12-31-2004

Automated evaluation of radiodensities in a digitized mammogram database using local contrast estimation

Min Taek Kim
Rowan University

Automated evaluation of radiodensities in a digitized mammogram database using local contrast estimation

by

Min Taek Kim

A Thesis Submitted to the

Graduate Faculty in Partial Fulfillment of the

Requirements for the Degree of

MASTER OF SCIENCE

Department: Electrical and Computer Engineering
Major: Engineering (Electrical Engineering)

Approved: Members of the Committee

_____ _____

In Charge of Major Work

_____ _____

For the Major Department

For the College

Rowan University
Glassboro, New Jersey
2004

ABSTRACT

Mammographic radiodensity is one of the strongest risk factors for developing breast cancer and there exists an urgent need to develop automated methods for predicting this marker. Previous attempts for automatically identifying and quantifying radiodense tissue in digitized mammograms have fallen short of the ideal. Many algorithms require significant heuristic parameters to be evaluated and set for predicting radiodensity. Many others have not demonstrated the efficacy of their techniques with a sufficient large and diverse patient database. This thesis has attempted to address both of these drawbacks in previous work. Novel automated digital image processing algorithms are proposed that have demonstrated the ability to rapidly sift through digitized mammogram databases for accurately estimating radiodensity. A judicious combination of point-processing, statistical, neural and contrast enhancement techniques have been employed for addressing this formidable problem. The algorithms have been developed and exercised using over 700 mammograms obtained from multiple age and ethnic groups and digitized using more than one type of X-ray digitizer. The automated algorithms developed in this thesis have been validated by comparing the estimation results using 40 of these mammograms with those predicted by a previously established manual segmentation technique. The automated algorithms developed in this thesis show considerable promise to be extremely useful in epidemiological studies when correlating other behavioral and genetic risk factors with mammographic radiodensity.

ACKNOWLEDGEMENTS

I would like to thank the faculty at Rowan University for all the help and support that they have given to me over the years. Over the years they have amazed me with their intelligence, their dedication, and most importantly their kindness and patience. I would especially like to thank my advisor, Dr. Mandayam, for the friendship and advice that he has given during the past years. Because of his guidance, I can be assured that I will look back to my time at Rowan and be proud of what I have accomplished. My special thanks go to Dr. Kadlowec and Dr. Polikar for serving on my defense committee.

I would like to thank Fox Chase Cancer Center for their funding through grants from the American Cancer Society and the American Institute for Cancer Research, Dr. Marilyn Tseng of FCCC for her patience and direction, and Dr. Celia Byrne for providing us with the validation data which has guided our work over the years. I would also like to thank Dan Barrot and Lindsay Burd for their help throughout the research.

I would like to thank my friends and family for all their support. I know I haven't always shown the love and appreciation that you deserve but I could not have made it without you. To my friends Kevin, Joe, and Jon; it's finally over and we made it.

Finally, and most importantly, I would like to thank my wife Johanna. You have given me love, support, and the comforting knowledge that you will always be there for me no matter what I do or how hard life is. Without you, I would not be the person I am today. You are and always will be the greatest treasure of my life. I love you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Cancer results from a mutation that has the potential to cause the uncontrollable growth and spread of unhealthy cells. If cancer does spread by metastasis, it may result in death to the individual. The mutations can be caused by both external stimuli and internal stimuli and may go undetected up to as much as ten years after the occurrence of the stimulus. In America today, cancer is the second leading cause of death, second only to heart disease. About 1.3 million new cancer cases were diagnosed in 2003. This year alone, approximately half a million Americans are expected to die due to cancer related deaths, accounting for more than 1,500 deaths per day [1].

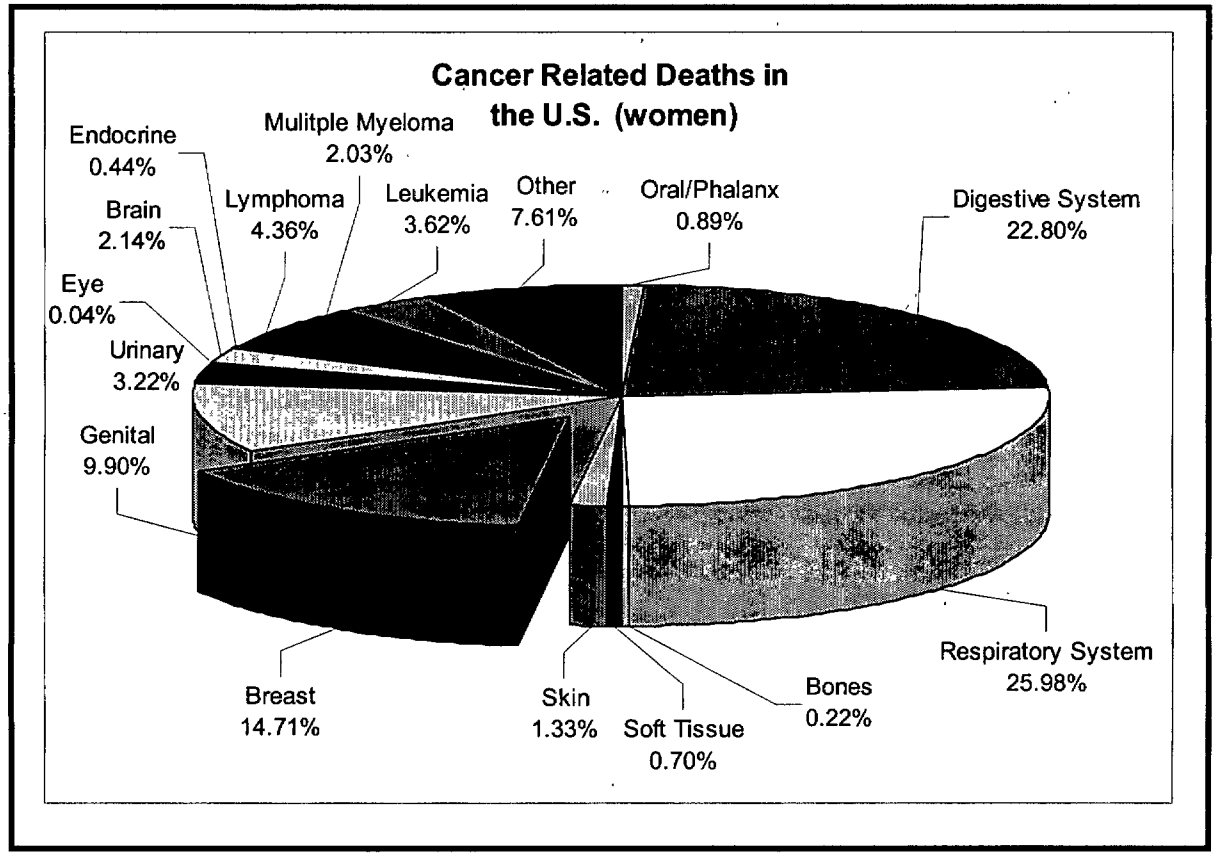


FIGURE 1.1 Cancer related deaths in the U.S.

For women, breast cancer is the third leading cause of death, only surpassed by lung and digestive cancers. It is anticipated that about 40,000 deaths will occur due to breast cancer in the United States this year, accounting for 15% of all cancer related deaths, as shown in Figure 1.1. Though not the leading cause of cancer related deaths among women, breast cancer is the most frequently diagnosed non-skin cancer today. As Figure 1.2 shows, it accounts for 33% of all new cancers. An estimated quarter of a million new cases of invasive breast cancer occurred among women in 2003. In addition to the invasive breast cancer, approximately another 56,000 new cases occurred among women in 2003, of which roughly 85% were ductal carcinoma in situ [1].

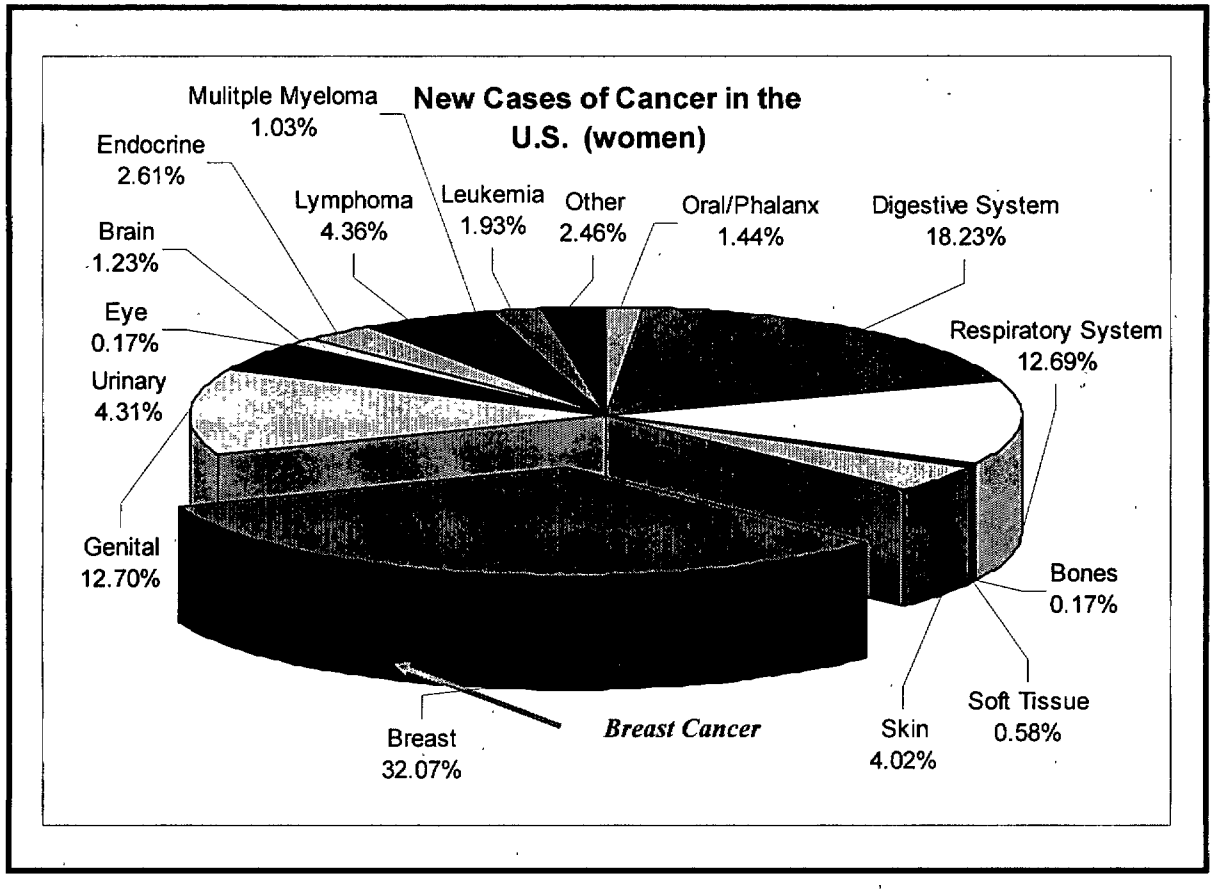


FIGURE 1.2 Estimated new cases of cancer in 2003.

The earliest sign of breast cancer is usually an abnormality that can be seen in mammograms before it is ever felt by the women or health care provider. Because early detection results in a better disease prognosis, mammograms have been shown to save lives and increase treatment options. According to the most recent studies, it has been shown that mortality due to breast cancer has declined 1.4% every year between 1989 and 1995 and 3.2% every year afterwards. These declines can be attributed to the increase of early detections measures such as regular mammography screenings [2].

Even though the exact mechanism of why a patient acquires breast cancer is not well known, there has been much research done to gain knowledge in potential risk factors. Studies have identified risk factors that can help doctors determine whether a woman has a higher probability of acquiring breast cancer. Knowledge of such risk factors can allow doctors to advise patients on how often they should be screened for early signs of breast cancer, making it more likely that any abnormal growth will be detected early. This thesis introduces an algorithm that can be used an important tool for doctors to analyze a patient's risk for acquiring breast cancer.

1.1 Breast Anatomy

The human breast, otherwise known as the mammary gland, is composed of various components. Each breast contains 15-20 sections called lobes that surround the nipple. Inside the lobes, there are 30-80 smaller structures called the lobules. Contained at the end of each lobule are bulbs which ultimately create the milk. These structures are linked together by small tubes called ducts, which carry milk to the nipples. The only muscle tissues located within the breasts are found in the nipple and around the lobules. A layer

of fat, also known as adipose tissue, surrounds the glands and ducts within the breast [3]. Figure 1.3 shows an illustration of the anatomy of the breast.

There are many factors that contribute to the actual size and density of the breast. Studies have shown that genetics, nutrition, and conditioning factors play a major role in the size, composition, and shape of the breast. In normal breast development, the

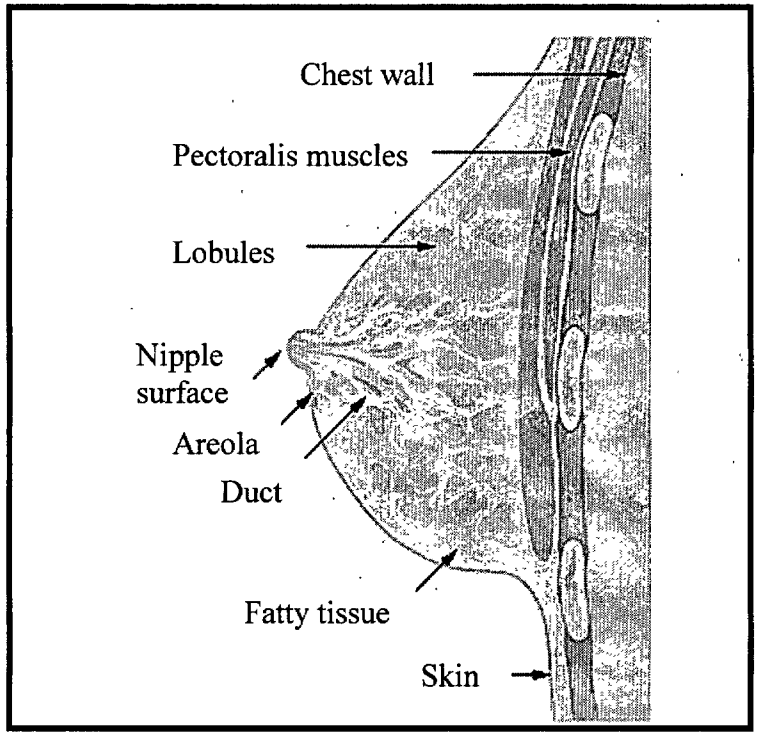


FIGURE 1.3 Breast anatomy.

effects of the major reproductive hormones estrogen, progesterone, and prolactin play the most important role [4].

The size and composition of the breast also changes throughout a women's life. The changes can occur due to the cyclic changes of menses, altered physiology and anatomy during pregnancy and lactation, and the aging factor. During the menstrual cycle, it has been demonstrated that the breast changes due to the ovarian hormone levels. The breast tissue has been found to be less dense during the follicular phase than in the

luteal phase among premenopausal women. During pregnancy and lactation, the breast tissue becomes denser due to the increased amount of glandular tissue contained within the breast. During menopause, the glandular components of the breast are replaced by fat and connective tissue, which can be attributed to the decline in ovarian functions [5].

1.2 Breast Density and Parenchyma

The density of a breast can be defined by the amount of its ducts, lobules, and interlobular fibrous tissue. There are many implications to the density of the breast. First, there are important clinical repercussions to dense breasts. Breasts that contain mostly adipose tissue are very radiolucent in nature and therefore easily reveal the presence of small tumors. Dense breasts are very difficult to evaluate with mammography. For the smaller tumors located within the breast, there is poor contrast in

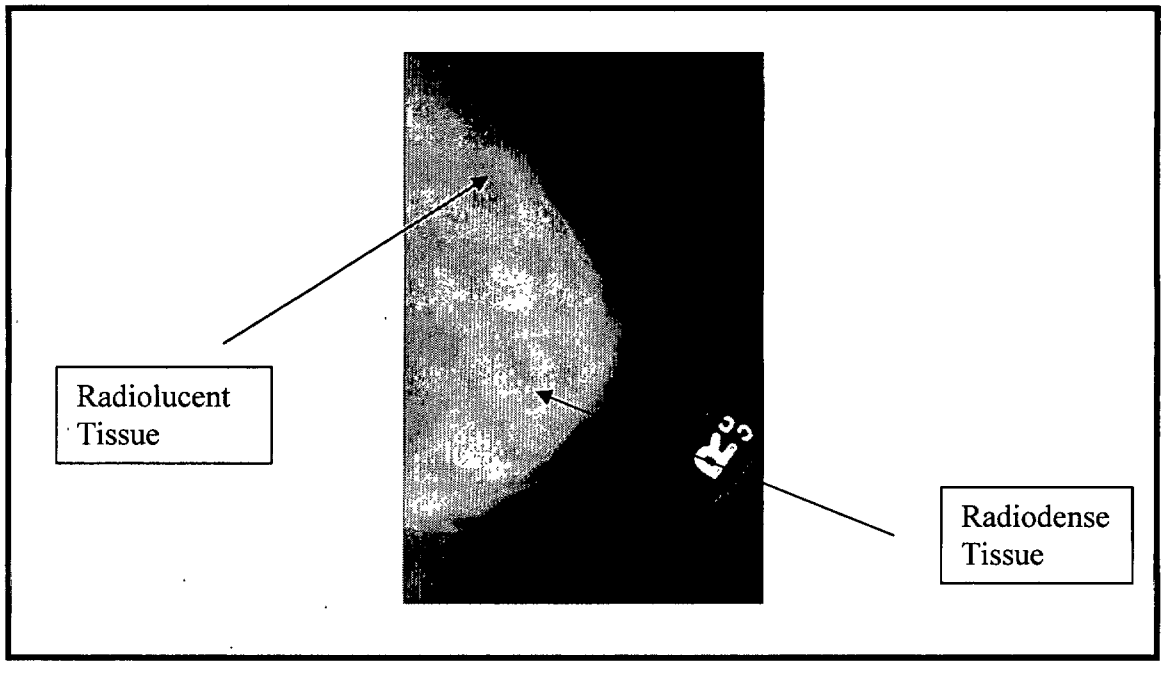


FIGURE 1.4 Mammogram showing the two different regions located within the breast, radiodense and radiolucent.

comparison to a radiodense mammogram. Figure 1.4 shows an example of radiodense and radiolucent tissue. The second and most important implication of high radiodense tissue in a breast is its link to an increased risk for breast cancer [6, 7]. Studies have shown that breast density is one of the strongest predictors of breast cancer. This can simply be attributed to the fact that there is more glandular tissue within a radiodense tissue, which in turn contains more cells that have the potential to turn into cancerous cells.

The mammographic parenchymal patterns have been shown to consistently change based on breast cancer risk factors such as age, menopause, social status, parity, and body size. It has also been shown that parenchymal density is also related to alcohol consumption, nutritional variables, family history of breast cancer, and race. Because it has been shown that radiodensity is one of the highest risk markers for breast cancer, understanding how it can be changed can play a vital role in cancer prevention. In one study, researchers investigated nearly 1000 pairs of monozygotic (identical) and dizygotic/fraternal (non-identical) twins in North America and Australia. Radiologists who were blinded to the patient information analyzed the mammograms for radiodensity. The percentage of dense breast tissue was correlated twice as strongly among monozygotic twin pairs as among dizygotic twin pairs, showing that radiodensity can be traced to a heritable gene. Based on this study performed by Boyd, *et al* it was determined that "the percentage of dense tissue on mammography at a given age has high heritability. Because mammographic density is associated with an increased risk of breast cancer, finding the genes responsible for this phenotype could be important for understanding the causes of the disease [15]."

1.3 Mammography

Mammograms are one of the most important tools for a doctor to evaluate, diagnose, and follow breast cancer. The recent decline in cancer related deaths can be attributed to the increase in mammography screenings, as well as improved cancer treatments. There is no evidence that suggests other imaging methods have reduced the mortality rate of breast cancer like mammography.

It is estimated that there are over 50 million mammograms are performed annually. The US Food and Drug Administration reports that mammography can find 85 to 90 percent of breast cancers in women over the age of 50 and can discover a lump up to two years before it can even be felt [8]. According to the American Cancer Society, when the breast cancer is confined to the breast, the five-year survival rate is close to 100% [9]. The early detection of the cancerous cells helps to reduce the severity of treatment in addition to minimizing the amount of pain and discomfort the women experiences. It has been stated that women over the age of 40 should have yearly mammograms as well as clinical and self examinations. These guidelines were developed because the occurrence of breast cancer for women between the ages of 20-29 is only about 0.3%, and about 77% percent of the women who have breast cancer are over the age of 50 [1].

1.4 Mammography Procedure

The Mammography Quality Standards Act (MQSA) passed by Congress in 1992 and administered by the Food and Drug Administration (FDA) sets certain guidelines for the certified technician when performing a screening. The breast must first be placed

between the two compression plates. The MQSA requires the compression of the breast to be a certain thickness. Unfortunately, this standard cannot always be met due to differences in the size and shapes of breast.

