC), Wright-Patterson AFB OH, 1995.

39. Metz, Charles E. "ROC Methodology in Radiologic Imaging," *Investigative Radiology*, *21*(9):720–733 (September 1986).

40. Metz, Charles E. "Some Practical Issues of Experimental Design and Data Analysis in Radiological ROC Studies," *Investigative Radiology*, *24*(3):234–245 (March 1989).

41. Michalewicz, Z. *GeneticAlgorithms + DataStructures = EvolutionPrograms*. AI Series, New York: Springer-Verlag, 1994.

42. Morrow, William M., et al. "Region-based contrast enhancement of mammograms," *IEEE Transactions on Medical Imaging*, *11*(3):392–406 (September 1992).

43. Niblack, Wayne. *An introduction to digital image processing*. Prentice-Hall International (UK) Ltd., 1986.

44. Nishikawa, R. M., et al. "Computer-aided detection of clustered microcalcifications on digital mammograms," *Medical and Biological Engineering and Computing*, *33*(2):174–178 (March 1995).

45. Nishikawa, Robert M., et al. "Effect of case selection on the performance of computer-aided detection schemes," *Medical Physics*, *21*(2):265–269 (February 1994).

46. Nishikawa, Robert M., et al. "Computer-aided detection of clustered microcalcifications: An improved method for grouping detected signals," *Medical Physics*, *20*(6):1661–1666 (November/December 1993).

47. Parker, Sheryl L., et al. "Cancer Statistics," *CA*, *46*(1):5–27 (January 1996).

48. Polakowski, William E. *Computer-Aided Diagnosis of Mammographic Masses*. MS thesis, AFIT/GEO/ENS/95D-02, Graduate School of Engineering, Air Force Institute of Technology (AETC), Wright-Patterson AFB OH, 1995.

49. Raff, Ulrich. "Visual data formatting." *The perception of visual information* edited by William R. Hendee and Peter N. T. Wells, 160–201, New York: Springer-Verlag, 1993.

50. Richardson, Jon T., et al. "Some Guidelines for Genetic Algorithms with Penalty Functions." *Third Int. Conf. Genetic Algorithms*. 191–197. 1989.

51. Rogers, S. K., et al. "Neural Networks for Automatic Target Recognition," *IEEE Transactions on Neural Networks*, *8*(7/8):1153–1184 (1995).

52. Rogers, Steven K., et al. "Artificial Neural Networks for Automatic Object Recognition," *SPIE Institute Series on Automatic Object Recognition*, 231–243 (April 1990).

53. Russ, John C. *The Image Processing Handbook*. CRC Press, 1992.

54. Schaffer, J. David, et al. "A Study of Control Parameters Affecting Online Performance of Genetic Algorithms for Function Optimization." *Third Int. Conf. Genetic Algorithms*. 51–60. 1989.

55. Schmidt, Robert A., et al. "Computer-aided diagnosis in Mammography," *RSNA Categorical Course in Breast Imaging*, 199–208 (1995).

56. Shaw de Paredes, Ellen. "Radiographic Breast Anatomy: Radiologic Signs of Breast Cancer," *RSNA Categorical Course in Physics*, 35–45 (1994).

57. Sickles, Edward A. "Auditing Your Practice," *RSNA Categorical Course in Breast Imaging*, 81–91 (1995).

58. Silverberg, E. and J. Lubera. "Cancer Statistics," *Cancer*, *39* (1987).

59. Smiley, Steven E. *Image Segmentation Using Affine Wavelets*. MS thesis, AFIT/EN/ENG/91D-50, Graduate School of Engineering, Air Force Institute of Technology (AETC), Wright-Patterson AFB OH, 1991.

60. Smith, Robert A. "The Epidemiology of Breast Cancer," *RSNA Categorical Course in Breast Imaging*, 7–20 (1995).

61. Sonka, Milan, et al. *Image processing, analysis and machine vision*. University Press, 1993.

62. Strang, Gilbert. *Linear Algebra and its Applications* (Third Edition). San Diego: Harcourt Brace Jovanovich, 1988.

63. Strickland, R. N. and H. I. Hahn. "Detection of microcalcifications using wavelets." *Digital Mammography: Proceedings of the 2nd International Workshop on Digital Mammography*, edited by Alastair G. Gale, et al. 79–88. York, UK: Elsevier, 10-12 July 1994.