A conventional mammography process is shown in Figure 1.5. The breast is placed between two plates and the X-ray image is a result of how much radiation passes through the breast. Different types of tissue have different attenuation rates; as a result, abnormally dense tissue can be detected. Radiodense tissue has a high attenuation rate, therefore it will absorb most of the X-rays and will show up as bright spots in the X-ray after it has been inverted. Abnormalities such as microcalcifications, masses, and cysts will also show up as very bright spots and can be detected easily as long as the rest of the

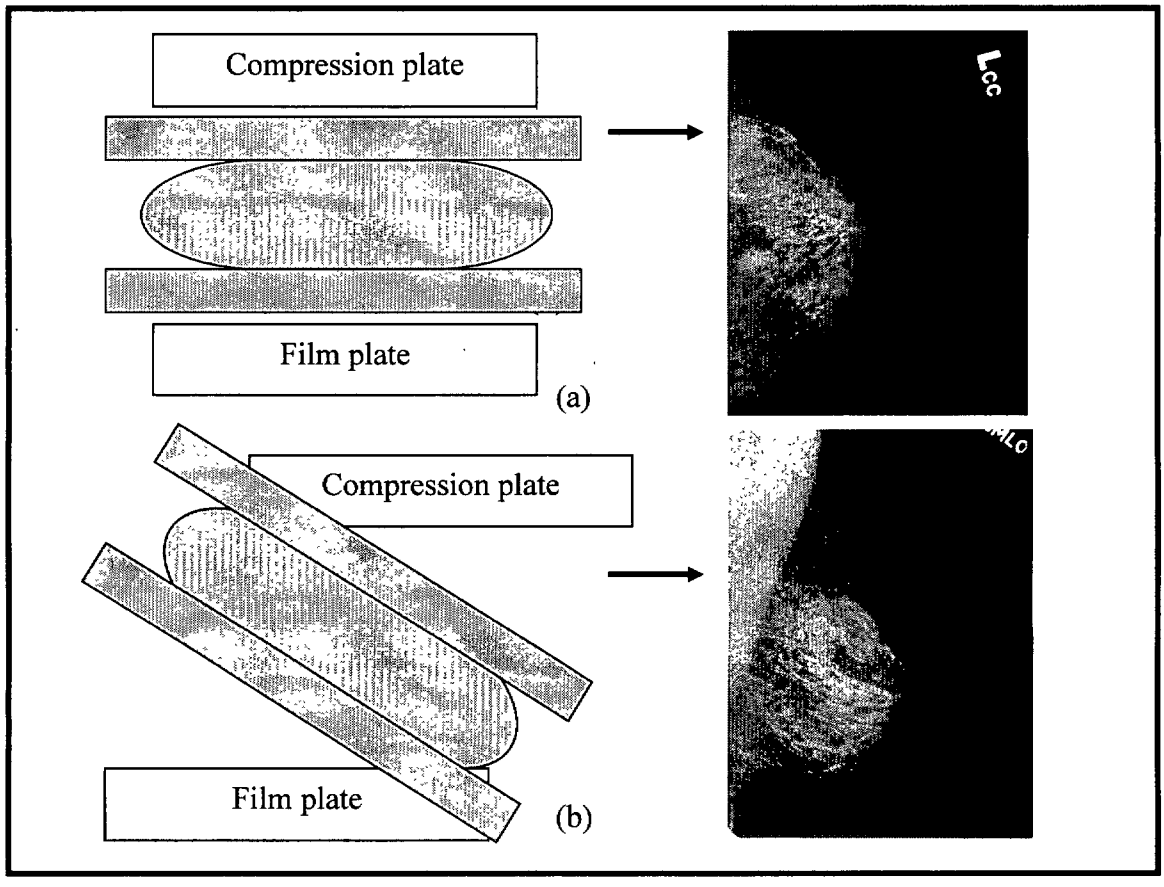
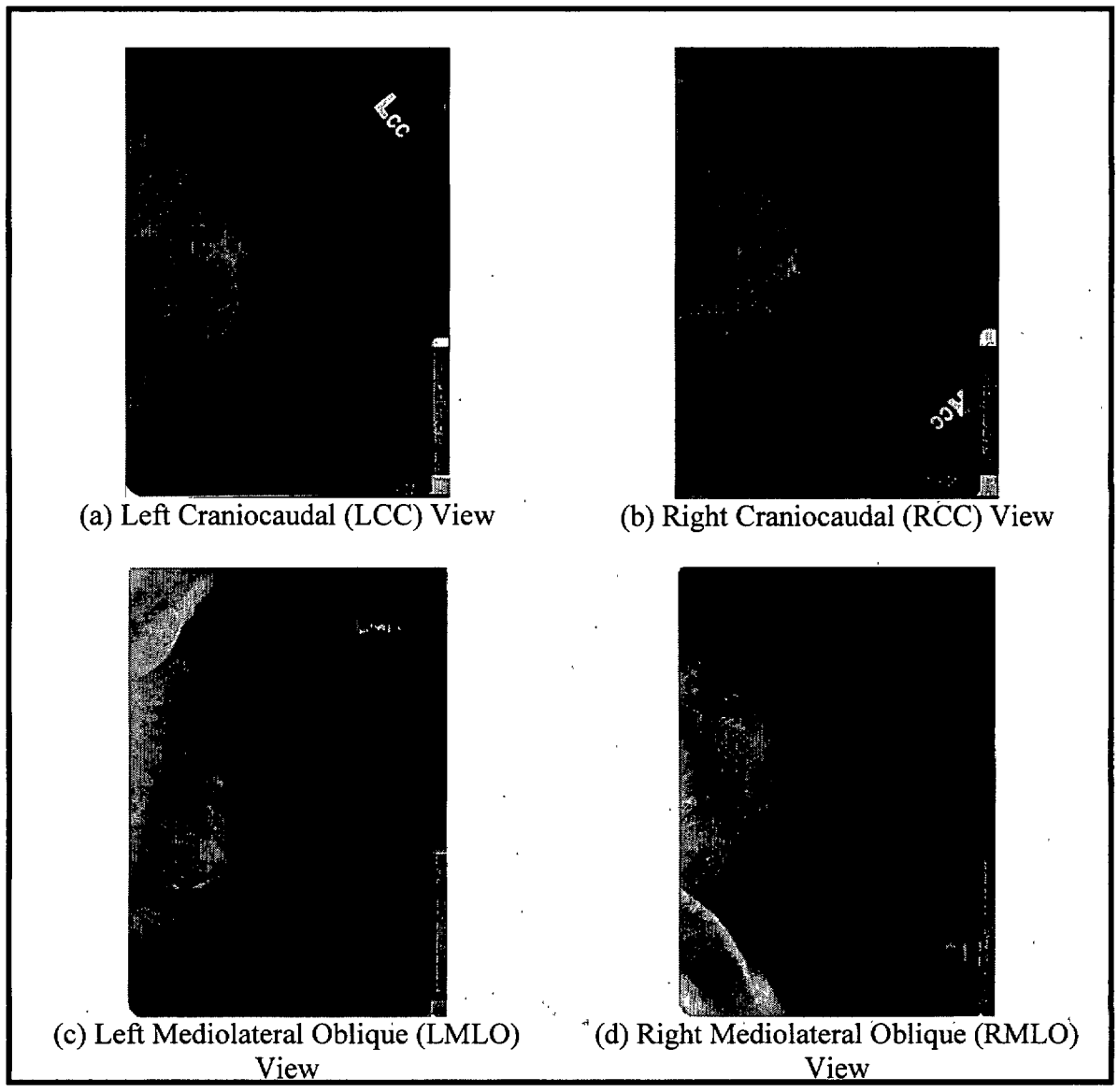


FIGURE 1.5 Mammogram procedure (a) craniocaudal (b) mediolateral oblique.

tissue is relatively radiolucent. This essentially creates problems for women with high radiodense tissue. The abnormalities may blend in with the high radiodense tissue and make it very difficult for analysis due to the low contrast. The adipose regions of the



(a) Left Craniocaudal (LCC) View

(b) Right Craniocaudal (RCC) View

(c) Left Mediolateral Oblique (LMLO) View

(d) Right Mediolateral Oblique (RMLO) View

FIGURE 1.6 Typical mammogram X-ray film set.

breast show up as dark regions in the mammograms. The parenchyma differences seen among women in mammograms are due to differences in the relative amounts of fat, ductal, and lobular tissue located within in the breast.

Mammograms are usually taken from one of two views for each breast; two for craniocaudal (CC) and 2 for mediolateral oblique (MLO) as shown in Figure 1.6. The mediolateral oblique view is taken from an angled projection. This method is sometimes preferred because it allows the technician to place more of the breast onto the film. Unfortunately, for some automated techniques, the inclusion of the chest muscle makes it very difficult for boundary estimation. The technician attempts to make sure that the compression of the breast is uniform throughout.

The craniocaudal view images the breast from above. This method does not have as much tissue as the MLO methods, but excludes the chest muscle from the image while including most of the breast parenchyma. Uniform compression in attempted throughout as with all procedures.

Even though the scanning procedure has evolved over the years allowing for semi-automation, it still requires the supervision of a trained technician for the images to develop properly. Most images that have problems occur due to incorrect positioning and compression. Radiologists are typically concerned with the presence of immediate dangers to the patient, malignancy and other abnormalities. But, as stated earlier, radiodensity has become an important part of any mammogram evaluation because of its correlation with high breast cancer risk. A highly dense breast is also a concern because of the complexity it adds due to the relative contrast between the abnormal regions and the radiodense regions.

As is the common theme throughout this thesis, variability plays a big role in the consistency of the image that is produced. A study has shown that there is inconsistency amongst radiologists in mammography screening [10]. The amount of inconsistency

varies between radiologists and depends on the study that uses the mammograms. There are many factors that contribute to the varying results and cause problems when a study spans multiple regions and uses different radiologists.

1.5 Risk Factors for Breast Cancer

The American Cancer Society characterizes risk factors into three broad categories. These three categories include: risk factors that cannot be changed, lifestyle-related risk factors, and risk factors with uncertain, controversial, or unproven effects on breast cancer risk. Risk factors that cannot be changed include gender, aging, genetic risk

TABLE 1.1 Breast cancer risk factors with their associated relative risk.

Risk Factor	High-Risk Group	Low-Risk Group	Relative risk
Age	Old	Young	> 4.0
Country of birth	North America, Northern Europe	Asia, Africa	> 4.0
Socioeconomic status	High	Low	2.0 – 4.0
Marital Status	Never married	Ever married	1.1 – 1.9
Place of residence	Urban	Rural	1.1-1.9
Place of residence	Northern US	Southern US	1.1-1.9
Race ≥ 45 years	White	Black	1.1-1.9
< 40 years	Black	White	1.1-1.9
Nulliparity	Yes	No	1.1-1.9
Age at first full-term pregnancy	≥ 30 years	< 20 years	2.0-4.0
Age at menopause	Late	Early	1.1-1.9
Weight, postmenopausal women	Heavy	Thin	1.1-1.9
Any first-degree relative with history of breast cancer	Yes	No	2.0-4.0
Mother and sister with history of breast cancer	Yes	No	> 4.0
Mammographic parenchymal patterns	Dysplastic	Normal	2.0-4.0

factors, race, and family history of breast cancer. Lifestyle-related factors like children, breast-feeding, alcohol, obesity, and physical inactivity are all risks factors that can be changed. Factors with unproven effects on breast cancer risk include antiperspirants,

smoking, and breast implants. In some studies, it has been shown that smoking can actually reduce the risk of breast cancer; but of course that increases a patient's risk of lung cancer, which is the number one cancer related death in America. Table 1.1 describes some breast cancer risk factors with their associated relative risk [11].

Many studies have shown that breast density may be one of the strongest markers for breast cancer risk analysis. The studies have shown that a strong correlation exists between breast parenchymal density on mammograms and breast cancer risk [12, 13, 14].

1.6 Levels of Breast Cancer and Survival Rates

Because of the importance of mammograms and other screening procedures, it is vital that the different stages of cancer and risks associated are discussed. The cancer in the breast is usually categorized through the use of stage definitions. Staging is the process by which a physician can assess the size and location of the cancer. The stage of the cancer plays an important role in determining the type of treatment needed. Most importantly, the stage of cancer determines the survival rate of the patient after treatment. If the cancer is detected before it has spread past the breast, the five year survival rate of the patient is 100%. The American Joint Committee on Cancer places the cancer in a letter category using the Tumor–Nodes–Metastasis (TNM) classifier system.

The first stage, stage 0, can be designated as in situ, the term in situ meaning in place. In this type of cancer, it is contained within the breast ductal system and has not spread beyond that. Up to twenty percent of diagnosed breast cancers are of this type. There are two types of cancers in this stage, lobular carcinoma in situ (LCIS) and ductal carcinoma in situ (DCIS). Even though the structure LCIS is similar to cancer, it does

not behave as a cancer and therefore is usually not treated as one. LCIS is usually not considered a malignancy by most doctors but can indicate a high risk for breast cancer. DCIS represents cancer cells that are confined to the milk ducts in the breast and have not spread to the fatty tissue or the lymph nodes. It can usually be detected in mammograms by the specks of calcium, known as microcalcifications. Stage 1 cancer occurs when the original tumor is about 2 cm in length or less and is localized within a specific area. Stage IIA usually denotes a tumor that is 2 to 5 cm in diameter and has not spread to the lymph nodes. Stage IIB indicates a cancer that is 2 to 5 cm in diameter and has spread to the axillary lymph nodes, or the tumor is over 5 cm but has not spread to any of the nodes. Stage IIIA indicates any type of cancer that has spread to the axillary lymph nodes as well as the axillary tissues. Stage IIIB indicates any breast cancer that has attached itself to the chest wall and has spread to the pectoral lymph nodes. Stage IV, which is the most severe of all stages, indicates that the cancer has spread to other parts of the body.

TABLE 1.2 Five year survival rates of various stages of breast cancer.

0	100%
I	98%
IIA	88%
IIB	76%
IIIA	56%
IIIB	49%
IV	16%

The stage in which the cancer is detected plays a big role in determining the survival factor of the patient. Table 1.2 provides the 5-year survival rates for cancers at different stages. The key thing to notice in the survival rates in the percentages for anything below stage III. Even at stage IIB, where the tumor has spread to some lymph nodes or the tumor is 5 cm in diameter, the survival rate is at 76%. This is the point where the tumor has grown to a size where it can be felt by the patient. Stages 0 and I are the most difficult to detect outside of mammography features, but by detecting in these early stages, survival in almost guaranteed. This stresses the importance in regular mammograms, as well as risk analysis, to insure the earliest possible detection of the breast cancer. It is suggested that women over 40 years of age get yearly mammograms, but if a women has a higher risk of acquiring breast cancer, steps should be taken earlier. A detailed risk analysis plays a significant role in determining when a woman should obtain regular mammograms.

1.7 Statement of Problem

The estimation of radiodense tissue has traditionally been a subjective determination by trained radiologists using Wolfe's Classification. The main problem is the variability associated with the subjective nature of radiologists [10]. The same image may produce different results depending on the radiologist, and sometimes for the same radiologist. Later work by Boyd, Jensen, *et al* improved on the subjectivity of Wolfe's work but was still based on observer decisions.

A completely automated system for estimating the amount of radiodense tissue indicated in mammogram images has considerable potential to provide a rapid, objective

and cost-effective procedure for estimating breast cancer risk. There have been several attempts by previous researchers to arrive at such a system.

There are many issues that must be taken into consideration when analyzing a digitized mammogram. One issue is that the quality of the image depends mostly on the technician who acquires the image. Factors such as kilo-voltage peak, milli-ampere-seconds, anode/filter combinations, focal spot, positioning, and compression all play a pivotal role in the contrast, sharpness, and resolution of an image. Even if these variables can be minimized with a standard procedure, there is still the issue of patient diversity. The previous algorithms created in this research attempt to account for these problems but have difficulties when new image types are introduced to the system.

Many of the algorithms that have been created only focus on a particular dataset or image type. This causes problems when new images from other datasets are introduced into the algorithm. The previous two algorithms created at Rowan University also suffer from the same issues as well as other problems introduced during implementation. These problems include:

1. The images were segmented from algorithms that assumed that the histograms of the mammogram were bimodal in nature, when in fact very few of the mammograms in the database followed that trend.
2. There were many portions of the algorithm that were based on user feedback rather than image statistics. These inputs prevented the algorithms from being fully automated.

3. The algorithms' success was only based on the final numerical output of the segmentations, rather than a combination of visuals and percentages, this introduced false positive results.
4. Many of the images were manually filtered; this introduced another step that made the algorithms non-automated.

This thesis addresses many of the problems associated in other algorithms and introduces several ideas that attempt to make the algorithm invariant to any non-tissue changes in the image. The specific aims of this thesis are:

1. To survey current techniques for the automated segmentation of radiodense tissue in a digitized mammogram.
2. To evaluate various texture-based filtering, image enhancement, and image processing methods for segmentation.
3. To validate results compared to expert-annotated images.
4. To implement a new algorithm for use on the FCCC database.

1.8 Scope and Organization of the Thesis

This thesis presents an algorithm that attempts to mimic the overall procedure that a radiologist performs when analyzing the mammograms. There are several ideas that are adopted from previous research done at Rowan University but are now implemented in novel ways. This thesis also introduces a new method of Local Contrast Estimation (LCE) which will allow for better inter-dataset performance. All implementations created include:

1. New image pre-processing methods that focus on the problems created during scanning procedures.
2. A new tissue segmentation procedure based on new image enhancement and image processing techniques.
3. A more accurate and more efficient method to create the parametric mask used to account for the artificial change in compression created during the mammogram procedure.
4. A use of LCE for the estimation of breast density.
5. An investigation of various texture segmentation methods as well as other image processing methods for radiodense tissue estimation.

All algorithms and implementations were developed using mammograms obtained from the Channing Laboratory of the Brigham and Women's Hospital and Harvard Medical School (Harvard) and mammograms obtained from Fox Chase Cancer Center (FCCC). The Harvard images were obtained pre-digitized in bitmap format. The FCCC mammograms were obtained from Fox Chase Cancer Center and had to be digitized on site using two types of scanners: the Agfa scanner and the Lumisys. The mammograms were evaluated by a radiologist as a benchmark for the results obtained in this thesis. The results of this research are intended for use in a study conducted at Fox Chase Cancer Center (FCCC), which examines correlation between dietary patterns and breast density.

This thesis is organized into five chapters. Following the introduction, Chapter 2 provides a description of previous techniques for the quantification of radiodense tissue

and details the mathematical concepts used in the algorithm, such as edge and texture detection. Chapter 3 gives an overview of the approach taken in this thesis. Chapter 4 details the approach taken and displays all results obtained. Finally, Chapter 5 provides a conclusion of the work presented in this paper and an outline of possible future work.

CHAPTER 2: BACKGROUND

This chapter is split into three major sections. Section 2.1 details some of the previous methods that have been used in the quantification of radiodense tissue in a digitized mammogram. Section 2.2 discusses many of the texture methods that are evaluated in this thesis. Section 3.3 explains all other mathematical techniques that were used.

2.1 Breast Density

In the mid 1970's, studies were done by J. Wolfe to determine the relationship between the amount of parenchymal tissue located in the digital mammogram and the associated risk of breast cancer [6, 7]. These studies show that there is indeed a relationship between the amount of radiodense tissue and the risk associated with acquiring cancer. Over the past few decades, several other studies have been performed to further investigate the relationship between radiodensity of breast tissue and increased risk for cancer. Many of these trials produced results that were similar to the finding that Wolfe established. [16-26]. Ever since Wolfe's investigations, it has been shown that woman with breast density of 60-75% or higher are at a 4-6 times higher chance of acquiring breast cancer than compared to women with little or no breast density [12]. Combined with these findings, several studies have been done attempting to find the correlation of breast density compared to other important risk factors. One study done by Boyd et al showed that breast density is a heritable trait, which correlates to the understanding that women with family members who have had cancer are at a much higher risk of acquiring breast cancer themselves [15]. The fact that breast density can be traced genetically,

which in turn may allow cancer to be traced through a family, permits for a better understanding of the pathogenesis of breast cancer.

As stated earlier, a completely automated system for estimating the amount of radiodense tissue indicated in mammogram images has considerable potential to provide a rapid, objective and cost-effective procedure for estimating breast cancer risk. Unfortunately, because of the many issues that must be taken into consideration when analyzing a digitized mammogram, an automated algorithm must take into account that mammograms from varied locations will be inherently different. Even if these variables can be minimized with a standard procedure, there is still the issue of patient diversity. The Mammography Quality Standards Act was passed in 1992 in Congress to address some of the programs associated with the variability in mammograms, but the problems still exist.

2.1.1 Manual Techniques

The estimation of radiodense tissue has traditionally been a subjective determination by trained radiologists using Wolfe's Classification. Wolfe's Classification involves classifying mammograms into 4 separate categories:

N1: The breast is comprised entirely of fat.

P1: The breast has up to 25% nodular densities.

P2: The breast has over 25% nodular mammographic densities.

DY: The breast contains extensive regions of homogeneous mammographic densities.

Wolfe's Classification was popular for many years but had its limitations. The main dilemma with the method, a problem that it shares with other methods, is the subjective nature of the evaluation. Studies have shown that screenings were inconsistent from radiologist to radiologist [27]. The same mammogram produced varying results. Later work by Boyd, Jensen, *et al* improved on the subjectivity of Wolfe's work but was still based on observer decisions [28].

The Breast Imaging Reporting and Data System (BI-RADS) incorporates a four-category classification scheme for mammographic density, as recommended by the American College of Radiology [29]. The classifications include:

1. The breast is almost entirely fat.
2. There are scattered fibroglandular densities.
3. The breast tissue is heterogeneously dense. This may lower the sensitivity of mammography.
4. The breast is extremely dense, which could obscure a lesion on mammography.

This classification system is more a measure of how likely a lesion can be hidden within the breast tissue due to the high radiodensity. Radiologists typically estimate the breast density on mammograms in clinical practice based upon this BI-RADS classification. This method has many similarities with Wolfe's Classification technique, including the problem of inter-observer variability.

One main problem with previous classification methods is that they are too broad for use in research pertaining to radiodensity. In one study, the breast density was classified into six different categories [30]. Again, this method is based on the amount of perceived radiodensity in the breast tissue and is classified into percentage categories. The mammograms are classified as being 0 %, greater than 0 to less than 10 %, 10 to less

than 25%, 25 to less than 50 %, 50 to less than 75 %, or 75 to 100 % dense. Even though there are now six different categories, there still remains the problem of inter-observer variability, similar to the other methods. Because of the many problem associated with manual techniques, research has been done to create semi-automated and automated techniques to address the issues.

2.1.2 Semi-Automated Segmentation Techniques

The classifications in the previous methods are all based on the subjective opinions of the radiologist. The issue is that the definitions of radiodensity are broadly defined in previous methods and can be interpreted differently by each radiologist. To remedy this problem, several methods have been introduced to create interactive rule based systems for estimation of radiodensity.

Yaffe, Boyd *et al* developed an interactive software to identify radiodense regions within the breast [31, 32]. This program is still based on the subjective evaluation of a radiologist, but allows the physician to quantify the actual number with great detail. First, the digital mammogram is digitized into 4096 (2^{12}) discrete grayscale levels per pixel. Using the program, the radiologist segments the image into 3 separate regions: radiodense, radiolucent, and non-tissue. An initial threshold is chosen for each region to first establish a boundary between non-tissue and tissue regions. Afterwards, a second threshold is chosen by the user to distinguish between the radiodense and radiolucent regions. This method is referred to as the "Toronto" method. Figure 2.1 demonstrates a step by step illustration of the procedure.

Magnetic Resonance Imaging (MRI) can also be used as a method for the quantification of radiodensity, although a very expensive one [33]. An MRI is scanned by laying the patients in a prone position under two surface coils that are placed q above

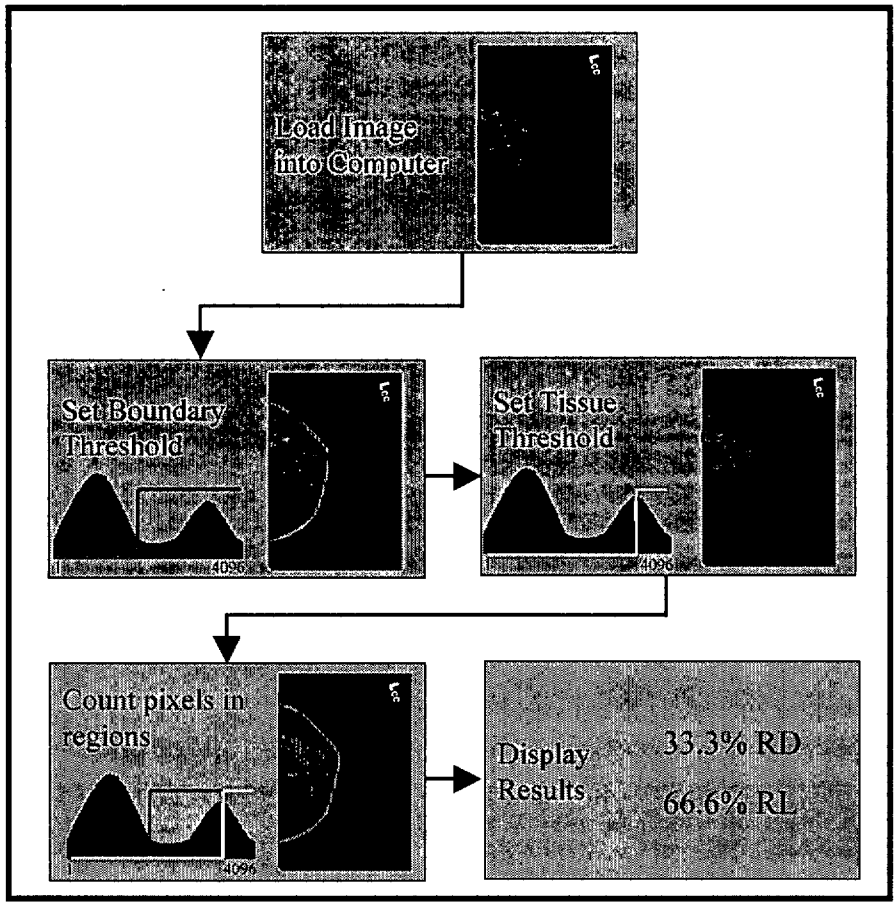


FIGURE 2.1 Block diagram of the Toronto interactive computer method for determination of radiodense tissue percentages.

each other. The breast is scanned from the cranial to caudal direction with constant thickness using an imaging device. A program is used with the MRI's to obtain an analysis of the radiodensity, similar to how the Toronto method is done. This technique has potential, but because there is no single, standardized and generally accepted technique for all breast MR imaging examinations, it is not widely used [34].

2.1.3 Automated Segmentation Techniques

Although semi-automated techniques have been created to address many of the problems with inter-observer variability, variability still persists. The standardized hardware and software is used differently depending on the person. Therefore, research has been done to create automated techniques that remove any subjective observations in the analysis. A table of some methods is shown in Table 2.1.

TABLE 2.1 Summary of automated segmentation techniques in breast density analysis.

Proponents	Approach	Advantages	Disadvantages
Lou and Fan [35]	Adaptive fuzzy K-means technique to classify pixels as radiodense.	7.98 % error among 81 mammogram images. 18 seconds process time per image.	
Zou et al. [36,37]	Rule based histogram classifier	Maximum difference 20% from expert analysis.	No object method for validation.
Bovis and Singh [38]	Classification using texture analysis.	91 % correct classification.	Relies on knowledge of the region to be segmented. Classifier is based on simplistic measures of texture.
Saha, Udupa, et al. [39]	Scale-based fuzzy connectedness models	Estimates correlate strongly with analysis by radiologist.	Does not automatically exclude pectoral muscle.
Neyhart et al. [40] Eckert et al. [41]	Constrained Neyman-Pearson decision function w/wo Compression Adjustment	Automated technique	Performance fit to database tested with. Weak inter-dataset performance.

Lou and Fan created an automated technique for the quantification of radiodense tissue [35]. Their process uses adaptive histogram equalization techniques along with an

adaptive fuzzy K-means technique to classify pixels. Within their own database, they achieved a performance error of 7.98 %.

Zhou et al. developed a rule based histogram classifier to determine the amount of dense area within a mammogram [36, 37]. Based on the BI-RADS breast density ratings, a histogram classification is used to place the image into one of the four categories. The threshold is obtained using a discriminant analysis method or a maximum entropy principle method, the choice based on the BI-RADS category. Compared to the estimates of two radiologists, their showing resulted in a 20% max difference. They also showed that the estimates of the two radiologist differed by 20%, which causes difficulties in validating any algorithm that is created.

A method has been created to classify the mammograms into density categories similar to the ACR BI-RADS classification [38]. Three methods are used to obtain features: Spatial Grey Level Dependency (SGLD) matrices, fractal dimension, and statistical gray-level measures.

The SGLD matrices model the correlation between pixels within the breast region. The SGLD matrix is the joint probability occurrence of gray-levels i and j for two pixels with a defined spatial relationship in an image. Calculating some measure of scatter of the SGLD matrix around the main diagonal will analyze the texture coarseness. By the end, fifteen statistical measures are extracted from this SGLD matrix. The fractal dimension is calculated for every pixel in the region and the mean value over all pixels is used as the fractal dimension feature. Statistical features like the mean, homogeneity, standard deviation and skewness of grayscale values of the breast tissue region are all used as features.

Based on the many features obtained from the three methods, the images were classified into three categories, 'dense', 'glandular' or 'fatty', by using supervised learning techniques. The K-Nearest neighbor classifier obtained the best results, scoring a 91% recognition rate. The researchers discuss two problems with these techniques [38]. First, radiologists analyze texture variations in a complex and subtle manner and classifiers based on simplistic measures of texture do not perform well; second, the differences in imaging conditions lead to a non-rigid variation in the mammographic intensity distribution. These two conditions are stated as reasons for why textures may not be a valid method for this research.

One of the more promising methods developed recently is an automated method has been developed by researchers at the University of Pennsylvania. They use principles of fuzzy connectedness to quantify radiodensity [39]. One of the main principles in this method is the idea that objects within the image stay together as a whole. These objects can be easily identified by an observer, but computers have trouble due to the many variation created within the image. For classification, the researchers define a fuzzy affinity, which is used to define a relationship between two pixels. Based on this affinity, the correlation of pixels is determined.

First a region of reference is identified using the highest 32% of intensities. Next, a "fuzzy connectivity scene" is generated to determine the likelihood that a region will be correlated with the reference region. A fuzzy connectivity scene is created for the dense region of the breast tissue and segmented using an automatic threshold selection method. It has been shown that this method correlated highly with the observation of a trained radiologist.

The Neyhart and Eckert methods are both automated techniques based constrained Neyman-Pearson decision rules. These techniques will be discussed further in Chapter 3 and Chapter 4 [40, 41].

2.2 Textures

Texture refers to the properties that represent the surface and structure of the object. It is basically a way to measure the object's three dimensional surface characteristics from a two dimensional image. Usually, these characteristics can be used for processes such as image segmentation based on texture, as well as image classification. A formal definition of texture segmentation could be the following: a partitioning of an image into regions, each of which contains a single texture distinct from its neighbors. From a definition standpoint, segmentation based on textures seems very simple and straightforward. Unfortunately, textures exhibit various characteristics that are not always the same based on the orientation and size of the texture. Therefore, it usually takes extensive testing before the proper patterns can be discerned and used. To start with, consider the two terms "texture" and "segmentation" separately [42].

Mathematically, image segmentation is well-defined. For any image I(x,y), which consists of an array of pixels, the goal is to label the pixels into some group. Figure 2.2(a) shows an example of an image with three separate groups. It can be seen visually, that there are three separate groups in the image by either distinguishing between the shapes or the colors of the objects. Unfortunately, this is a very simple example and does not represent what is typically seen in a image segmentation problem. An object can be considered a region that consists of a connected group of pixels that share the same

statistics. This sounds simple enough, but actually choosing what statistic to use can drastically change the results obtained from an image. Figure 2.2(b) can be used as an example. The stones on the object are a good example of how the statistics chosen for comparison can affect the segmentation results.. If one type of statistic is chosen, the result may classify each individual stone as an object. On the other hand, another type of statistic may group all the stones into one object. Therefore, there are many results that can be obtained from segmentations and no one result is the right one. Basically, the right result is the one that is most suitable for the user.

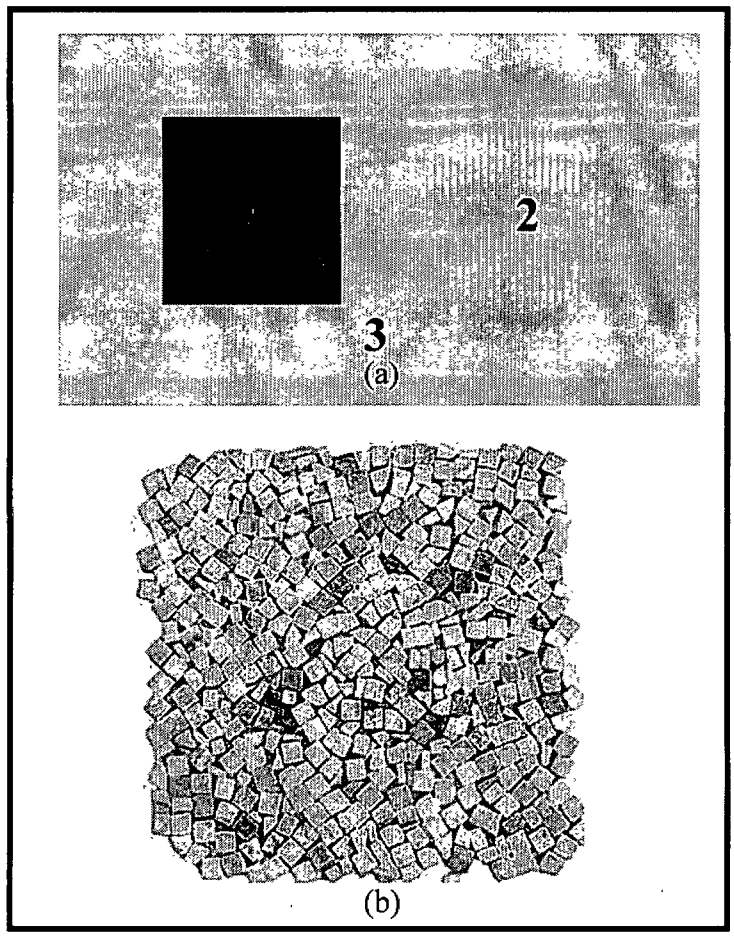


FIGURE 2.2 (a) Three different objects in the image can be classified by using either the shape or color. This is an example of a simple segmentation problem. (b) The image is more complex and can have different segmentation results based on the statistic chosen.

The concept of texture is even more complicated than that of segmentation. Even though image segmentation has a formal definition, texture does not have one due to its variability. A typical definition in the field is "one or more basic local patterns that are repeated in a periodic manner." However, it is not always clear exactly what the pattern is or how it is repeated. It is also unclear whether all objects have textures or some contain no texture at all. Simply, a texture consists of objects called texture primitives or texture elements, also referred to as texels. The texels can be considered as the smallest element that can be considered that texture. As stated previously, textures are very scale dependent. The main goal in texture description is to create a mathematical representation of the phrases that are usually commonly used to describe texture, phrases such as: fine, coarse, grained, smooth, etc

There are two approaches to defining texture, which may be thought of as "top-down" and "bottom-up." Top-down models claim that there is a basic element, called a texel, sometimes referred as a primitive. These elements are defined by a placement rule, which defines how and where the elements are placed. This definition works well if the texture consists of modular primitives, such as, bricks or piles of pennies. The bottom-up approach claims that texture is a property that can be derived from the statistics of small groups of pixels, such as mean and variance. This works better for textures like sand and grass where it is difficult to see individual elements. Because of the wide variation of texture definitions, there is no defined delineation between the two types of methods.

2.2.1 Texture Description Methods

Because of all the variability, no one texture model has been shown to give great results all the time. Many methods have been created from scratch and many have been

borrowed from other fields other than engineering. Models are often conceived by thinking of the image not as a set of pixels, but rather as a function, $I(x,y)$, defined over a continuous region of the plane. The process in which the model is discretized and then implemented is of no less importance than the theory behind it. This section will detail the methods that have been used in this thesis for analysis.

2.2.2 Gabor Filtering

Gabor filters work on the principle that each texture will contain within it some region of frequency that is different from another texture's frequency regions. By finding this region and filtering everything else in an image, it has been shown that the textures can be segmented using Gabor filters [43].

A real impulse response of the Gabor filter in the spatial domain is defined by:

$$h(x,y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp\left\{-\frac{1}{2}\left[\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right]\right\}\cdot\cos(2\pi\mu_0 x) \qquad (2.1)$$

which is a sinusoidal modulated Gaussian with a spread of σ_x and σ_y in the x and y directions respectively that has been modulated by a frequency of μ_0. The real impulse response can be seen Figure 2.3(a). In the frequency domain, the Gabor filter becomes two modulating frequency shifted Gaussians, as shown in Figure 2.3(b). The equation of the frequency response is given by:

$$H(u,v) = \exp\{-2\pi^2[\sigma_x^2(u-u_0)+\sigma_y^2v^2\} + \exp\{-2\pi^2[\sigma_x^2(u+u_0)+\sigma_y^2v^2\} \qquad (2.2)$$

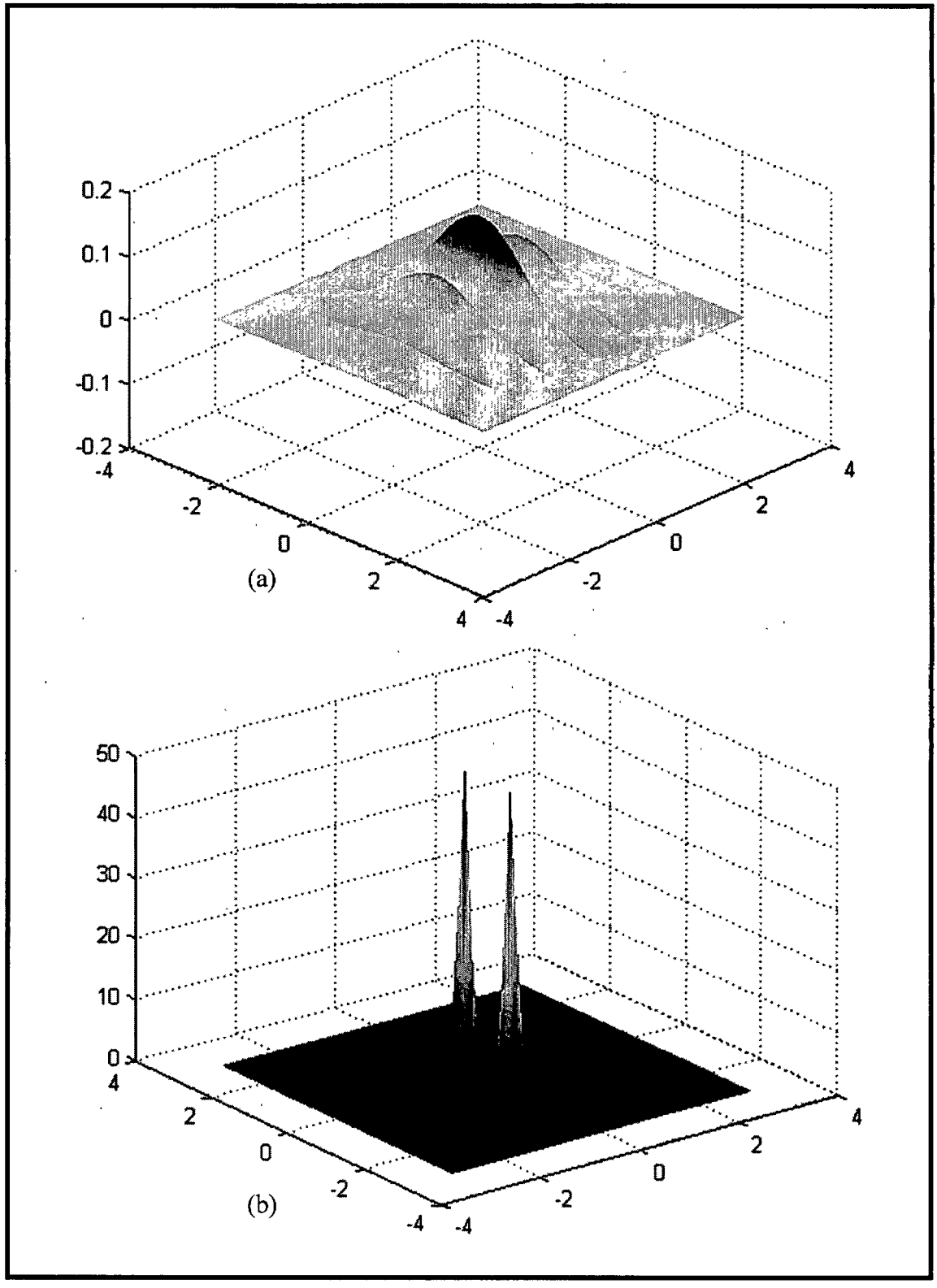


FIGURE 2.3 (a) A plot of the impulse function of the Gabor filter. (b) A frequency domain plot of the Gabor Filter.

The Gabor filter can be used to obtain texture segmentation in both the spatial and frequency domain. In the spatial domain, it functions as a wavelet decomposition and can be used to obtain features based on post-filtered images. In the frequency domain, the Gabor filter can be used to remove the unwanted frequency components located in the image that corresponds to the unwanted textures.

2.2.3 Co-occurrence Matrices

Co-occurrence matrices, also called Spatial Grey Level Dependence Matrices, attempt to capture texture using a sparse representation [44]. Each matrix in the set corresponds to an offset (e.g. 1 pixels down and 1 pixel to the right). The entry in row i and column j of each matrix is the number of pixels in the image of grey level i that have a neighbor of grey level j in the direction of the offset. From these matrices a number of statistical quantities can be measured, such as mean, variance, entropy, energy, contrast, and correlation. Figure 2.4 shows an example of how a zero degree co-occurrence matrix works. Figure 2.5 shows the equations used to obtain the features.

$$P_0(a,b) = |\{[(k,l),(m,n)] \in D : k-m=0, |l-n| = d, f(k,l)=a, f(m,n)=b\}|$$

$$\begin{matrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 2 & 2 & 2 \\ 2 & 2 & 3 & 3 \end{matrix} \qquad P_0(a,b) = \qquad \begin{matrix} 4 & 2 & 1 & 0 \\ 2 & 4 & 0 & 0 \\ 1 & 0 & 6 & 1 \\ 0 & 0 & 1 & 2 \end{matrix}$$

FIGURE 2.4 A co-occurrence matrix based on a 0, 1 positioning, which is a comparison of pixels that are to the right of another pixel.

Essentially, the main problem with the co-occurrence matrices is the number of calculations required for each region scales according to how many gray levels there are. For example, when there are four gray levels as shown in Figure 2.4, a matrix that is 4 by 4 is formed as the output, which is 16 numbers that must be calculated. If analyzing images that have 256 gray levels, the matrix formed becomes very large and over 65,000 numbers must be calculated. For simpler textures that are more distinct from each other, computation speed can be increased by quantizing the number of gray levels to a smaller number.

$$\sum_{a,b} P_{\phi,d}^{2}(a,b)$$

$$\sum_{a,b} P_{\phi,d}(a,b)\log_2(P_{\phi,d}(a,b))$$

$$\sum_{a,b} |a-b|^{k}\, P_{\phi,d}^{\lambda}(a,b)$$

$$\sum_{a,b,a\neq b} \frac{P_{\phi,d}^{\lambda}(a,b)}{|a-b|^{k}}$$

FIGURE 2.5 Co-occurrence features; energy, entropy, moments, inverse moments, respectively.

Once all the features are obtained, both the supervised and un-supervised methods can be used to obtain a classification of textures.

2.2.4 Law's Texture Energy Measures

Law's texture energy measures is a method to create simple spatial filters that can be used to measure statistics such as average gray-level, edges, spots, ripples, and waves in texture [46]. These measure are derived from three simple vectors: L_3 = (1,2,1), E_3 = (-1,0,1), and S_3 = (-1,2,1), which stand for averaging, edges, and spots respectively. By

the convolution of the vectors with themselves and the other 2 vectors, 5 new vectors are obtained, as shown in Figure 2.6(a). From these 5 vectors, 25 different filters can be obtained by a mutual multiplication, the first term as a column and second term as a row, as shown in Figure 2.6(b). The result of the convolution between the image and a filter

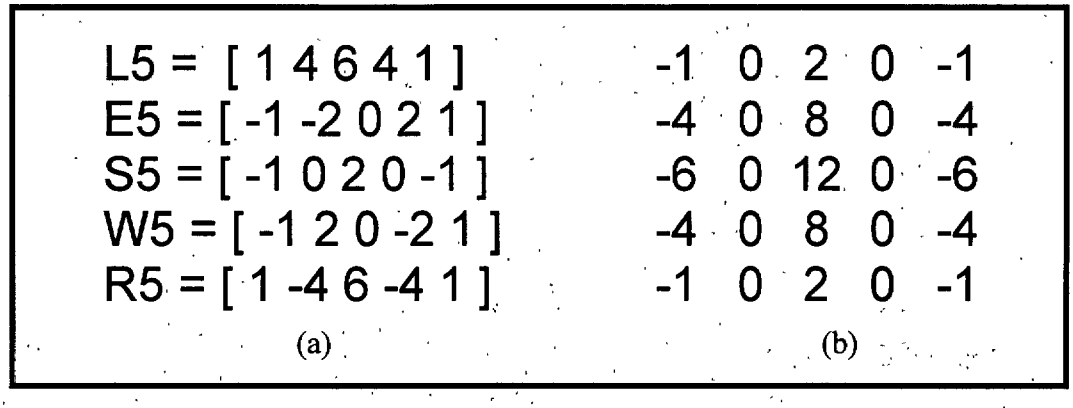


FIGURE 2.6 (a) 5 vectors created by the convolution of Law's 3 energy measures. (b) One of the filters created using L5 and S5.

combined with an energy calculating statistic can be used to obtain 25 different features. These 25 features can be used in clustering techniques for classification.

2.3 Other Methods

Other methods other than textures were used in this thesis. These methods include: gray level connectivity, image morphology, radial basis function networks, and K-means clustering. This section will explain in detail these methods.

2.3.1 Gray Level Connectivity

The connectivity between pixels is a fundamental concept that is used to define various digital image concepts such as regions and boundaries. Two pixels are considered

connected if they satisfy certain location criteria as well as gray level criteria. For instance, in a binary image with values 0 and 1, two pixels may be 4-neighbors, but they are said to be connected only if they have the same value.

V is usually considered to be the conditions in which the gray level relationships are defined. For example, in a binary image, V={1}. The 1 defines that if two pixels are considered neighbors by location and they both have values of 1, they are connected. In gray scale images, the idea is similar, but the definition of V can be anything from a single value to a region of numbers. For example, in the adjacency of pixels with a range of possible gray-level values 0 to 255, set V could be any subset of these 256 values.

There are three types of adjacency:

1. 4-adjacency. Two pixels, p and q, with values from V are 4-adjacent if q is in the set $N_4(p)$, where $N_4(p)$ is the set of pixels defined by $(x+1, y), (x-1, y), (x, y+1), (x, y-1)$
2. 8-adjacency. Two pixels, p and q, with values from V are 8-adjacent if q is in the set $N_8(p)$, where $N_8(p)$ is a combination of $N_4(p)$ and $N_D(p)$, which is a set of pixels defined by $(x+1, y+1), (x+1, y-1), (x-1, y+1), (x-1, y-1)$
3. M-adjacency (mixed adjacency). Two pixels, p and q, with values from V are m-adjacent if:
    a. q is in $N_4(p)$, or
    b. q is in $N_D(p)$ and the set $N_4(p) \cap N_4(q)$ has no pixels whose values are from V.

Mixed adjacency is a modification of 8-adjacency that was created to account for the problems that can occur when using 8-adjacency. For example, consider the pixel

arrangement shown in Figure 2.7(a) where V={1}. The three pixels at the top of Figure 2.8(b) show the problem that can occur when using 8-adjacency, as indicated by the dashed lines. By using m-adjacency, it is insured that a pixel will not be used twice in determining connectiveness.

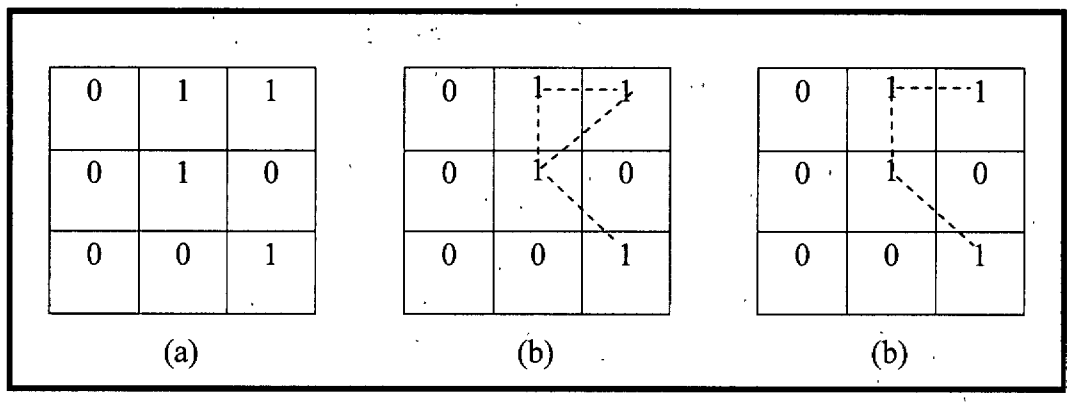


FIGURE 2.7 (a) Example arrangement of pixels. (b) 8-adjacency. (c) m-adjacency.

Connectivity is commonly used in segmentation problems after an image has been either changed into a binary image or when there is a gray level image that does not vary much in each region. Because of the fundamentally simple nature of the connectivity definition, the algorithm has a hard time analyzing images that have a high level of variance throughout the image [46].

2.3.2 Image Morphology

Dilation and erosion are two basic morphological operations that are commonly used in image processing. These two methods can be valuable for extracting image components that are useful for representation and description. Morphology can provide boundaries of objects, their skeletons, their convex hulls and can also be used in thinning applications as well as bridge gapping. Morphological operations are based on the simple concept of

expanding and shrinking objects that are related. The operations are usually performed on binary images but are occasionally adapted for grey level images [47].

The dilations and erosion transformations are based on the interaction of an image A and a structuring set B, often called the structuring element. Structuring element B is often a circular disc in the image plane but can be any shape. The structuring element can be any number of dimensions, but is restricted to two dimensions for images. Structuring element B can be viewed as a convolution mask. Four important terms come into play for both dilation and erosion: translation, reflection, complement, and subtraction. If we consider A and B to be subsets of Z^2, the translation of A by x is denoted A_x and is defined by the equation:

$$A_x = \{c : c = a + x, for\ a \in A\} \tag{2.3}$$

The reflection of B is denoted by $\hat{B}$ and is defined by the equation:

$$\hat{B} = \{x : x = -b, for\ b \in B\} \tag{2.4}$$

The complement of A is denoted by A^c, and the difference of the two sets is defined by A – B.

Dilation of an object A by a structuring element B is given by the equation:

$$A \oplus B = \{x : \hat{B}_x \cap A \neq \varnothing\} \tag{2.5}$$

The result of this operation is the reflection of B around its origin and then shifting this reflection by x. Figure 2.8 demonstrates a dilation procedure.