64. Strickland, Robin N. and Hee Il Hahn. "Wavelet transforms for detecting microcalcifications in mammograms," *IEEE Transactions on Medical Imaging*, *15*(2):218–229 (April 1996).

65. Swett, Henry A., et al. "Computer Vision and Decision Support." *The perception of visual information* edited by William R. Hendee and Peter N. T. Wells, 272–315, New York: Springer-Verlag, 1993.

66. Tahoces, P. G., et al. "Enhancement of chest and breast radiographs by automatic spatial filtering," *IEEE Transactions on Medical Imaging*, *10*(3):330–335 (September 1991).

67. Yaffe, Martin J. "Digital Mammography," *RSNA Categorical Course in Physics*, 275–286 (1994).

68. Yoshida, Hiroyuki, et al. "Automated detection of clustered microcalcifications in digital mammograms using wavelet transform techniques." *Image Processing2167*. Medical Imaging, edited by Murray H. Loew. 868–886. Newport Beach, CA: SPIE, February 1994.

Vita

Captain Ochoa obtained a BSEE from the University of Arizona and a commission as a 2nd Lt in the United States Air Force in September of 1992. From January 1993 through May 1995 he was a Senior Imagery Exploitation Engineer in the Imagery Systems Division, National Air Intelligence Center, Wright-Patterson Air Force Base (WPAFB), OH. Currently, he has been assigned to the Air Force Institute of Technology (AFIT), WPAFB, OH. At AFIT he has been pursuing an MS in Electro-Optics, specializing in pattern recognition. As of January 1997 he may be reached at the Air Force Information Warfare Center, Kelly AFB, TX.

Permanent address: 5290 Cobb Dr
Dayton, Ohio 45431

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 1996	Master's Thesis

4. TITLE AND SUBTITLE

CLUSTERED MICROCALCIFICATION DETECTION USING OPTIMIZED DIFFERENCE OF GAUSSIANS

5. FUNDING NUMBERS

6. AUTHOR(S)

Edward M. Ochoa
Captain, USAF

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology, WPAFB OH 45433-6583

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT/GEO/ENG/96D-13

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Maj. Jeffrey W. Hoffmeister, MD
2255 H Street
Armstrong Laboratory / CFAHV
Wright-Patterson AFB, OH 45433-7022

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)*

Clustered microcalcifications are one of the earliest indicators of breast cancer, and are detected only by mammography; 30 to 50 percent of nonpalpable cancers are mammographically visible on the basis of microcalcifications alone. Furthermore, for early breast cancers, screening studies suggest that 70 to 90 percent were detected based on microcalcifications alone. This research proposes the following methodology for clustered microcalcification detection. First, preprocess the digitized film mammogram to reduce digitization noise. Second, spatially filter the image with a difference of Gaussians (DoG) kernel. To detect potential microcalcifications, segment the filtered image using global and local thresholding. Next, cluster and index these detections into regions of interest (ROIs). Identify ROIs on the digitized image (or hardcopy printout) for final radiologic diagnosis. Finally, to improve detection rates, globally optimize detection parameters using a genetic algorithm (GA), then locally optimize using the simplex method.

The database of 56 digitized (12 bit, 100 μm) full-breast (20×10 cm^2) film mammograms contained 63 biopsy-truthed clustered microcalcification ROIs over 28 cases. This technique demonstrated a true positive (TP) case detection rate of 96.4 percent (27/28), and TP ROI (54/63) and TP image (48/56) detection rates of 85.7 percent with 5.75 false positives (FPs) per full-breast image.

14. SUBJECT TERMS

Pattern Recognition, Breast Cancer, Clustered Microcalcifications, Medical Imaging, Genetic Algorithms, Difference of Gaussians

15. NUMBER OF PAGES

97

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	UL

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Block 12b. Distribution Code.

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

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

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

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

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

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

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