Erosion of an object A by the structuring element B is defined by the equation:

$$A \theta \mathrm{B} = \{x : B_x \subseteq A\} \tag{2.6}$$

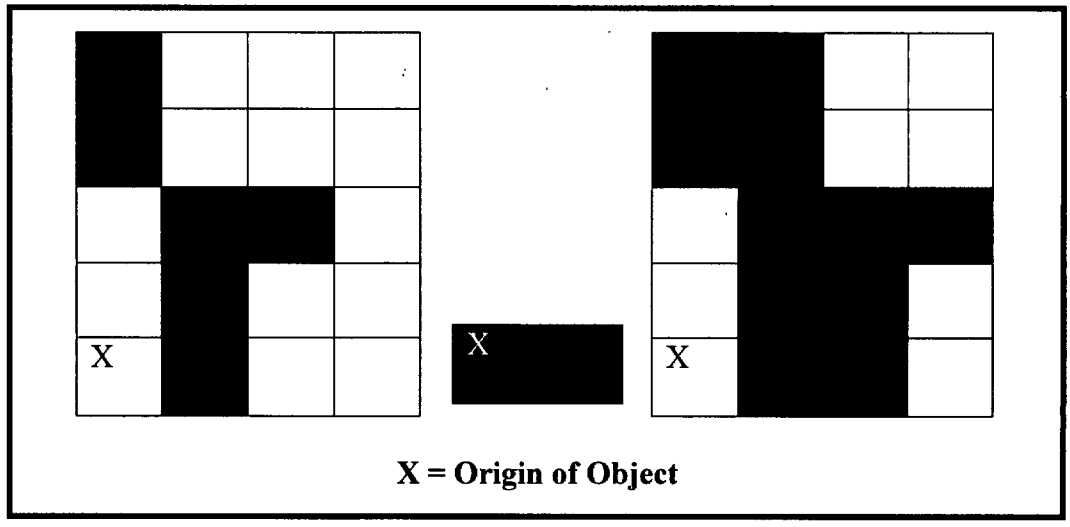


FIGURE 2.8 Example of dilation.

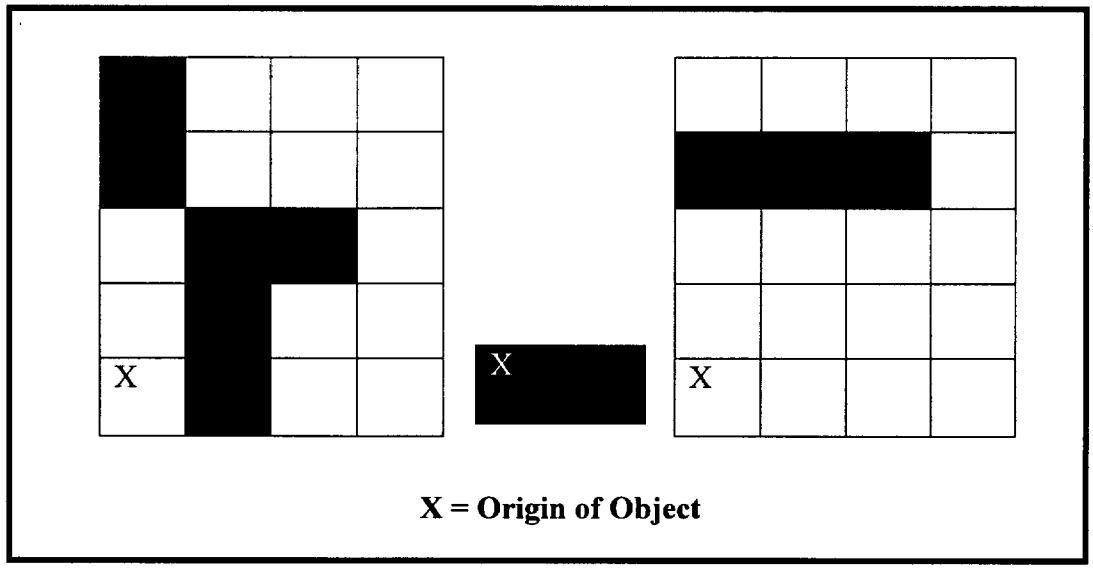


FIGURE 2.9 Example of erosion.

Erosion combines two sets using vector subtraction of set elements and is the dual operator of dilation. An example of erosion is shown in Figure 2.9.

Erosion and dilation can be used in a variety of ways, in parallel and series, to implement other transformations such as thickening, thinning, and skeletonisation.

2.3.3 RBF Networks

Radial basis functions (RBF) are characterized by a two-layer feed-forward neural network. The network contains a set of inputs and outputs with hidden layer nodes in between that function as the processing units. Each of these hidden layer nodes implements a radial basis function. Figure 2.10 shows what the architecture of an RBF looks like [48].

RBF networks can be used in two ways:

1. Data modeling for function approximation. For this method, the inputs are considered the missing samples of the time series or the spatial series. The output is the value that RBF network approximates.

2. Pattern classification application. For this application, the network may have any number of inputs that correspond to the features of a particular class. The outputs correspond to the different classes that can be possible in the data.

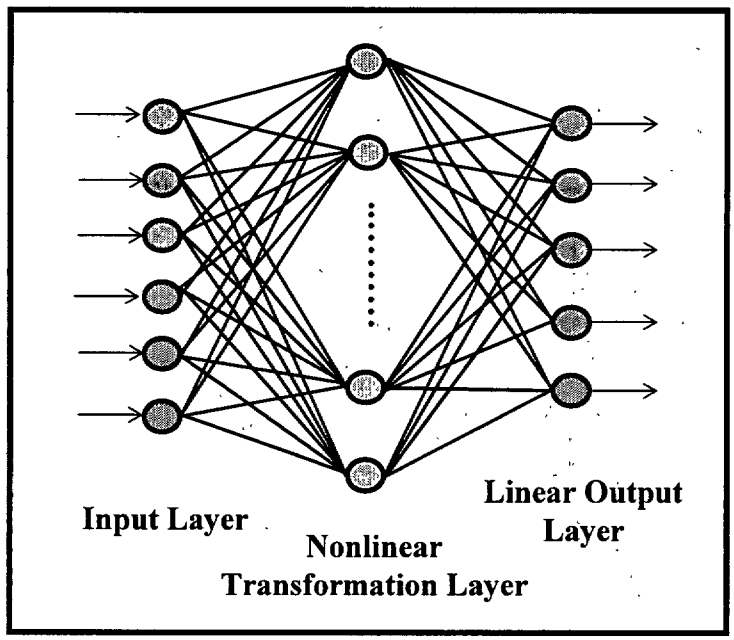


FIGURE 2.10 Radial Basis Function neural network architecture.

The activation function most frequently used for the RBF network is the Gaussian. The Gaussian activation function is given by the equation:

$$\phi_j(X) = \exp[-(X-\mu_j)^T \sum_j^{-1} (X-\mu_j)] \tag{2.6}$$

for j=1,…,L, where X is the input feature vector, L is the number of hidden units, μ_j and $\sum_j$ are the mean and covariance matrix of the *j*th Gaussian function. Geometrically, the RBF becomes a Gaussian in multidimensional space, whose dimension is given by the number of the inputs. The mean vector, μ_j, represents the location in each dimension and the $\sum_j$ models the shape of the Gaussian. Statistically, the activation function is a model of the probability density function.

The output layer is modeled as a weighted sum of the hidden layer outputs as shown in the equation below:

$$\Psi_k(X) = \sum_{j=1}^{L} \lambda_{jk} \varphi_j(X) \tag{2.7}$$

for k=1,…,M where λ_{jk} are the output weights, each corresponding to the connection between the hidden unit and an output unit. M represents the number of output units.

RBF networks are more commonly used in curve fitting and generalization operations. Typically, when it comes to classification, Multi-Layer Perceptron (MLP) networks are more commonly used.

2.3.4 K-means Clustering

A very popular method for classifying data in an unlabeled data set is widely known as *k*-means clustering. This technique attempts to discriminate between *k* different classes based on the center of each cluster. By approximating the mean of each cluster, the samples can be categorized by finding the closest cluster mean to it. If a sample of data is closer to a particular mean than it is to the others, most likely it will belong to that class. The proximity of data to the mean of the clusters can be calculated using the squared Euclidean distance given as $\|\mathbf{x}_k - \hat{\mu}_i\|^2$. Once this has been completed the means of the newly classified data can be calculated and the process will repeat itself. This iterative step of recalculating the means makes the result more and more accurate with each step. Eventually the means of the clusters will no longer change and the *k*-means algorithm will have classified all of the data. As long as the number of classes is known, *k*-means is effective at separating the data, even if no other information about the dataset is available [48].

The algorithm describing *k*-means clustering is shown in Figure 2.11, where *n* is the number of samples in the dataset and *c* is the number of clusters.

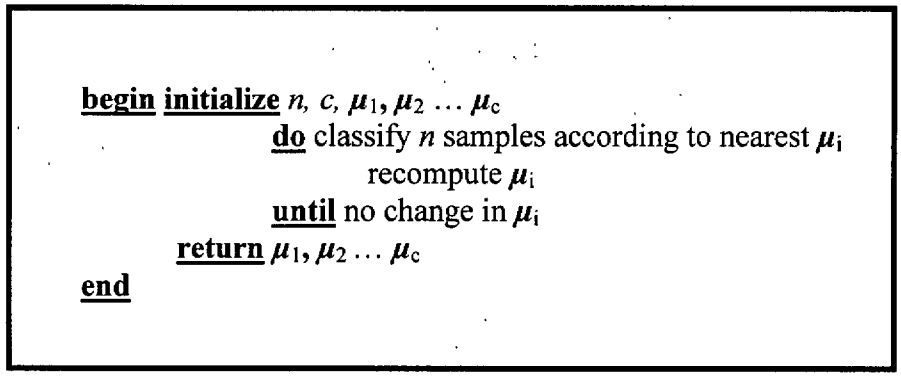


FIGURE 2.11 Pseudocode for k-mean clustering.

CHAPTER 3: APPROACH

The initial goal of this thesis was to investigate the use of textures for the automated segmentation of radiodense tissue from a digitized mammogram. As time passed, the scope expanded due to many of the issues and problems associated with the actual use of the textures. From the beginning of the experimentation phase to the implementation of the final algorithm, there were many pathways investigated. Some of these steps led to ideas that would finally be implemented in the Global Contrast Estimation for Radiodense Tissue Segmentation algorithm. Other paths were not directly applied to the algorithm, but still benefited this thesis by providing a good idea of for the final direction of this algorithm. The approach phase was split into three sections. Section 3.1 involved the investigation of textures as a possible use for the automated segmentation of radiodense tissue. Section 3.2 involved the investigation of other pre-processing methods as a way to make the segmentation procedure more Bayesian, rather than constrained Neyman Pearson. Section 3.3 involved the final implementation of the Global Contrast Estimation Algorithm, a significant deviation from the methods created previously by the research groups at Rowan University [40, 41]. In addition to explaining the final algorithm, the rest of this chapter details all methods that were tested throughout the entire research process.

3.1 Texture Evaluation

Using the original scope of the thesis as a basis for research, textures were the first segmentation method investigated in order to create an automated segmentation of radiodense tissue from the mammograms. As was explained in Chapter 2, since there

were many different definitions of textures as well as different methods for segmentation several of them had to be tested before any conclusion could be made about textures. The rest of this section details all the experiments that this thesis conducted with textures. The results of the experiment are shown in Chapter 4.

3.1.1 Regional Variance Statistics

Before any of the texture methods were tested, an investigation of simple 2-dimensional characteristics was done for two main reasons: to see if simpler statistics would allow for preprocessing of the radiodense region and to give a preliminary indication of whether or not texture segmentation would work in segmenting the mammograms into the radiodense and radiolucent regions. The 10 Harvard and 34 FCCC images were used for imaging of the simple characteristics as well as the texture methods introduced later on in this chapter. The database was used for two different methods in the regional variance testing.

First, the images were tested globally to see how they would look when each pixel was replaced by a new value that was equivalent to the variance function of that region. Variance comparison was chosen because this method eliminated the mean or gray level characteristics, meeting a main goal of this thesis which was to find a method to compare other characteristics other than just the gray level value. Figure 3.1 demonstrates a pseudocode implementation of the regional variance imaging procedure.

Even though these new variance based images were not expected to be much better than the new images, the hope was that they would show that the radiodense regions would have some characteristics in common with each other while being

somewhat different from the radiolucent regions. Figure 3.2 shows a demonstration of what this variance function did on canonical images obtained from the SIPI database [49]. A mask size of 9 x 9 was used for the implementation shown.

```
                    Regional Variance Imaging
Image_old \\ the original image
Image new \\ matrix of zeros the size of Image_old
for i=1:1:N \\ length of the rows of Image_ old
        for j=1:1:M \\ length of the columns in Image_old
                Image_new( i , j )=variance(Mask(Image_old( i , j )))
                \\ Where mask is a region defined by the area around
                the coordinates i and j
        end
end
```

FIGURE 3.1 The pseudocode implementation of how the new variance based images were obtained.

The textures shown in Figure 3.2(a) are simple textures because compared to each other many of their texture characteristics are vastly different. These textures also do not have multiple levels of primitives, which would have created more complicated results. Because the textures are simple, the resulting image from the regional variance imaging, as shown in Figure 3.2(b) ends up being separable. The hope was that the results for the mammograms would look similar to the results of the canonical images.

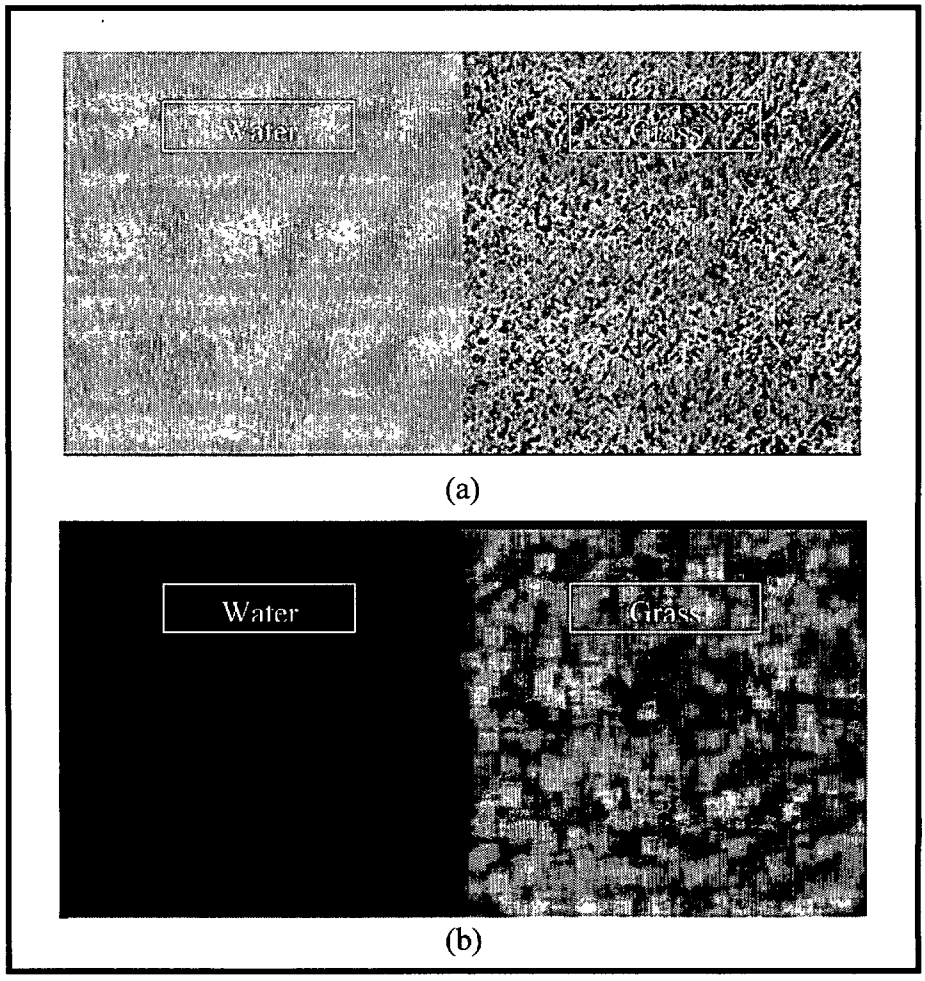


FIGURE 3.2 (a) Two textures obtained from the SIPI database. (b) Results of the regional variance imaging function for the two canonical textures.

The next step was to measure the statistics of the two regions, radiodense and radiolucent, of the database. First, the images were segmented into their various regions based on the percentages given by Dr. Celia Byrne. Since the actual profile of the radio-density regions obtained by Dr. Celia Byrne was never provided, this procedure had to be performed by adjusting the threshold of the binary segmentation until the amount of white pixels within the image was the same percentage given by Dr. Byrne. Because the masks used were generated by the new algorithm demonstrated in this thesis, there could be some small error associated with the actual segmentation, though the error should be

very minimal. The radiodense and radiolucent regions were separated from each other in order to calculate their individual variances. From these variances, several numbers were calculated: the average variance of the radiodense regions, the average value of the radiolucent regions, the variance of the 'variances', and the average difference between the two regions. Figure 3.3 demonstrates a pseudo implementation of the procedure that was done.

The main reason for obtaining these statistics was to see if the variances of the regions stay consistent through each image and if there any discernable differences between the variances of the two regions. As mentioned earlier, it was the hope of this

Database Variance Statistics

Database *\\ the 10 Harvard and 34 FCCCimages*

Threshold *\\ the thresholds based on the correct percentages given by Dr. Celia Byrne*

for i=1:1:N *\\ number of images in database*

radiodense = segment(Database(i), Threshold(i))

\\ group all pixels that are above the threshold

radiolucent = segment(Database(i), Threshold(i))

\\ group all the pixels that are below the threshold

radiodense_variance(i) = variance(radiodense)

radiolucent_variance(i) = variance(radiolucent)

\\ find the respective variances of both sections

Continued

```
end
variance_radiodense_variance = variance(radiodense_variance)
|| find the variance associated with the variance of all radiodense
 regions
variance_radiolucent_variance = variance(radiolucent_variance)
|| find the variance associated with the variance of all radiolucent
 regions
average_radiodense_variance = average(radiodense_variance)
|| find the average associated with the variance of all radiodense
 regions
average_radiolucent_variance = average(radiolucent_variance)
|| find the average associated with the variance of all radiolucent
 regions
difference_regions = difference(average_radiolucent_variance,
                                average_radiodense_variance)
|| find the difference between the two averages
```

FIGURE 3.3 The pseudocode implementation of the function that calculated variances of the two different regions as well as the statistics associated with them.

thesis that not only would these bits of information give a preliminary indication of whether or not the statistics between the regions were different, but also, that the information would stay consistent between all the images of the database. It was a measure of the intra- and inter-regional similarities. The results for this procedure are shown in Chapter 4 of this thesis.

3.1.2 Gabor Filtering

As stated in the background, Gabor filtering for textures has become more widespread in the past few years. The fact that it uses the frequency components as a basis for segmentation allows for the use of fast and efficient algorithms to achieve the final goal. Due to these reasons, Gabor filters were chosen as the first implementation of texture based segmentation for the mammograms. By having the Gabor filters, it also allowed the comparison between a frequency based method as well as spatially based ones, which will shown later in this thesis.

To use the Gabor filters as an 'automated' procedure in the segmentation of radiodense tissue, the images had to contain one fundamental statistic, the two regions used for separation had to contain some regions in the frequency domain that were drastically different from each other. Since this was the first time that the statistics of the two regions was being obtained in this work, this was not known. Therefore, the first step in this procedure was to obtain the frequency profiles of the two regions, radiodense and radiolucent, and determine which frequencies, if any, were different enough from each other in which they could be used for Gabor filtering. One constraint was that since it was not possible for the region to change from image to image, the choice of region would have to work with all images, which would allow the procedure to be automated.

Various regions of known radiodense and radiolucent were selected to test for the frequency content. 50 regions were selected in sizes of 30 x 30 pixels for both the radiodense and radiolucent portions of the 10 Harvard images. Only the Harvard set was used for this implementation of the code. Once the frequency profile of the regions was obtained using a 2-dimensional FFT, the average profile was computed for both the

radiodense and radiolucent regions. These two profiles were then compared to find the regions that were the most different from each other. Figure 3.4 demonstrates this procedure.

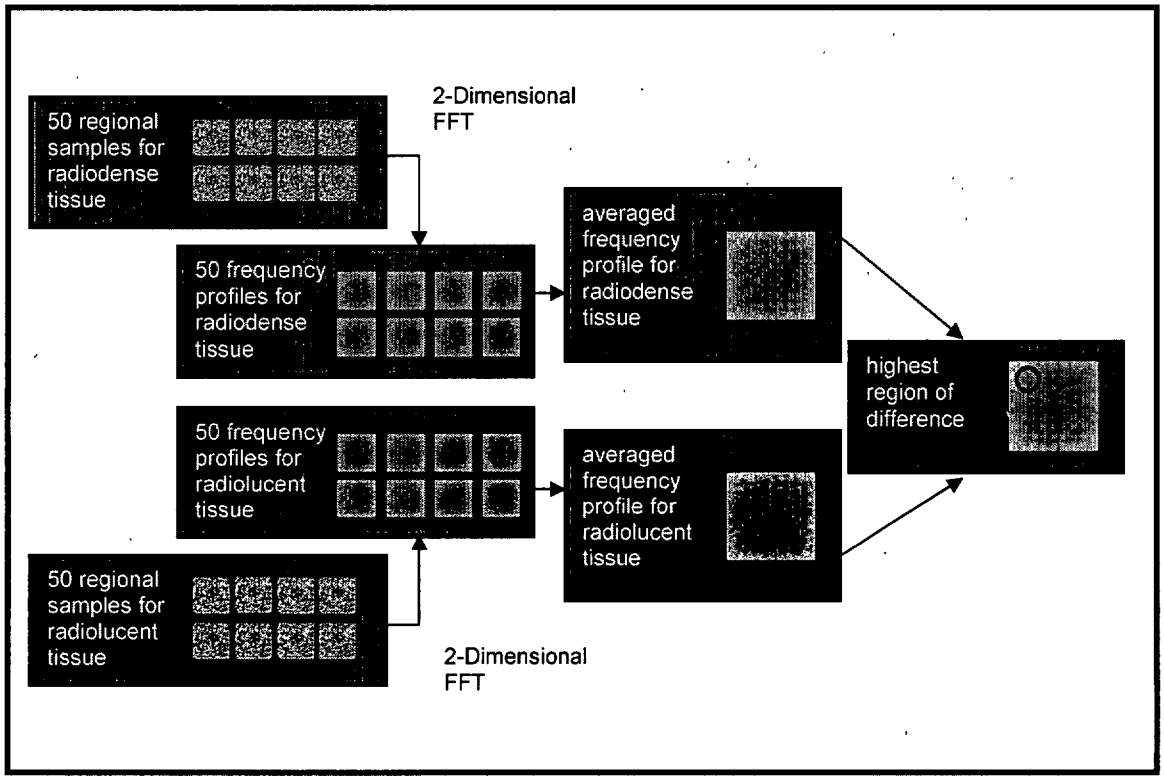


FIGURE 3.4 The procedure used to obtain the optimal region for placement of the Gabor filter in the frequency domain for texture segmentation of the radiodense tissue from the radiolucent tissue.

The hope was that the region of frequencies chosen from the samples of the known radiodense and known radiolucent regions would stay consistent throughout all the images, which would then allow for a decent segmentation of both regions based on frequency.

Once the region of maximum difference was chosen, the algorithm was exercised on the database with different size Gabor filters, which were chosen to make sure all scenarios were investigated. The results are shown in Chapter 4.

3.1.3 Co-occurrence Matrix

The co-occurrence matrix texture segmentation was chosen because of the many various features that could be obtained from the matrix, which lends itself to an automated segmentation through some form of statistical clustering. The co-occurrence matrix also allowed for a spatial transformation view of textures. The one major problem with the co-occurrence matrix was that it was very computationally expensive depending on the number of gray levels in the image. As the background chapter on co-occurrence explained, for every N number of gray levels, there was N^2 numbers that had to be obtained for the co-occurrence matrix for each corresponding point on the image. Given that at 256 gray levels each implementation could take up to an hour depending on the speed of the computer, various measures had to be taken in order to speed up the actual implementation. The co-occurrence procedure was split into two parts, the implementation of the actual algorithm and the automated clustering of the features.

The co-occurrence algorithm was created to take into account various known issues. First, as shown in Figure 3.5, even though the mammogram image was rectangular, the actual tissue information was located in a much smaller area. For example, the resolution of the image was 926 x 676, which corresponded to 625,976 pixels. Therefore, it required that many number of calculations if the co-occurrence algorithm was done on each pixel, regardless of whether it was needed or not. The number of actual pixels within the tissue region was 156,679. This meant that unless steps were taken to only calculate the features in the tissue region, this image would waste approximately 75% of the calculations of the X-ray region. To accommodate for this, the information of the mask was sent in to the co-occurrence matrix so that it would

only calculate the important areas of the image. The second issue was taking into account the symmetric nature of the co-occurrence matrix. As shown in the Figure 3.6, the information contained within any matrix is symmetric along the northwest-southeast diagonal, therefore if implemented properly; the number of calculations could be reduced by almost half at most.

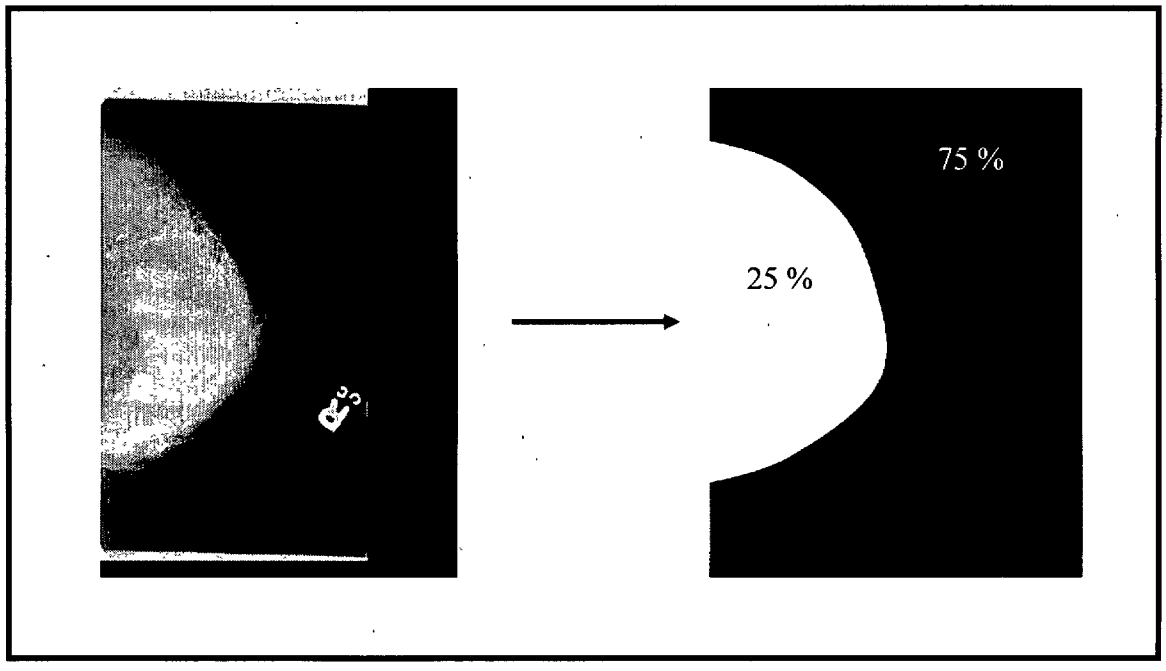


FIGURE 3.5 The breast is only located on a small percentage of the mammogram.

The code was implemented so that only the portions on the matrix that were below the diagonal or on it were calculated. It was because of the place keeping that the number of calculations were not completely reduced by half. Figure 3.7 shows a pseudocode implementation of what was done to obtain the co-occurrence matrix.

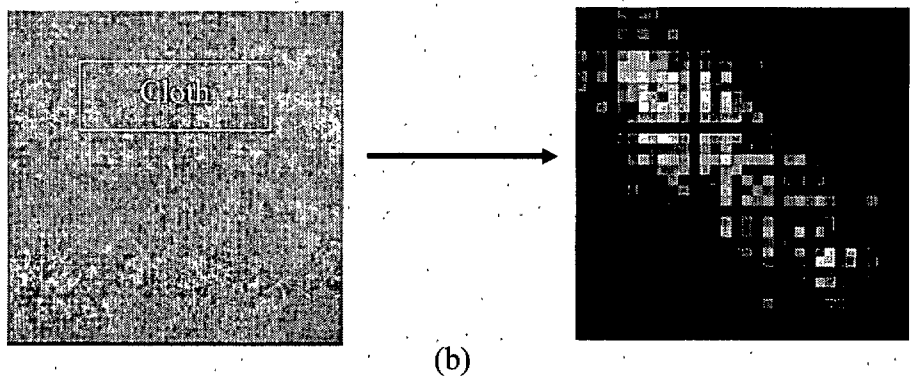
Once the features; the energy, the entropy, the inverse difference, and the contrast, of the co-occurrence features were calculated, they were prepared for the automated clustering technique. Because these features were images, they had to be changed into

single vectors before the algorithm could be implemented. Therefore, the non-tissue portion of the images had to be made into a third class to make the registration of pixel locations easier for the algorithm. Of course, if it was shown visually that the segmentation based on the co-occurrence matrix features worked properly, the final version of the code could have been implemented by only outputting the percentage of

$$P_0(a,b) = |\{[(k,l),(m,n)] \in D : k-m=0, |l-n| = d, f(k,l)=a, f(m,n)=b\}|$$

$$\begin{matrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 2 & 2 & 2 \\ 2 & 2 & 3 & 3 \end{matrix} \qquad P_0(a,b) = \qquad \begin{matrix} 4 & 2 & 1 & 0 \\ 2 & 4 & 0 & 0 \\ 1 & 0 & 6 & 1 \\ 0 & 0 & 1 & 2 \end{matrix}$$

(a)



(b)

FIGURE 3.6 (a) Symmetric co-occurrence matrix of a 4 level texture. (b) The co-occurrence matrix of a 32 level cloth texture.

Co-occurrence Feature Extraction

Feature *\\ the feature that is being calculated in the co-occurrence matrix*

Continued

```
Image \\ the image being analyzed
Degree \\ The angle of co-occurrence being measured
Mask \\ the tissue location information
Levels \\ the number of levels in the image
for n=1:1:N \\ number of rows in the image
    for m=1:1:M \\ number of columns in the image
        if Mask( n , m ) == 1
            Feature( n , m ) = Calc_Cooc(region(Image( n , m )))
        else
            Feature( n , m ) = 0
        end
    end
end
            Calc_Cooc(region(Image( n , m )))
Temp \\ the region being analyzed
Co_Matrix \\ a Levels x Levels matrix
Feature_output \\ the feature output, whichever one it may be
for i=1:1:Levels
    for j=1:1:i,
        \\ only calculate up to the column that matches the current
        row to take advantage of the symmetric nature of the
        co-occurrence matrix
        Co_Matrix( i , j ) = Co_occurrence( i , j , Temp )
```

Continued

```
                Co_Matrix( j , i ) = Co_Matrix( i , j )
        end
end
Feature_output = Statistics(Co_Matrix)
\\ the feature_output may be a vector or a scalar
```

FIGURE 3.7 The pseudo-code implementation of the feature extraction based on the co-occurrence matrix.

radiodensity. But the main purpose of this thesis was to take a different approach to analyzing the images. The scenario where the experimental percentage was close to Dr. Celia Byrne's but didn't make any sense visually wanted to be avoided at all costs. The results for this procedure will be shown in Chapter 4 of this thesis.

3.1.4 Laws Texture Mask

Law's texture mask was the third implementation of texture segmentation done in this thesis. It offered a simple spatial filtering approach to the problem, unlike the other two methods introduced earlier which required some form of transformation. One of the reasons that Law's texture mask was chosen for this thesis was the many different features that could be obtained. As, shown in the background chapter, Law's mask has five different characteristics that can be convolved with one another to create a mask used for filtering. This allowed for 25 different features from the different combinations. Five of these features were the results of a filtering procedure using masks created by the convolution of a Law's texture characteristic by itself. For example, as shown in Figure

3.8, the gray level characteristic convolved with itself created the mask shown. As can be seen by the resulting mask, the results from these masks were rotation invariant so did not have to be added to any other result to make it so. The other 20 results obtained from the

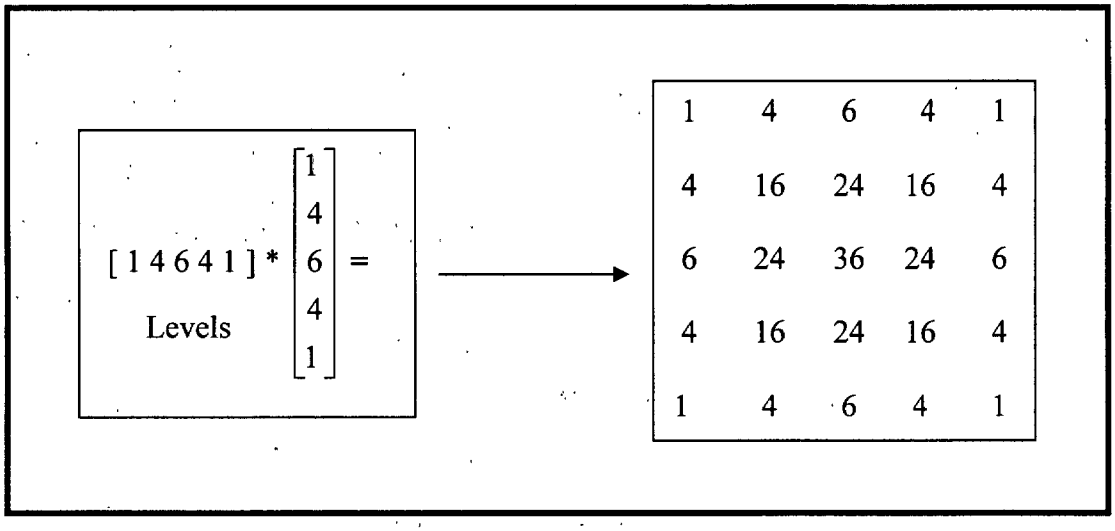


FIGURE 3.8 The mask created from the convolution of one Law's Texture Characteristic by itself.

masks of the inter-convolution of 5 characteristics had to be added together with it's corresponding rotated mask result for it to be rotation invariant. For example, the result from the Edges-Ripples mask had to be combined with the Ripples-Edges mask to make it rotation invariant. With the combinations, there were a total of 15 features that were then sent into the automated clustering algorithm, K-means. The results for this procedure will be shown in Chapter 4 of this thesis.

3.2 Other Image Processing Methods

Various other texture methods were experimented with after the initial three shown previously in Section 3.1. Because of the results obtained from the first three as well as

some preliminary results obtained from other methods, it was decided that textures was not the right path to follow. All results showed that the information contained within the two regions were similar to each other, and could not be discernable by anything else other than the gray levels. At this point, this thesis started investigating other approaches to this problem, mainly ones that emulated the thought processes of the radiologist involved. Rather than attempting to create an automated segmentation process based on statistics that weren't even considered by the radiologist, it was decided that it would be more ideal to find a method that closely followed the steps that were taken by the radiologist. It was not until the final stages of investigation that an algorithm was created that indirectly followed methodology taken by the radiologist. This portion of the procedure will detail all the intermediate steps of experimentation that led to the final implementation of the Local Contrast Estimation algorithm.

3.2.1 Non-Linear Transformations

Non-linear transformations were investigated not as a method for automated segmentation, but rather a method to preprocess the images. The images were sent through a power transformation to discern if it would help in analyzing the images. The theory behind the transformation was that since this thesis was working with images that had regions that needed to be distinguished by their perceived brightness, a better separation between the brighter and darker levels would make segmentation easier. For example, let us examine a vector of numbers, I_x that was used for demonstration. I_x contained 10 numbers ranging from 2 to 7 and was sent through a simple power transformation to the 4th, as shown in Figure 3.9.

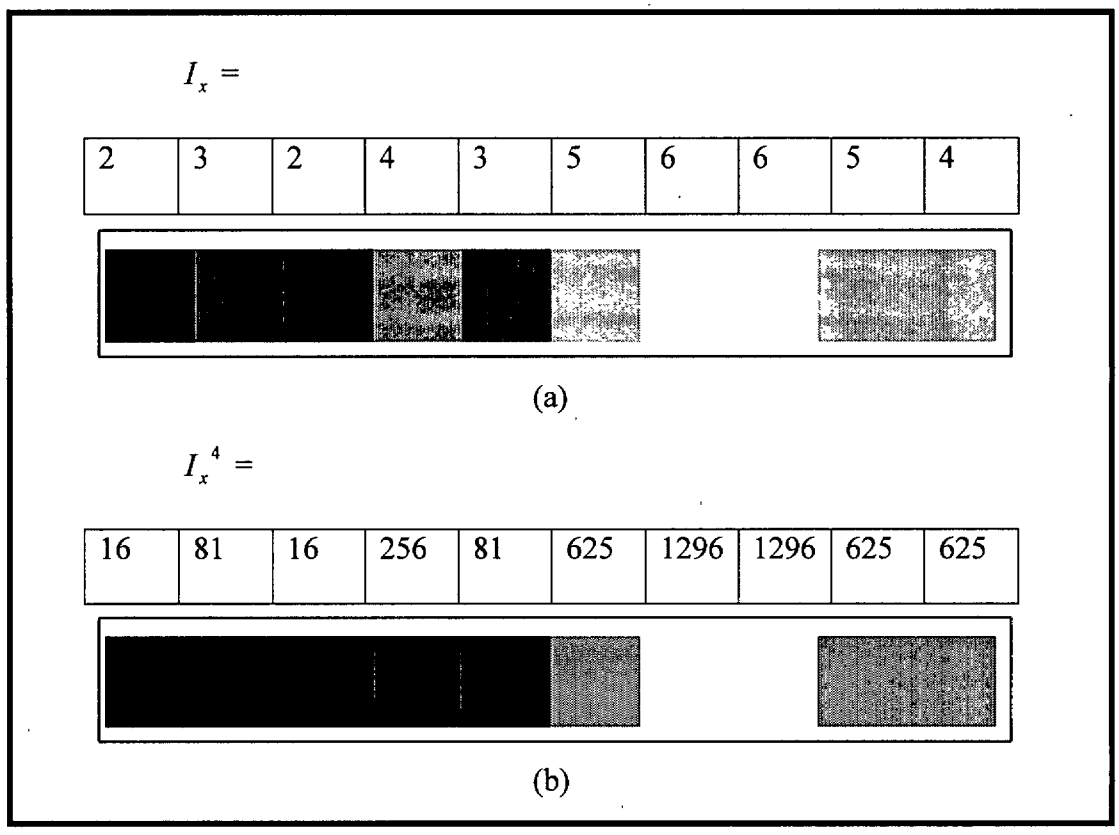


FIGURE 3.9 (a) Vector I_x and its pictorial representation. The image has been normalized before displaying. (b) Vector I_x after it has gone through a power transformation.

Let us assume that right half of the vector I_x was considered radiodense while the left half was considered radiolucent. The goal of the power transformation was to reduce relative brightness of the radiolucent region in relation to the radiodense regions. As Figure 3.9 (b) shows, this was the case. One of the major problems with this method was if there was a single pixel that was significantly brighter than the rest, which could have been caused by errors during the X-ray process or the scanning process, the power transformation readjusted the image so that only single pixel was bright, since the other areas of the image were suppressed. Because of these problems, the use of a median filter was introduced into the procedure. To further reduce the problems discussed, a

regional power transformation was applied to the images as well. The results for this procedure will be shown in Chapter 4 of this thesis.

3.2.2 Gray Level Connectivity

Gray level connectivity introduced an interesting idea for creating a solution to the problem of identifying the amount of radiodense tissue in a digital mammogram. Rather than just taking the gray level information or trying to obtain a statistic based on texture to classify the various components, in allowed for a more visual approach to the problem. Even thought the method was not successful in terms of segmentation of the radiodense tissue, it introduced many ideas that would later be used in the final algorithm.

One of the main problems with using a histogram based approach to the problem was that it did not take into account the regions and patches that the groups of pixels formed. By definition, radiodense tissue is comprised of all 'interconnected' components, and therefore it is highly intuitive to look at the mammograms in the same manner. For this reason, a radiologist does not look at the individual pixels separately; but rather as large groups with similar brightness, when the determination of radiodensity is made. For example, Figure 3.10 shows an image of a mammogram from the Harvard set as well as the mask of the radiodense region it created after it was segmented. The region defined by the segmentation mask created a region of radiodense pixels. Rather than being scattered throughout the image randomly, most regions exhibited some sort of pattern when it came to the location of the radiodense pixels. All mammograms from the test database followed the same pattern.

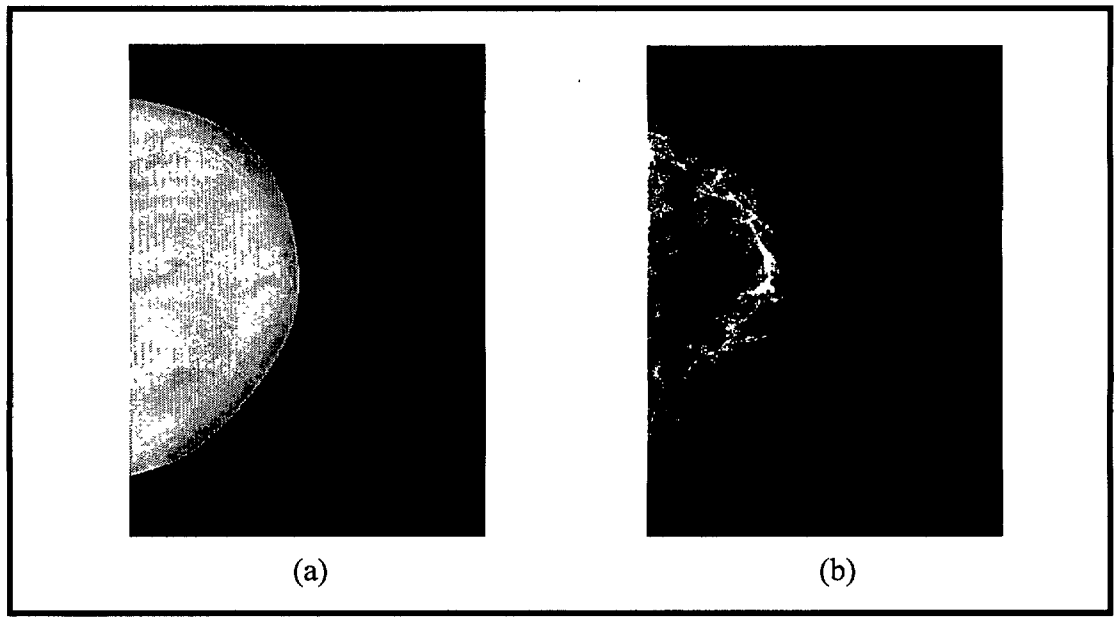


FIGURE 3.10 (a) Image with a radiodensity percentage of 1.5 percent from the Harvard database. (b) The basic shape or region of the radiodense tissue after segmentation. Aside from small portions, most regions are grouped.

Before the procedure for the gray level connectivity is explained, let us investigate another example of why the grouped regions will play the most important role in determining what regions are radiodense and radiolucent. Figure 3.11 demonstrates two examples that were created for this example. Each image contained same percentage of bright pixels and the same percentage of dark pixels. If we look at the images as mammograms and analyze them as such, it is easy to understand that it is highly unlikely that these two images would produce the same result, which would be 50% radiodensity. A radiologist would most likely have classified (a) as either a 100% radiolucent or a 100% radiodense image (again, this further shows that the analysis of the mammograms is a highly subjective procedure and could be different based on the radiologist), while classifying (b) as a 50% radiodense and 50% radiolucent image. Even though these two

classifications would be different to a radiologist, a histogram approach would disregard the regional information and only analyzed based on a total gray level pixel count.

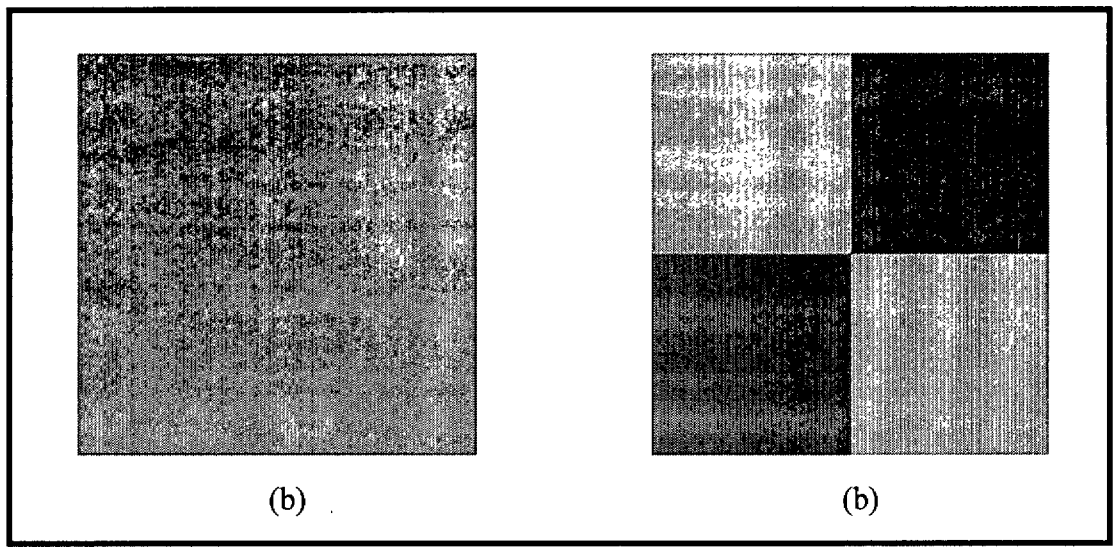


FIGURE 3.11 (a) An image made from equal amounts of dark and light pixels, but arranged in a checkerboard manner. (b) The same amount of bright and dark pixels as (a) but grouped.

The gray level connectivity approach was only implemented on a few of the images for the sake of evaluation. As will be shown in Chapter 4, the algorithm was too simple to take into account the amount of separate regions that some images contained. These separate regions still had to be classified based on their relative gray levels and that still led to the problem of establishing the correct threshold.

3.3 Local Contrast Estimation for Threshold Selection

There were many problems associated with the previous methods, the Constrained Neyman Pearson (CNP) [40] and the Spatially Varying Constrained-Neyman Pearson (SV-CNP) [41], that were implemented for this research:

1. The images were segmented from algorithms that assumed that the histograms of the mammogram were bimodal in nature, when in fact very few of the mammograms in the database followed that trend.

2. There were many portions of the algorithm that were based on user feedback rather than images statistics. These characteristic prevented the algorithms from being fully automated.

3. The algorithms' success was only based on the final numerical output of the segmentations, rather than a combination of visuals and percentages, this introduced false positive results.

4. Many of the images were manually filtered; this introduced another step that made the algorithms non-automated.

This thesis attempted to address many of these problems by introducing a new approach to the problem. Rather than taking components from the previous algorithms and adding new idea to them, the Local Contrast Estimation algorithm introduced in this thesis was created from scratch and therefore each step will be explained in detail. This section will outline this procedure as well as reviewing many of the problems that were inherent to the previous methods. Figure 3.12 shows a block diagram of the new algorithm introduced in this procedure.

After being scanned and digitized, the mammograms were preprocessed to get remove noise introduced during the procedure. After, the noise reduction, the tissue segmentation mask was created using a new method. After tissue segmentation, the

image was adjusted using a new compression mask. Finally, the Local contrast estimation of the image was done to find the correct threshold to use.

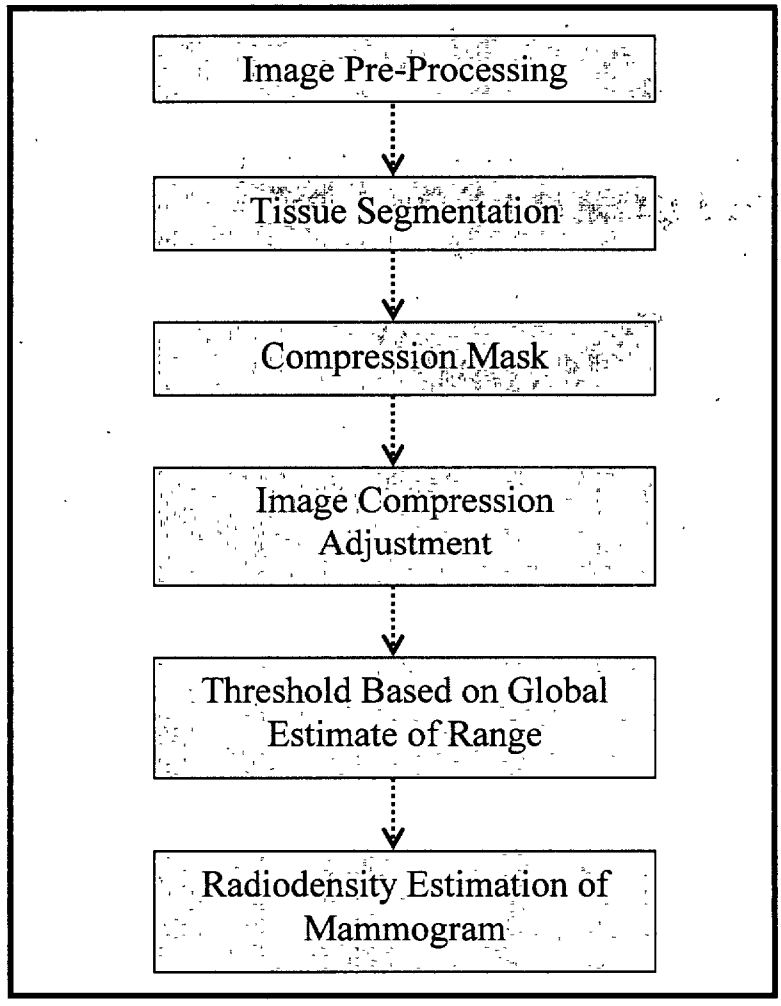


FIGURE 3.12 The block diagram of the Local Contrast Estimation procedure.

3.3.1 Evaluation of Previous Methods

One of the main problems with the previous algorithms was their tendency to base the accuracy of the algorithm on only the percentages, disregarding all the information that a visual representation of the segmentation would give. What this showed was that several parameters could be added to the system, making the test percentages very accurate for

the present database at that time, but the algorithm had a very hard time predicting the percentages of new images. For example, the CNP threshold was computed by the author by first obtaining a preliminary Bayesian Threshold; which introduced its own problems because the histograms of the images did not follow a bimodal trend, and this threshold was adjusted by several parameters to allow for a Neyman Pearson Threshold. The equation used in the procedure is shown below:

$$T_{CNP} = \frac{\mu_1 + \mu_2}{2} + \left(\frac{\alpha - \sigma^2}{\alpha}\right)\left(\frac{\mu_2 - \mu_1}{2}\right) \tag{3.1}$$

In this equation, values μ_1, μ_2, σ, were all obtained from the image statistics, and therefore allowed for the calculation through an automated process. Unfortunately, the α parameter was based solely on the author's input and therefore negated any chance of the algorithm being automated. Therefore, an analysis of general performance was done using the all datasets based on the parameters obtained from only the Harvard dataset.

The segmentation in the SV-CNP was done by adjusting the threshold obtained from the CNP and adjusting it based on the location of the pixel. This introduced several problems into the procedure. First, by using the CNP as a part of the method, this still left that supervised parameter within the process, and compounded the problems associated with both methods. By maintaining the α parameter and adding this new parameter based on the perceived compression, the algorithm was now burdened with several fitting variables.

The idea that a radiologist takes into account compression when analyzing a

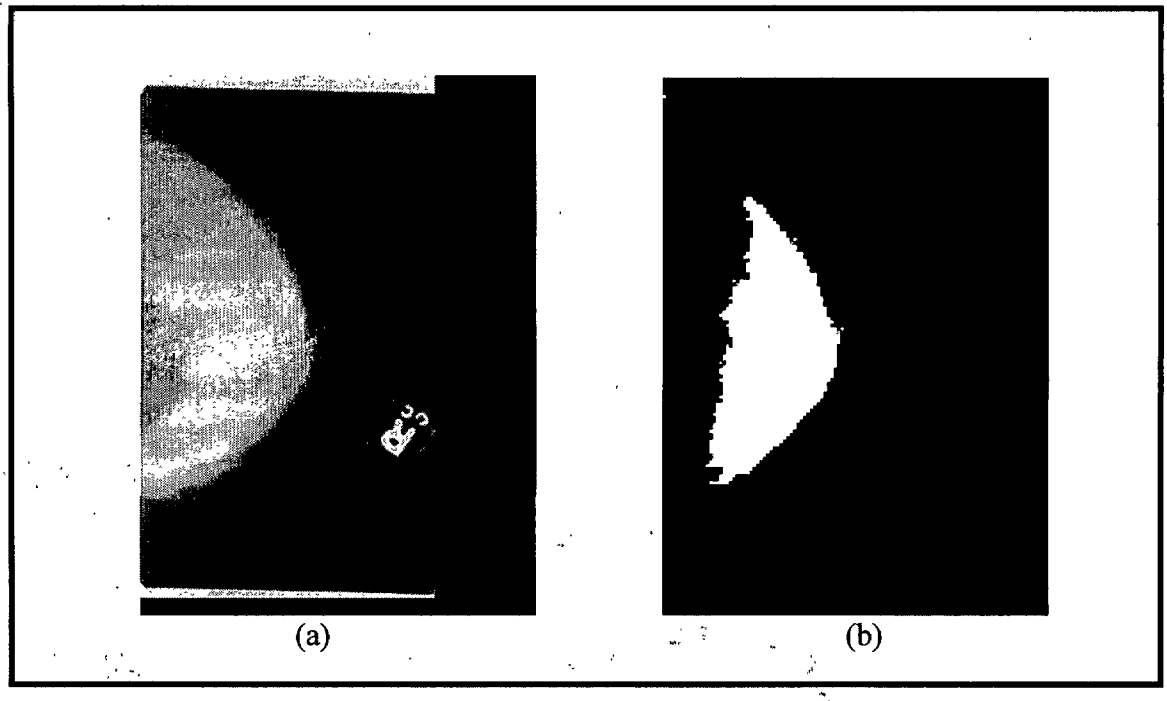


FIGURE 3.13 (a) Original Harvard image. (b) Image after segmentation using SV-CNP.

mammogram was a valid theory, but the amount can be entirely subjective based on the radiologist as well as the mammogram. The author of the SV-CNP took into account the compression without the knowledge of how much and how. Basically, the adjustments were based on the researcher's assumptions rather than actual physics. This extra parameter was a way to fit the results without worrying about the basic heuristics of the final image. As Figure 3.13 will show, even though the percentage obtained from the SV-CNP was close to the one given by the Toronto method, the actual segmentation image obtained from it was incorrect. The radiodensity region has been over fit due to the high amount of compensation due to the compression mask. As a result, in this thesis, an analysis of all the datasets was done with visual and percentage quantification.

Another problem associated with the previous two algorithms was that they were not very noise resistant. The proponents of both previous methods had many of the

mammograms filtered and de-noised manually before being introduced into their algorithm. Even though this sufficed for the purposes of testing, the algorithms had trouble analyzing any new image that was introduced to it; the new images had to be filtered manually as well. Because the main focus of this work was to create an algorithm that could be used on a large database, the manual preprocessing of features were not acceptable. This thesis created many features within the coding that made it very resistant to noise as well as several filtering procedures that made sure that any image analyzed would not suffer from noise problems.

3.3.2 Image Preprocessing

In an ideal world, all images would have been scanned in the same way every time and there would have been no problems when it came to noise in the overall system. Unfortunately, that was not the case and therefore the images always contained some sort of noise due to inconsistent scanning and as well as sensor problems. This first stage of preprocessing was designed to remove some of the major noise contained within the image that would make it difficult for the masking algorithm to segment the tissue from the X-ray. Figure 3.14 demonstrates how many of the mammograms looked before preprocessing.

The major problem with some of these mammograms was the white borders found within the edges. This type of error was created during the creation of the mammograms and therefore could not be controlled during the scanning or the analysis portion of the procedure. These edges varied between thicknesses of just a few pixels to

several hundred pixels depending on how the mammograms were created. As long as the breast region and the white stripes were not connected, this issue did not cause any major

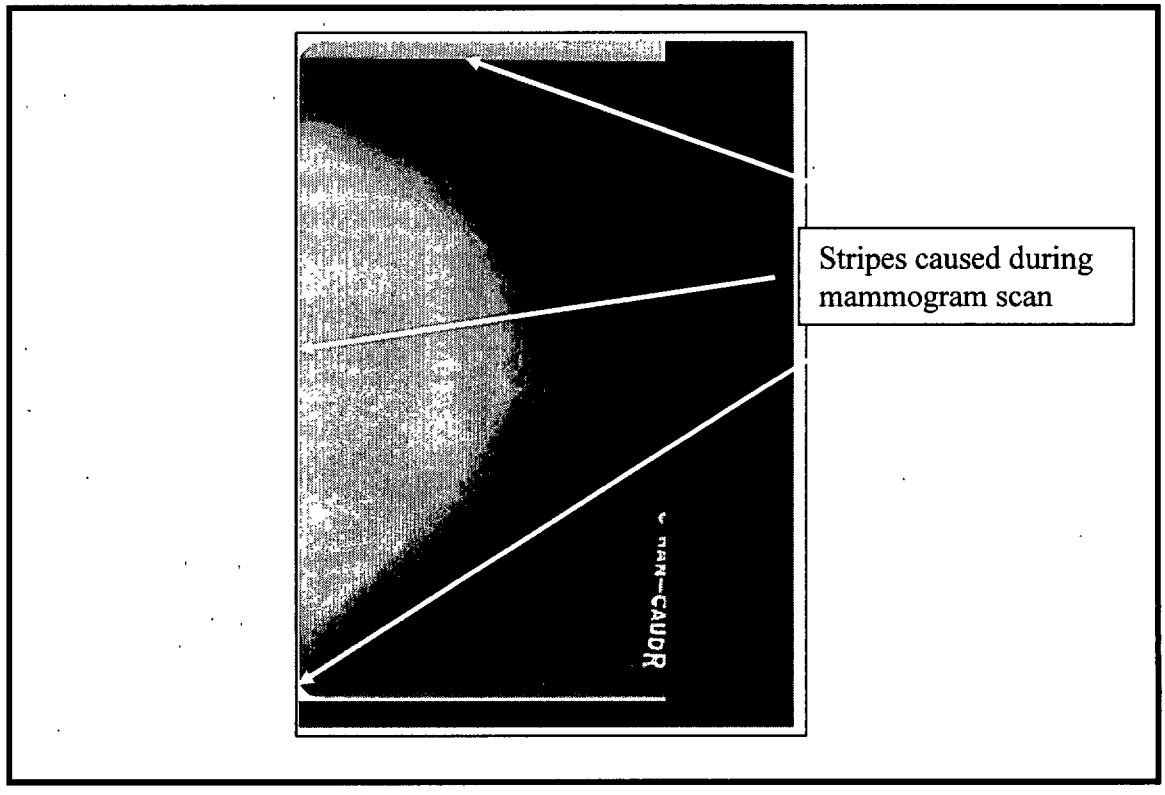


FIGURE 3.14 Raw scanned mammogram image.

problems in the implementation of the masking algorithm. But in some cases, the white stripes were large enough that they were connected to the tissue region. This caused the tissue segmentation algorithm, which will be discussed later in this chapter to perceive this region as tissue rather than X-ray region due to the large amount of bright pixels. Depending on the size of this error, the final percentage given for the radiodensity could have been off by a considerable amount.

To remedy this problem, an algorithm was created to remove as many of the stripes as possible. A pseudocode representation of the algorithm is shown in Figure

3.15. The algorithm was implemented to take into account many of the characteristics that all mammograms had.

First, the algorithm rotated the mammograms so that the tissue was always present on the left half of the image. This was chosen as the standard when the research started before this thesis and therefore all algorithms were created based on it. The rotation was based on some statistic of *A,* which was one of 4 regions, where the regions were defined as 50% blocks of the image from the edge. An example is shown in Figure 3.16. First the regions with the shortest length were disregarded as being as the portion desired. With the remaining two regions, the region with the highest average was chosen as the left portion of the image. This portion of the algorithm was added to make sure that all images were correctly oriented whether or not they were digitized properly.

Mammogram *\\ the mammogram being analyzed*

Mammogram = Orient(Mammogram) *\\ the mammogram is rotated so that it is always has the same orientation*

Threshold = statistics(left_half_mammogram) *\\ a initial threshold is selected based on the left half of the mammogram. 1.5 Sigma from the mean is chosen as the threshold.*

Temp=imbw(Mammogram_Threshold) *\\ the mammogram is turned into a black and white image based on the threshold obtained.*

for i=1:1:length(rows) *\\ analyze each column of the image*

White(i) = Connected_white(Temp(i))

Continued

```
\\ the last pixel of the first group white pixels in chosen as the value
end

Transition=Valley(White) // the two valleys of the function White, one from
the left and one from the right is chosen as the boundary points. All
information before the first point and after the second point is zeroed.
```

FIGURE 3.15 The pseudocode implementation of the stripe removal algorithm.

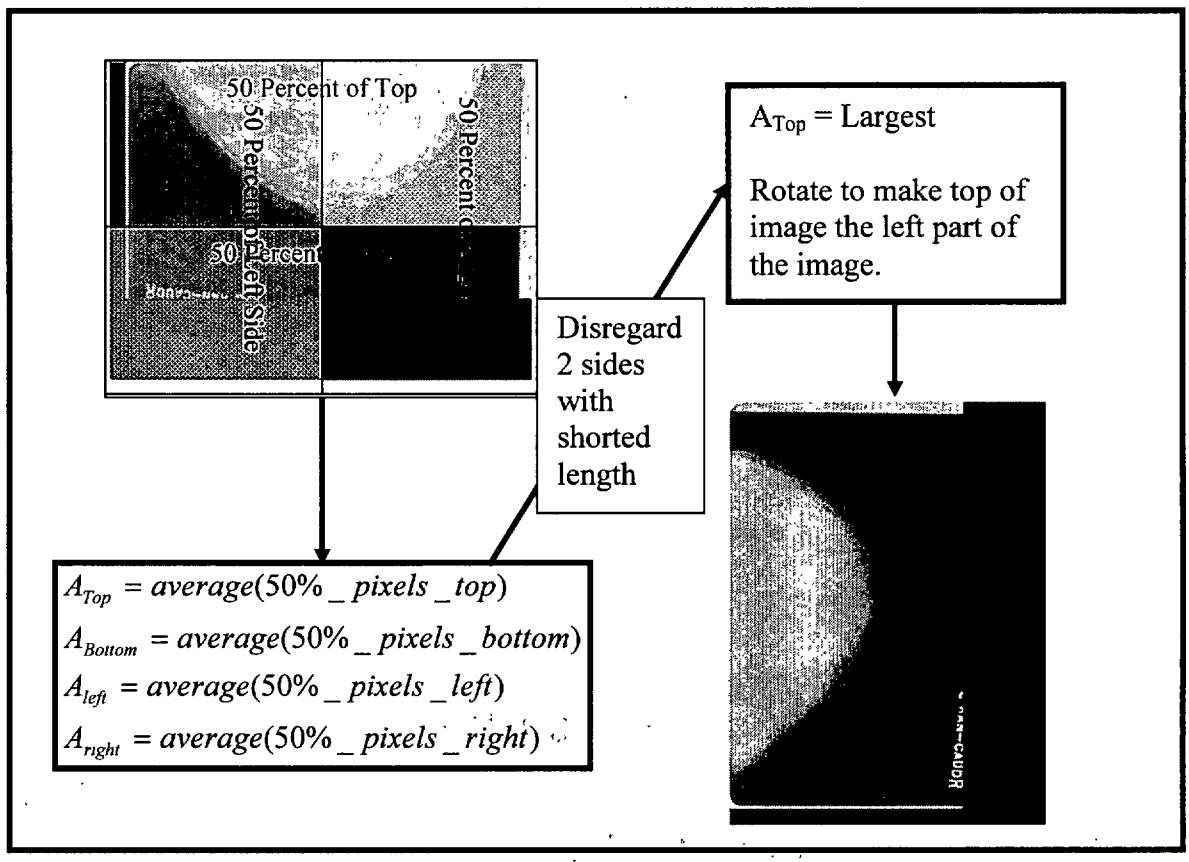


FIGURE 3.16 Detection of image position for proper rotation.

Once the mammograms were oriented correctly, the standard deviation and the mean of the left 50% of the image were obtained. This information was used to obtain a preliminary threshold, which was chosen as the point 1.5 sigma away from the mean, for a preliminary segmentation of the mammogram. The ideal here was that since the stripes located on the mammogram were almost always the brightest pixels on the image, this threshold would segment out a majority of non-stripe regions from the image, thus allowing the next portion of the algorithm to locate regions this regions for removal.

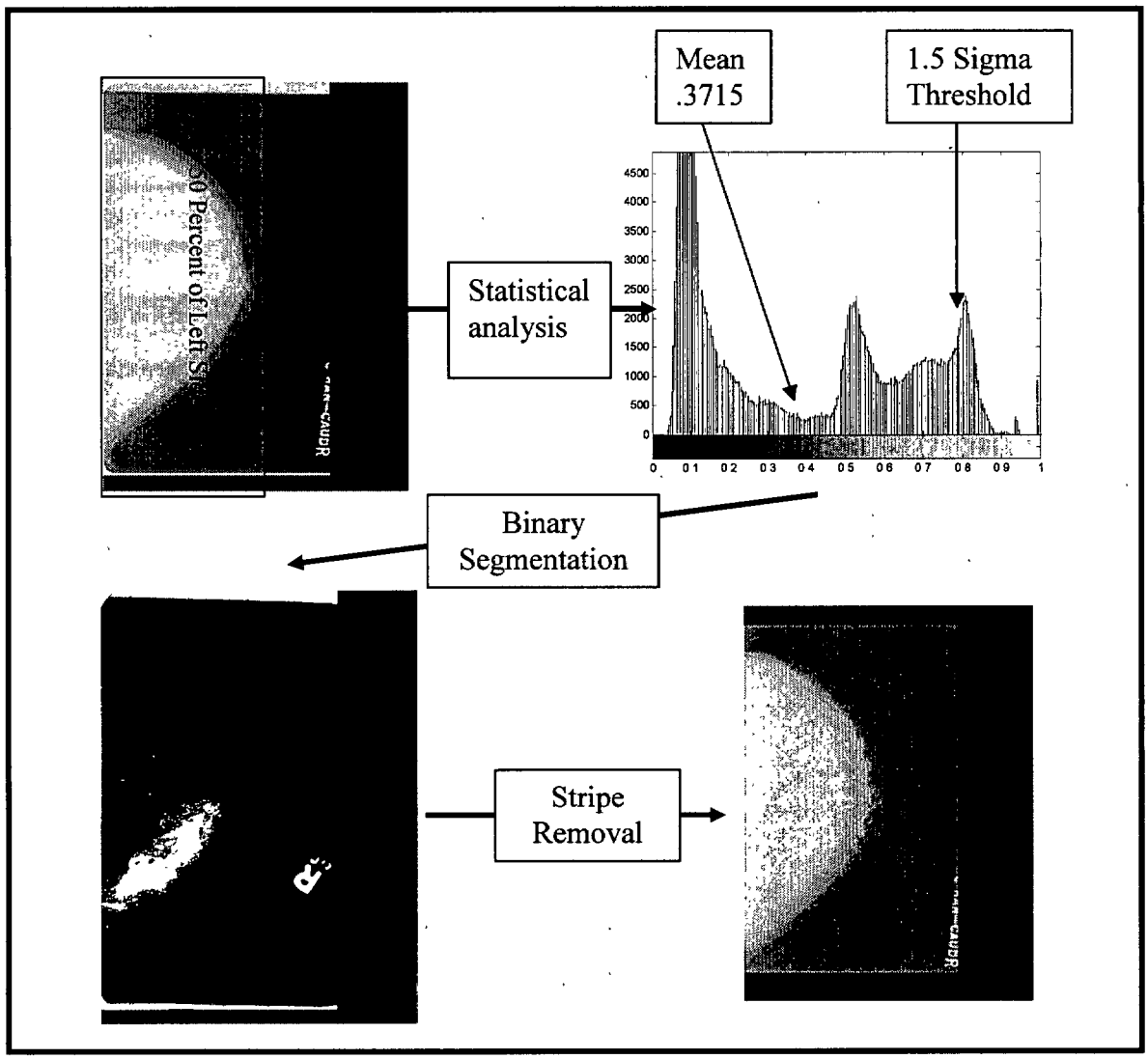


FIGURE 3.17 The preprocessing procedure performed a basic stripe removal procedure on all mammograms before tissue segmentation.

Once the image was segmented by the threshold obtained, each column was analyzed for the first group of white pixels. The last pixel of these groups was used for the determination of the valleys. Anything above the uppermost valley and below the bottommost value was zeroed out. Figure 3.17 shows an example of this procedure. This procedure was not expected solve all problems associated with the stripes. Therefore, other noise resistant features were added to the segmentation mask introduced later on. Other noise problems of the mammogram were taken into account during the implementation of the other parts of the algorithm.

3.3.3 Tissue Segmentation

As stated before in this thesis, there were many problems when dealing with the mammograms obtained for the database. These images were very noisy and therefore, any algorithm created for the analysis of radiodensity had to be created so that it could work in the worst case scenarios. One of the major changes and advances that this thesis attempted to make was to make the code very robust and resistant when handling these types of situations. In this thesis, both image processing and image enhancement techniques were used to create a more defined and detailed mask than was created in the previous versions of this research. The previous algorithms for masking were not very robust and needed supervision for obtaining the correct results. Because the main goal of this research was to create an automated algorithm, this was not acceptable.

First let us look at what done previously for the tissue segmentation algorithm and why it was inadequate for the entire database. Figure 3.18 outlines what was previously done in the SV-CNP.

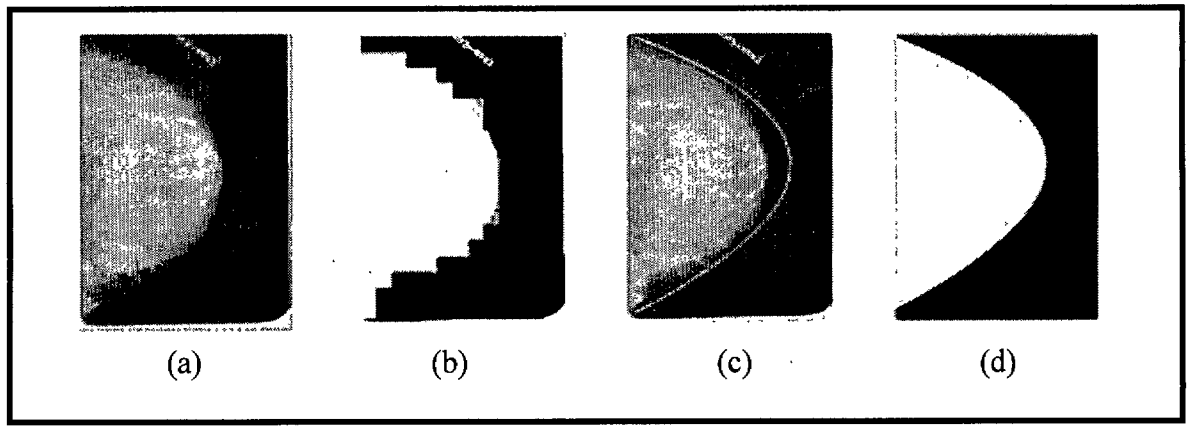


FIGURE 3.18 (a) The original image. (b) The image after a 50 x 50 bimodal histogram distribution evaluation (c) The RBF created by the using the last white box. (d) The segmentation mask created.

In the SV-CNP, the initial breast boundary was estimated by using a form of bimodal histogram estimation. The image was divided into 50 x 50 segments and tested to see if each block exhibited a bimodal Gaussian histogram or a relatively right shifted histogram, in then which it was considered part of the tissue and was assigned a value of 1. If the histogram was unimodal and further to the right compared to the other histograms, the block was assigned a value of 0. Figure 3.19(b) represents the histogram of the black bounded region shown Figure 3.19(a), which was a region located well within in the breast tissue. It can be seen that the region exhibited a unimodal Gaussian distribution as well as being centered more to the right of the histogram. In the region defined by the white bounded box in Figure 3.19(c), which was an area where the tissue met the rest of the X-ray, the corresponding histogram in Figure 3.19(d) showed that the region was more bimodal as well as being centered more to the left of the histogram. Once this was done for all 50 x 50 segments in the image, the image shown in Figure

3.18(b) was obtained from the procedure. Since this was a rough estimate of the breast edge, the edges were smoothed by using an RBF network.

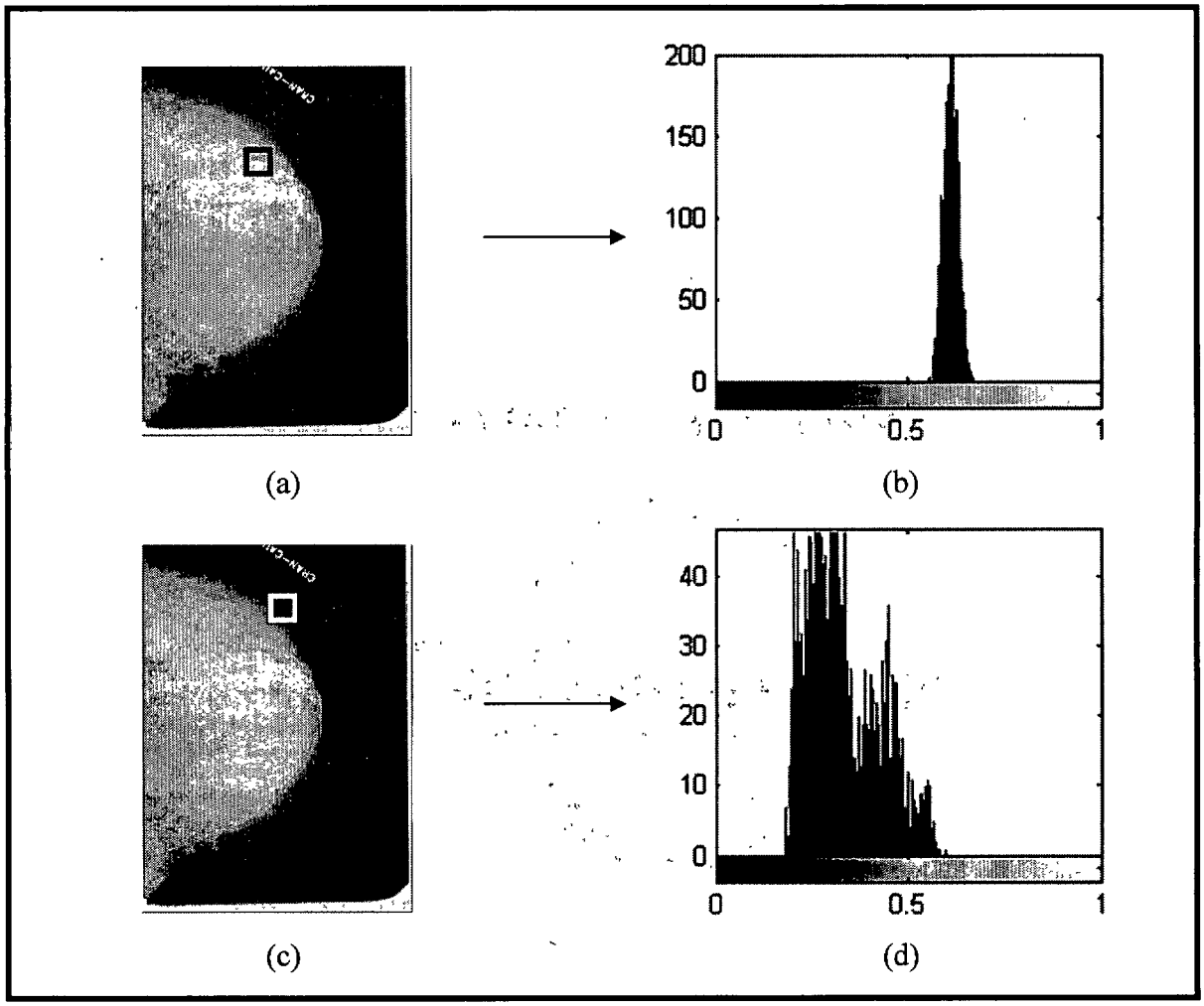
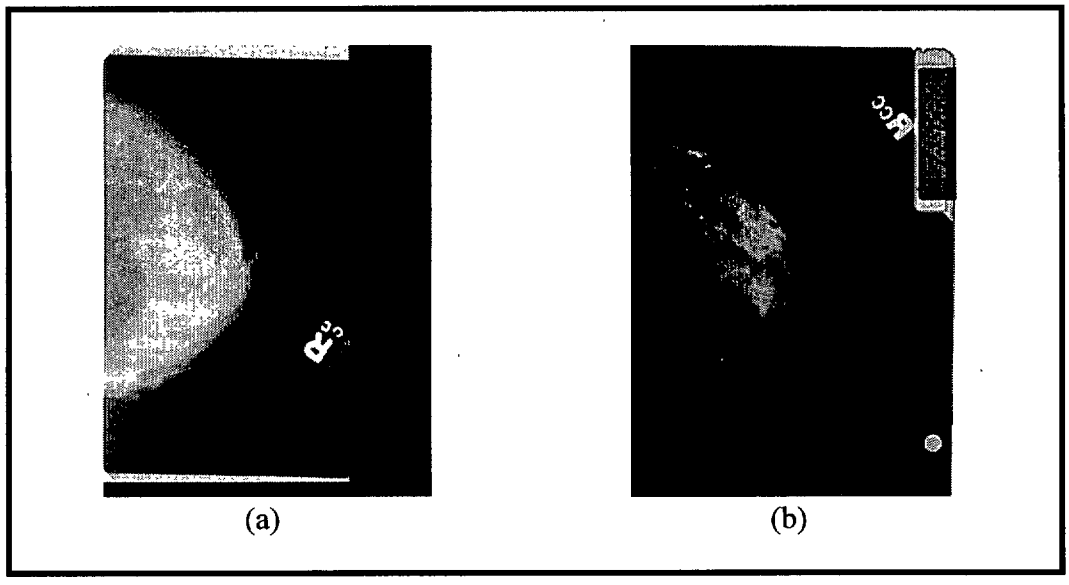


FIGURE 3.19 (a) A region located within the tissue. (b) The histogram of the tissue region. (c) A region of transition between the X-ray and tissue. (d) The histogram of the corresponding region.

This method made several assumptions that caused it to fail when new images that it wasn't tested with were added to the database. First, it assumed that all mammograms would exhibit the differences in terms of the tissue regions and the X-ray region. Unfortunately, this was not the case. Even in the datasets used for validation in

this thesis, there were two sets of mammograms, the Harvard Set, and the mammograms obtained from Fox Chase Cancer Center, and they were vastly different from each other.



(a) (b)

FIGURE 3.20 (a) Mammogram from Harvard database. (b) Mammogram from FCCC database.

Figure 3.20(a) shows a typical image from the Harvard Database. The X-rays from the Harvard set were very detailed in the distinction between the tissue region and the X-ray region. In other words, the overall means of the X-ray region and the breast tissue were significantly different from each other. Unfortunately, when dealing with the images in the FCCC database, as shown in figure 3.20(b), this was not always the case. The X-rays from the FCCC database all had several problems associated with them. First, unlike the Harvard sets, the FCCC had the issue of having very low means of the radiolucent regions. This made it very difficult to distinguish between the tissue region and the radiolucent regions. Second, many of the X-rays contained white stripes that ran down the middle of the X-ray, as shown in Figure 3.21. These stripes in many cases were

brighter than many of the regions on the breast tissue itself. If these stripes were not connected to the breast tissue itself, it did not cause too many problems. Unfortunately, many of the stripes were large enough and close enough that they ended up entering the tissue region, and therefore the previous algorithms had trouble creating accurate masks.

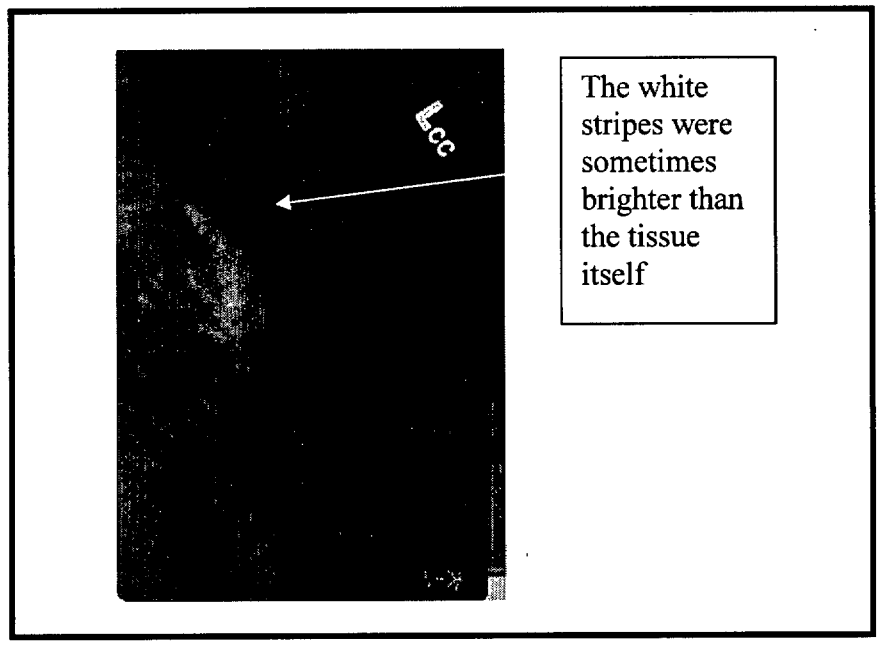


FIGURE 3.21 The images from the FCCC database had problems with stripes running through the image.

Because of these problems, the previous method produced many tissue masks that were off quite significantly, which would have introduced a high percentage of error in the final evaluation. The new tissue masking algorithm demonstrated in this thesis attempted to address these problems by being very adaptive and resistant to noise.

A. Image Enhancement for Tissue Segmentation

There were five major steps to the image enhancement portion of the tissue segmentation process:

1. Histogram stretching (or equalization depending on the image)

2. Binary transformation

3. Morphological operations,

4. Labeling of connected components

5. Extraction of the final binary image for use in the estimation of the breast edge.

The histogram stretching/equalization were performed so that the binary transformations could be done using the same thresholding techniques for any image. The binary transformation was done using either Otsu's method, applied on images obtained from medical grade scanner (optical density greater than or equal to 3.85) or using an experimentally determined method, involving the derivative of the image's histogram to find the optimal threshold, which was used on images obtained from a standard scanner (optical density less than or equal to 3.85). In the second method, the gray level at which the difference in pixel counts above an experimentally determined gray level value was less than or equal to a chosen value was used as the binary threshold for the enhanced image (for the images this has been tested on, the chosen difference in pixel count was 250, and minimum gray level was 0.5, with 256 gray levels ranging from 0 to 1). Figure 3.22 shows the histogram of a mammogram to which this second technique was applied to, its derivative and the optimal binary threshold for the image.

Next, morphological operations, such as erosion and dilation, were performed in order to better separate the breast from the other objects present in the mammogram (patient information, view, white borders if the image was not cropped after scanning). In the final step, the different objects in the enhanced image were labeled, and only the

connected component containing the largest number of member pixels was retained. An example of a typical mammogram during enhancement is shown in Figure 3.23, where the order of operations from left to right, top to bottom was: original image, image after histogram equalization, binary transformation using the histogram derivative method of thresholding, image after erosion/dilation, labeling of connected components, final enhanced image.

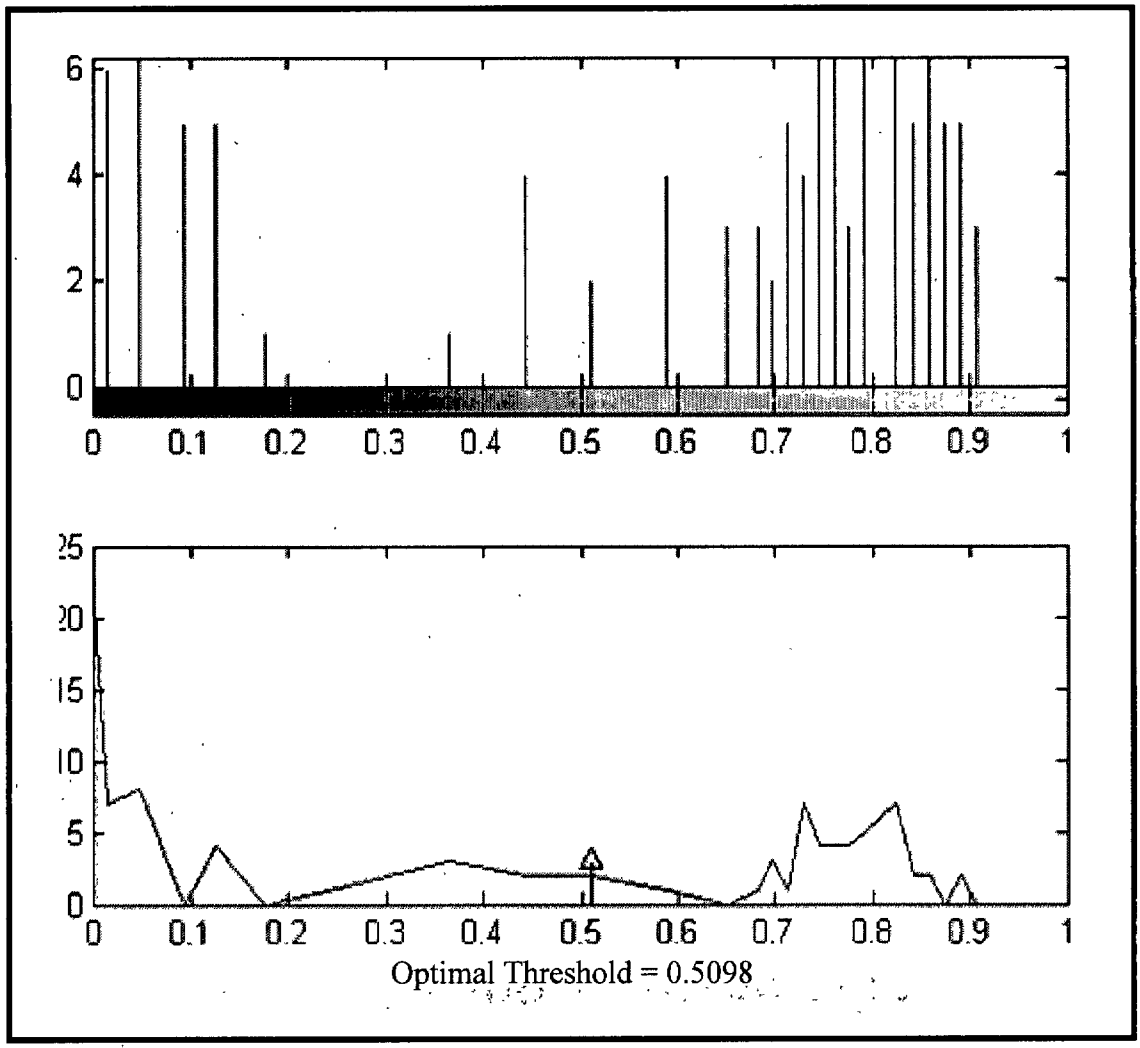


FIGURE 3.22 Histogram of a mammogram, its derivative and optimal binary threshold.

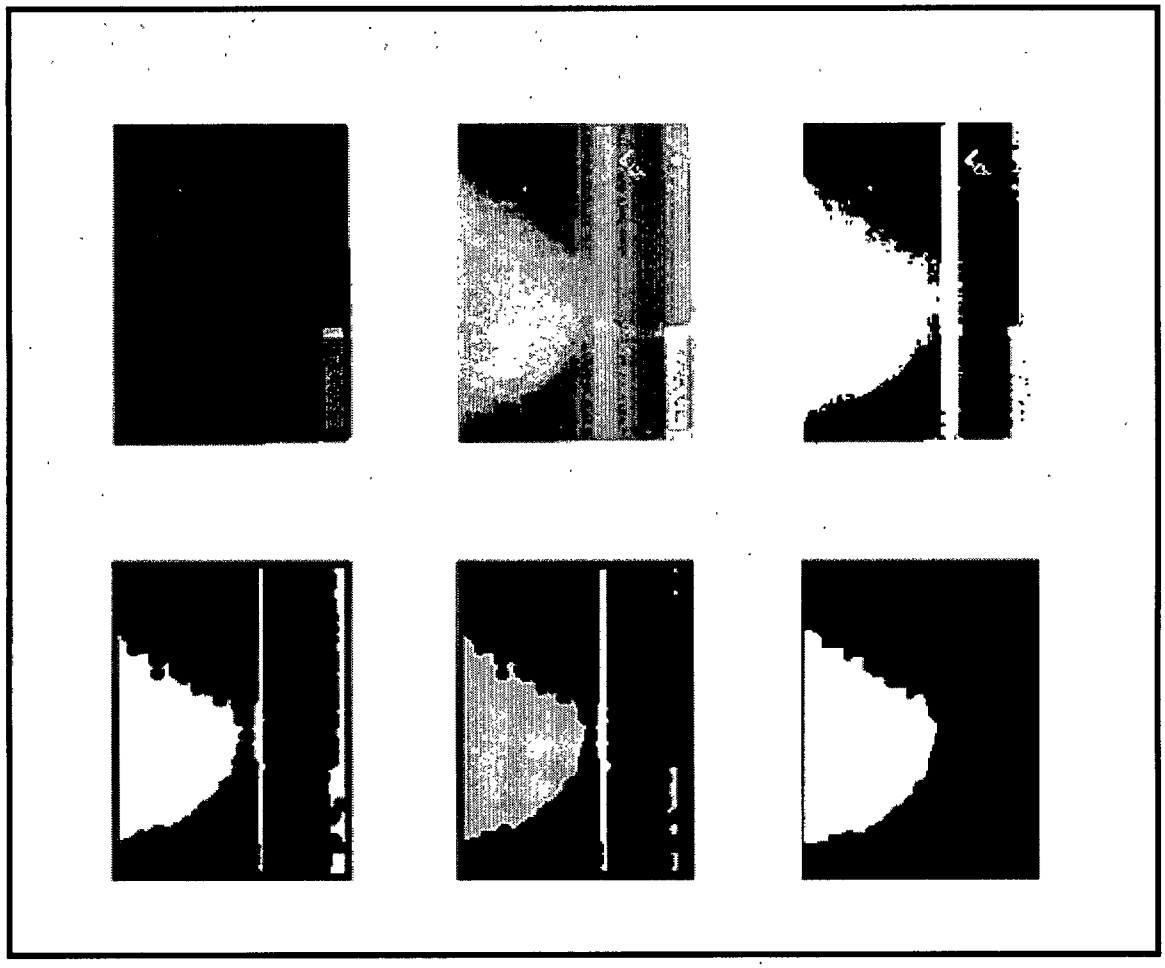


FIGURE 3.23 Typical mammogram scanned using a non-medical grade scanner subjected to the image enhancement process.

A. Generalized RBF Mask for Tissue Segmentation

The next step in the algorithm was to generalize the mask created through the image enhancement procedures, which would hopefully remove the noise that has been introduced throughout the procedure. First, the algorithm converted the 2-dimensional image obtained from the enhancement procedure into a 1-dimensional function that represented the preliminary edge of the breast. This function was then averaged by groups of four, which gave an initial generalized function of the breast edge. Taking into account the actual size of the images create by the image enhancement method, 400 by

300 pixels, several different sizes were experimented with. The assumption was that the breast edge could be looked as a very low frequency function and any high frequency component would most likely have been due to noise introduced into the system. From experimental analysis, it was shown that groups of four provide the closest results. It was from this averaged function, that the final breast edge was calculated through an iterative function that incorporated RBF networks, which is explained next.

The heuristics of the breast as well as "several real world scenarios" enabled several educated. First, as was discussed earlier, the breast edge would not vary erratically because it was a very low frequency function. This allowed any high frequency changes to be viewed as noise. Second, overall both sides of the breast would always be somewhat symmetric to each other due to the circular nature of the breast. Based on these assumptions, the RBF mask was created to be an iterative function where the breast edge was determined pixel by pixel rather than as a collective group in a single run. A flow of the RBF procedure that was created is shown in Figure 3.24.

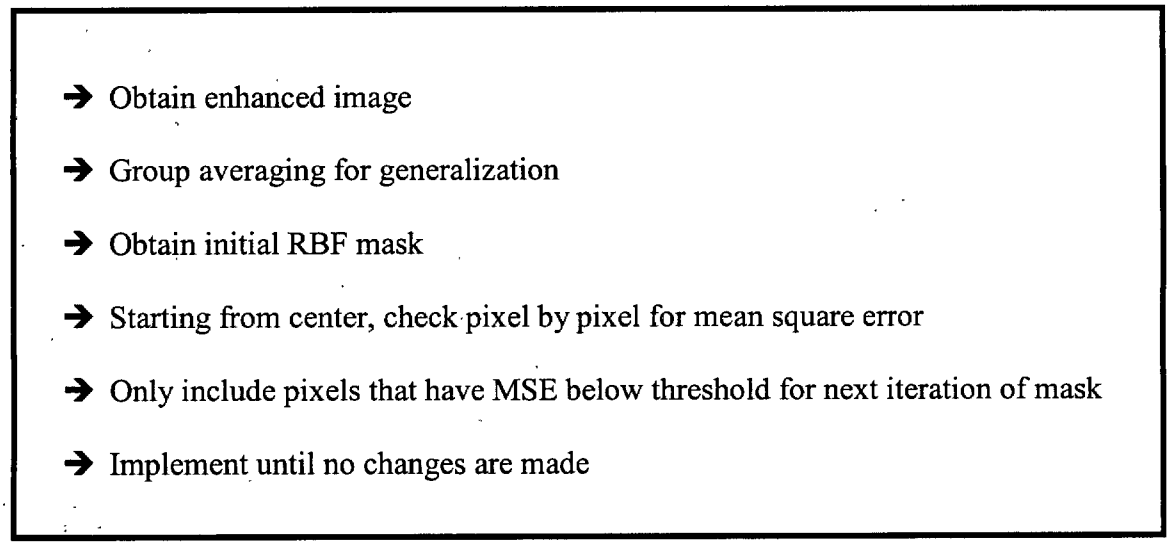


FIGURE 3.24 The flow of the iterative RBF estimation of the breast edge.

There were several conditions that were introduced into the iterative RBF network. First, the initial five points and last five points were always set at zero. This condition was based on the fact that all of the breast tissue had to be located on the X-ray; otherwise, any radiodensity percentage would not be accurate to begin with. The zero bounding also accomplished a second function by preventing the neural net from overestimating at the edges.

As the results will show, this method for tissue segmentation was more robust compared to the previous versions.

3.3.4 Compensation for Tissue Compression

Compression adjustment was based on an idea that was introduced in the previous research done at Rowan University. The previous method created is shown in Figure 3.25. In the compression model introduced by the SV-CNP technique, there were several problems. First, and most importantly, the model of compression decay in the algorithm was based on assumptions rather than actual physical models. This problem has not been addressed in this thesis for several reasons:

1. Obtaining physicals models of breast tissue and observing compression was not in the scope of this thesis
2. Actually testing women for compression of breast tissue was not possible considering the legal and medical issue associated.

The second problem was the actual implementation of the compression algorithm. The method in which the algorithm was implemented contained several components that

were not very accurate and added even more error to a system that was already laden with them. As stated before, the degree in which compression changes the amount of radiodense tissue is still unknown. But for algorithmic purposes, the code for compression was recreated to provide a more accurate model.

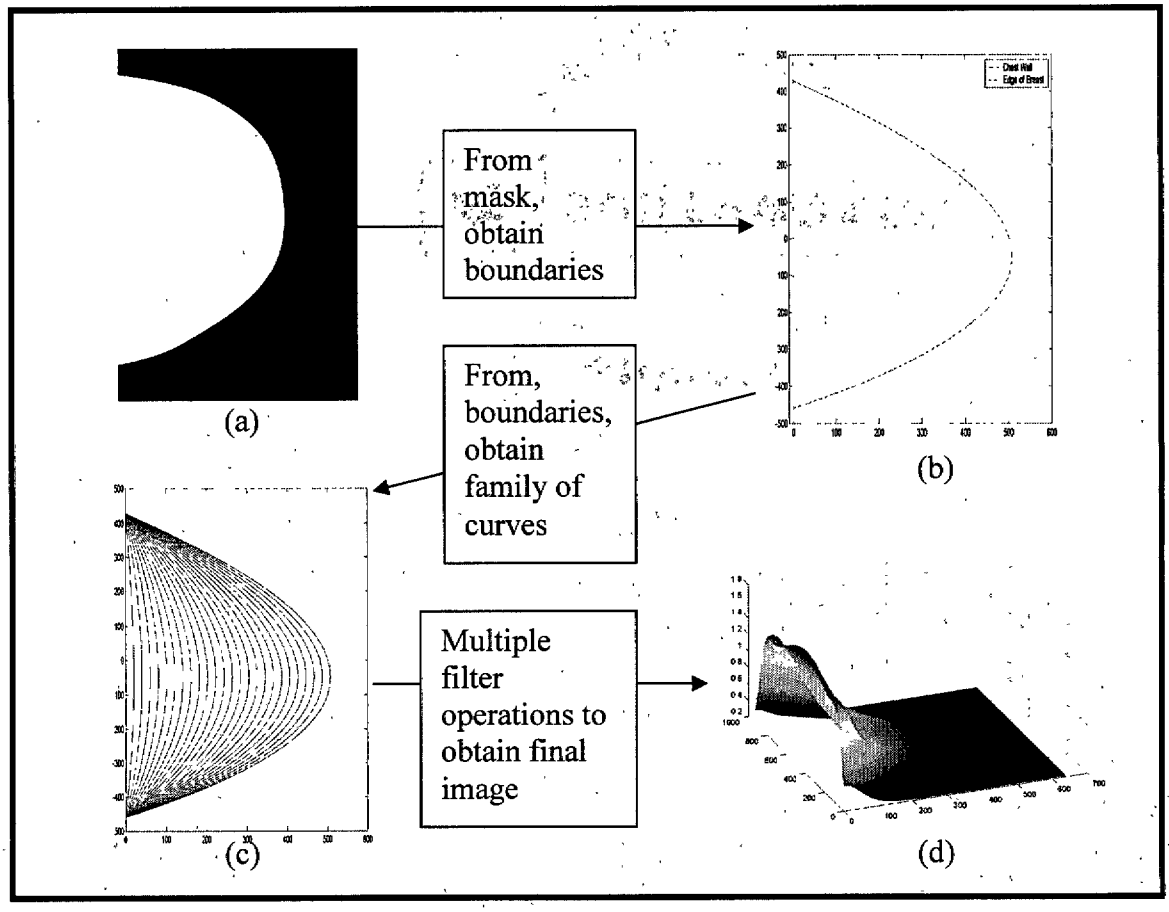


FIGURE 3.25 The compression model used in the SV-CNP. (a) The tissue segmentation mask created in a previous step. (b) The boundaries of the tissue segmentation mask. (c) The family of curves generated between the two boundaries. (d) The final compression mask obtained after multiple filtering operations.

Previously, once a high number of curves were generated for the mask, a Gaussian decay was interpolated onto the family of curves to represent the compression experienced during the scanning procedure. These curves were then filtered many times

to theoretically create a smoothed over mask that would represent the compression decay. Unfortunately, due to issues with resolution and sampling, the algorithm ran into several problems. When there were too many curves, the center was represented adequately, but the edges became saturated and a separation of lines became lost. If there were too few lines, the center of the tissue has too many non-filled pixels that would result in poor results during the filtering phase. Most of the compression mask ended up looking like the image shown in Figure 3.26. As Region A in Figure 3.26 shows, the centers of the masks were overall lower in value than the sides because of the separation of lines; which was a product of using an averaging filter for smoothing. The filtering itself caused many problems. Region B of Figure 3.26 shows how even though the values at the chest wall should all have been similar; the filtering operation introduced a false variation in values.

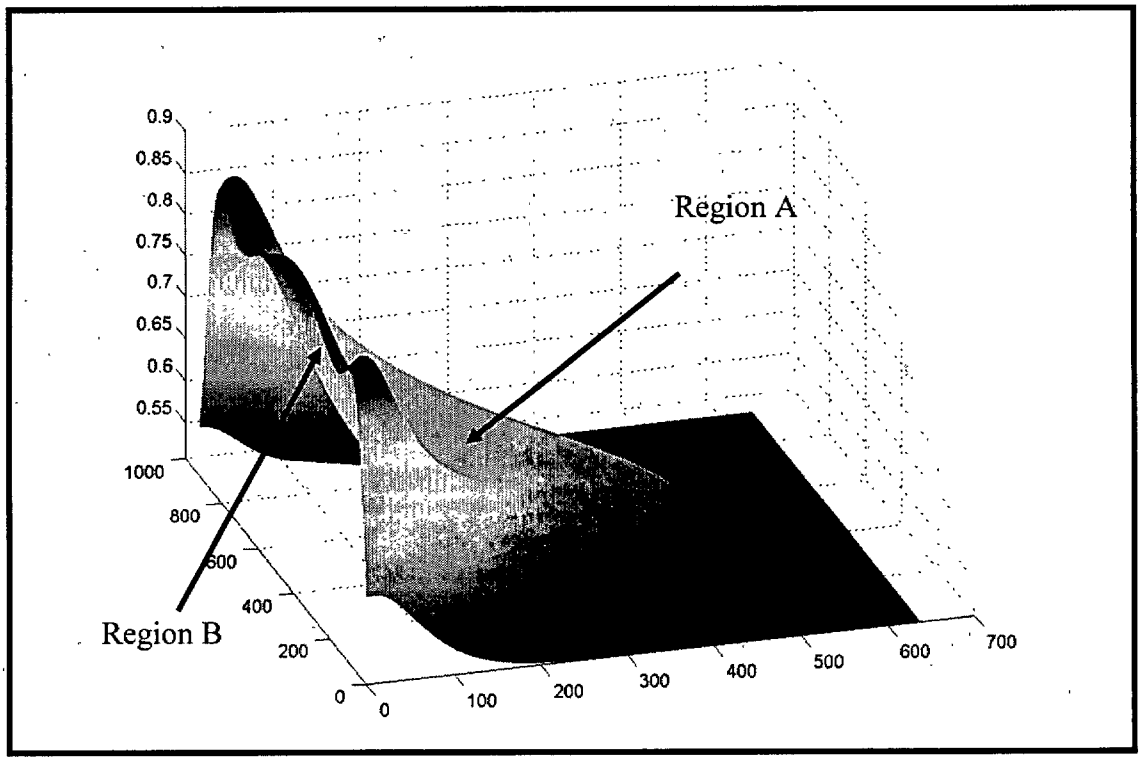


FIGURE 3.26 Final compression mask created by the SV-CNP.

The SV-CNP used the Homotopy Continuation algorithm (3.2) to model the family of curves between the breast edge boundary and the chest wall boundary.

$$y_t = f_{chest} + \left(\frac{t}{N}\right)\left(f_{edge} - f_{chest}\right) \tag{3.2}$$

Essentially, this equation was a model of only a single row, making the analysis of each row independent from each other. This independence combined with the fact that the function was only based on the current family line and the distance to the edge of the

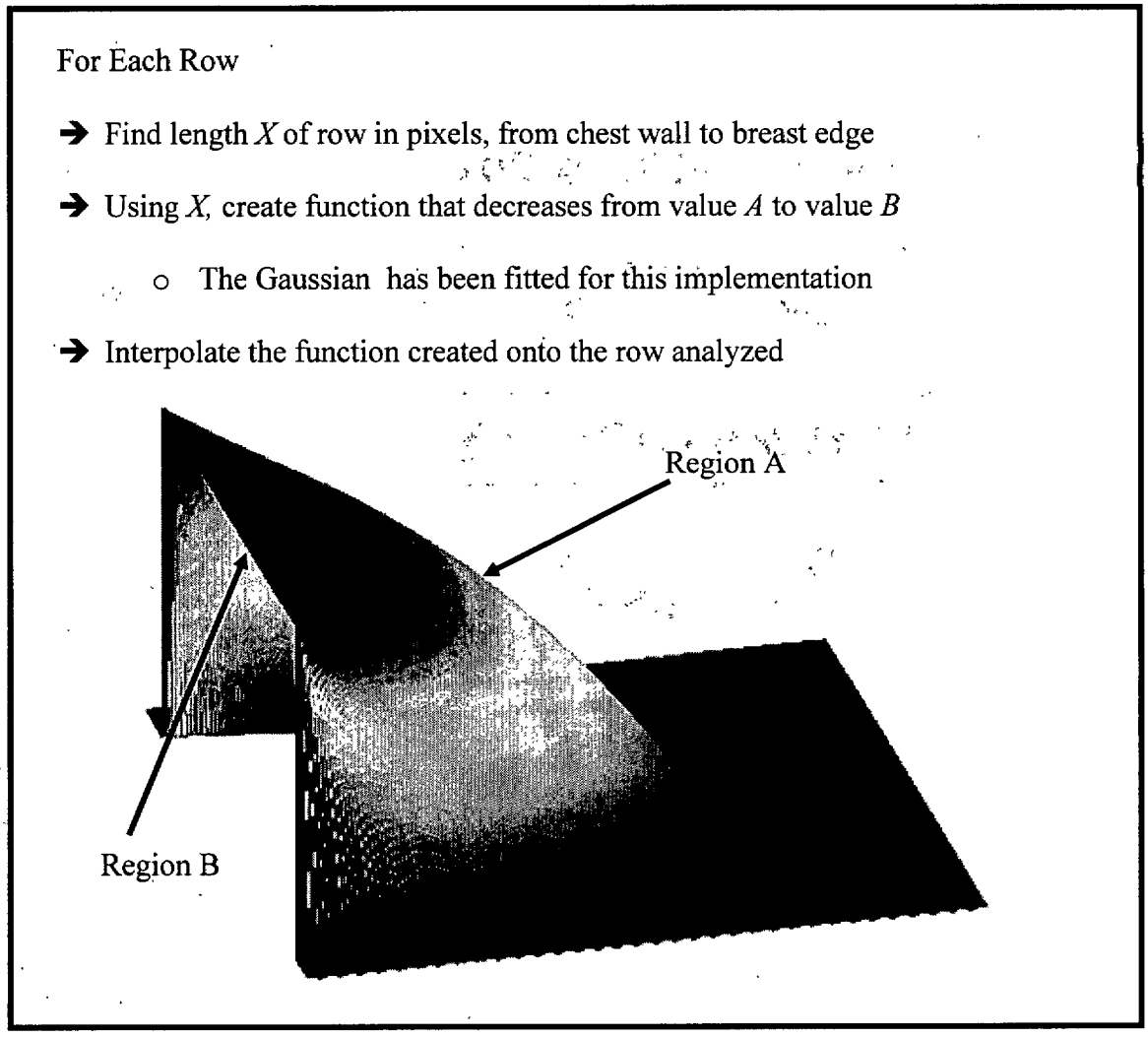


FIGURE 3.27 The new line by line implementation of the compression mask algorithm.

row, allowed for a line by line implementation as shown in Figure 3.28. The new implementation of the algorithm allowed for the filtering process to be eliminated, which in turn removed all the problems associated with it. Figure 3.27, region A is no longer drastically lower than the rest of the tissue and region B is uniform throughout as it should be.

The application of the mask also differed in the implementation developed in this thesis. Rather than applying the compression to the threshold, the actual image was adjusted. Because it was still unknown how much the compression played a role in the analysis of the radiodense breast tissue, the compression was set as low as possible. The images were put through an element by element multiplication with the masks, which the result was then subtracted from the original mask. The results are shown in Chapter 4.

3.3.5 Local Contrast Estimation (LCE)

From the various experiments that were done with regional analysis, it was decided that this thesis need to take a different approach to the problem of radiodensity segmentation. As stated before, the previous algorithms were laden with too many heuristic variables that did not correlate well with the image statistics. Even though the results from the other approaches investigated in this thesis did not fare well, the basic principle of analyzing the two regions, the radiodense and radiolucent, as groups rather than separate pixels was used in the development of the Local Contrast Estimation.

The Local Contrast Estimation worked under one main concept: similar tissues of the breast are interconnected to each other. As stated in Chapter 1, the breast is composed of three different types of tissue; fibrous, adipose, and glandular. The

radiodensity of a mammogram is essentially an analysis of how much glandular tissue is located in the mammogram compared to other types of tissue. Since the main function of the glandular tissue is to create milk during and immediately following pregnancy, the tissue is all interconnected.

Since the interconnectivity is taken into account by the radiologist during the analysis of the mammogram, the approach this thesis has taken was that the best estimate could only be achieved if the methodology was the same. A preliminary test for the interconnectivity was done in Section 3.2.2, but without any good results. The idea was fundamentally sound, but the images found within the two databases were too random to use gray level connectivity. Also, unless the algorithm produced only 2 connected regions, which it never did, using this method also left the problem of classifying the many different regions. Local Contrast Estimation introduced in this thesis solves this problem by analyzing the boundaries created by these regions rather than the regions themselves. Figure 3.28 shows a flow of how LCE works.

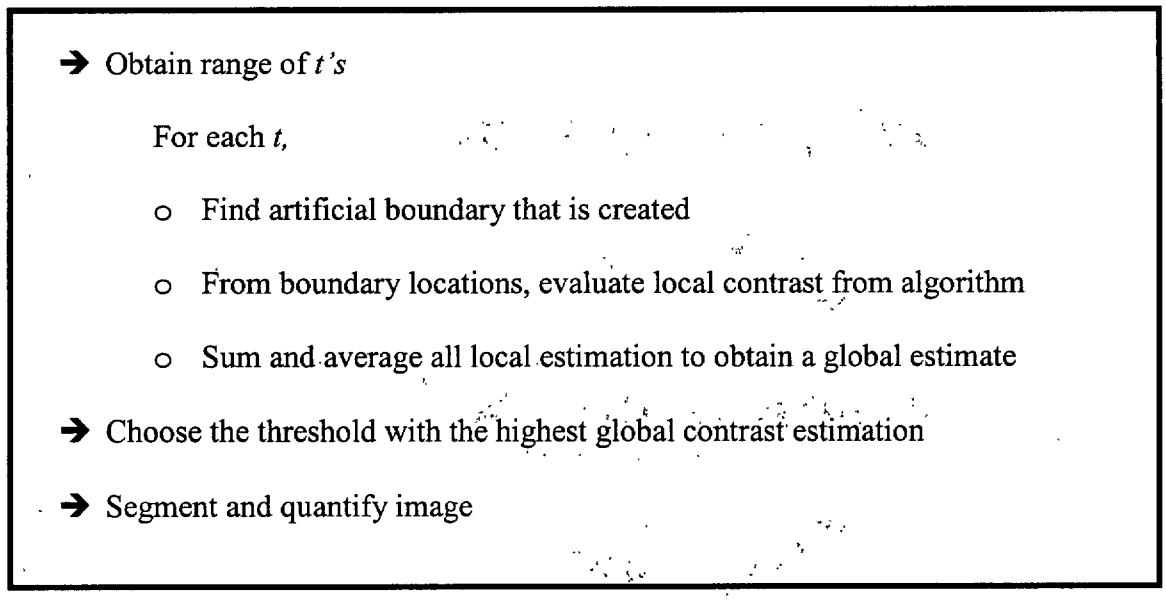


FIGURE 3.28 The flow of the Local Contrast Estimation.

The equations that define this system are shown below:

For every mammogram, X, which is a discrete image of M x N represented by a set gray levels $G = [0,1...,L-1]$, it is segmented by threshold T_{LCE}.

To obtain T_{LCE}, first a range of threshold are obtained using equation 3.2:

for $n=0:N$,

$$\begin{aligned} T(n) &= G_{mean} - a * G_{std} + n * \delta \\ T(N) &= G_{mean} + b * G_{std} \end{aligned} \tag{3.2}$$

where G_{mean} is the mean of all pixel values in the breast regions, G_{std} is the standard deviation of the pixels in the breast region, a and b are experimental percentages, N = number of thresholds, and $\delta = \frac{b * G_{std} + a * G_{std}}{N}$.

For each $T(n)$, the binary segmentation is obtained using equation 3.3:

$$s_k = T(r_k) = \begin{Bmatrix} 1; if\ r_k \geq T(n) \\ 0; if\ r_k < T(n) \end{Bmatrix} \tag{3.3}$$

where r_k = the pixels of the mammogram and $T(n)$ = to the current threshold.

From s_k, a boundary function is obtained by using equation 3.4:

$$\begin{aligned} R_v(x,y) = [(&-s_k(x-1,y-1) - 2s_k(x,y-1) - s_k(x+1,y-1) + \\ &s_k(x-1,y+1) + 2s_k(x,y+1) + s_k(x+1,y+1)]/9; \\ R_h(x,y) = [(&-s_k(x-1,y-1) - 2s_k(x-1,y) - s_k(x-1,y+1) + \\ &s_k(x+1,y-1) + 2s_k(x+1,y) + s_k(x+1,y-1)]/9; \\ R = &\sqrt{R_v^2 + R_h^2} \end{aligned} \tag{3.4}$$

where R_v = the vertical edge component, R_h = the horizontal edge component and R = to the total edge component.

For each $R(x,y) > 0$, a J x K sub image of $X(x,y)$ centered around x and y is analyzed using equation 3.5:

$$v(a) = \gamma^2 X(x,y) \qquad a = 1,2,...,A \tag{3.5}$$

where A is the total number of instances where r(x,y) > 0

The global contrast is defined by equation 3.6:

$$C(n) = v(a)\big|_{average} \tag{3.6}$$

The final predicted threshold is defined by:

$$T_{LCE} = T(n) \quad n \text{ where } \{C(n) = \max(C(n))\}$$

The first step in the algorithm is to estimate the range of T's that will be used in analyzing the local contrast. After the initial separation of the tissue portion of the X-ray with the noise portion, a histogram analysis is done to produce a mean and standard deviation of the distribution. Based on this distribution, the thresholds are determined by equation 3.2. This procedure is shown in Figure 3.29.

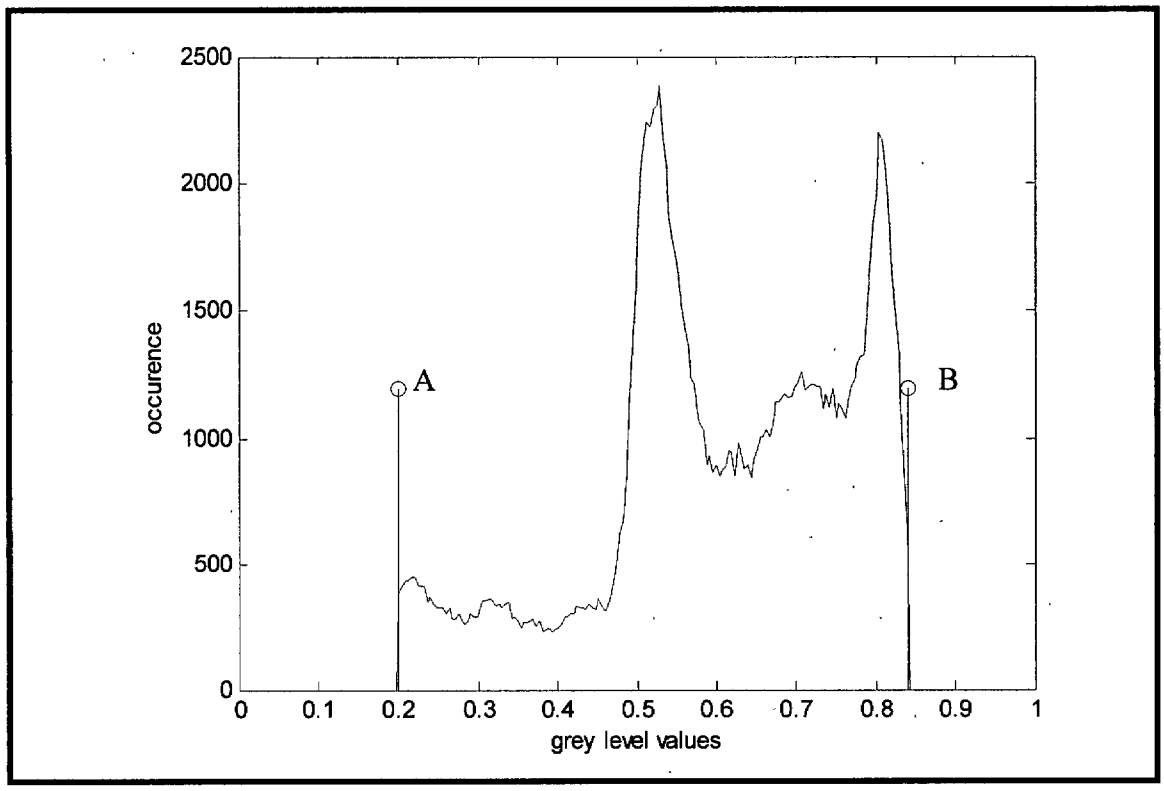


FIGURE 3.29 The range estimation of the threshold values. A is the lowest threshold used. B is the highest threshold used.

The next step in the Local Contrast Estimation is the binary segmentation of a mammogram using a threshold *T*, as shown by equation 3.3. After segmentation, a mask that outlines the artificial boundary created by the segmentation is obtained using equation 3.4. Once the boundary image *B(i,j)* is obtained using these criteria, the locations are sent to the next part of the algorithm for the contrast analysis, as shown in Figure 3.30.

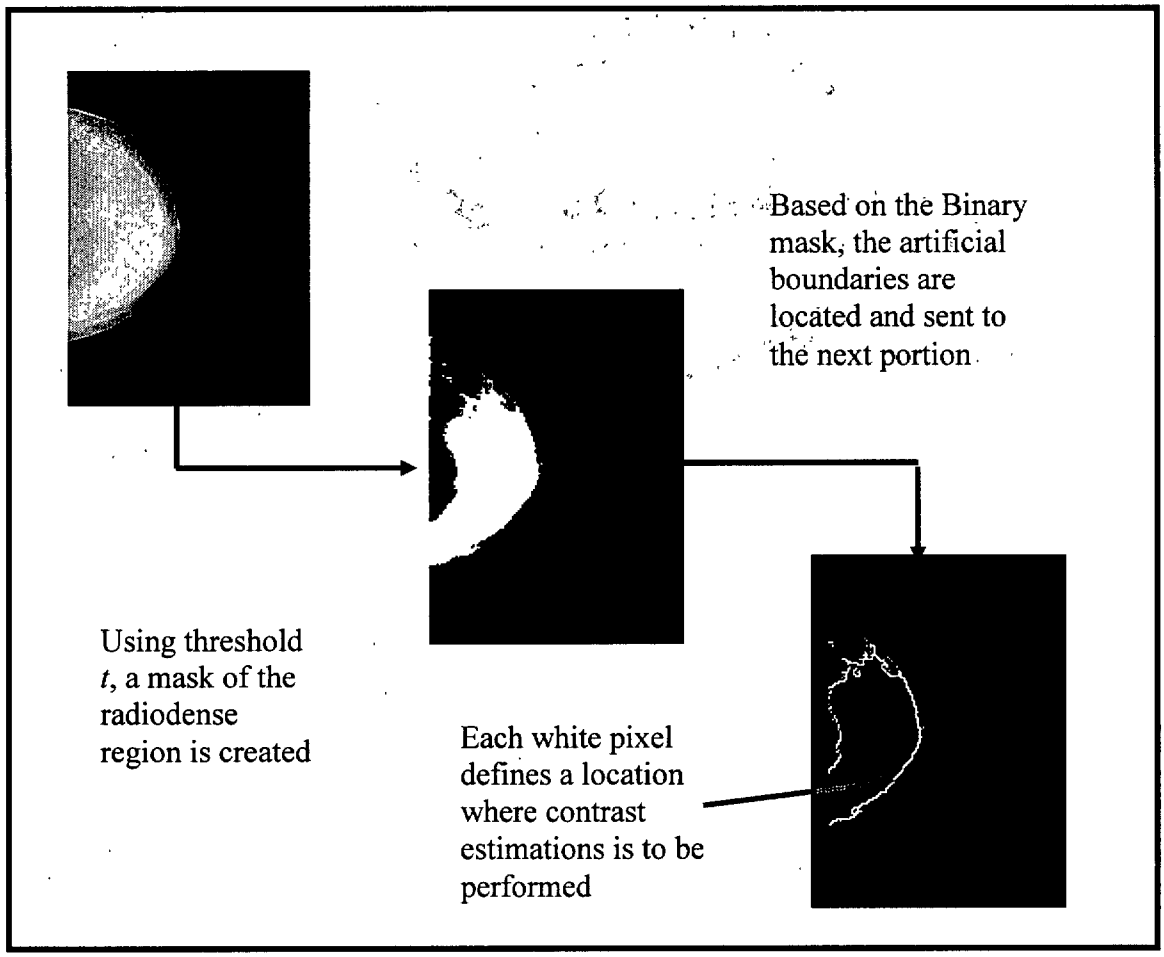


FIGURE 3.30 The artificial boundary estimation *B(i,j)* that was obtained based on a binary segmentation using threshold *t*.

The contrast at each location defined by the previous step is analyzed using four different collections of pixels. Four regions are used so that any type of edge: vertical, horizontal, northwest-southeast diagonal, and northeast-southwest could be analyzed as shown in Figure 3.31. These collections of pixels are then analyzed using the LCE, which the implementation is shown in Figure 3.32(a). Once all the local contrast estimations are estimated, they are all summed and averaged to obtain the global contrast estimate, as shown in Figure 3.32(b).

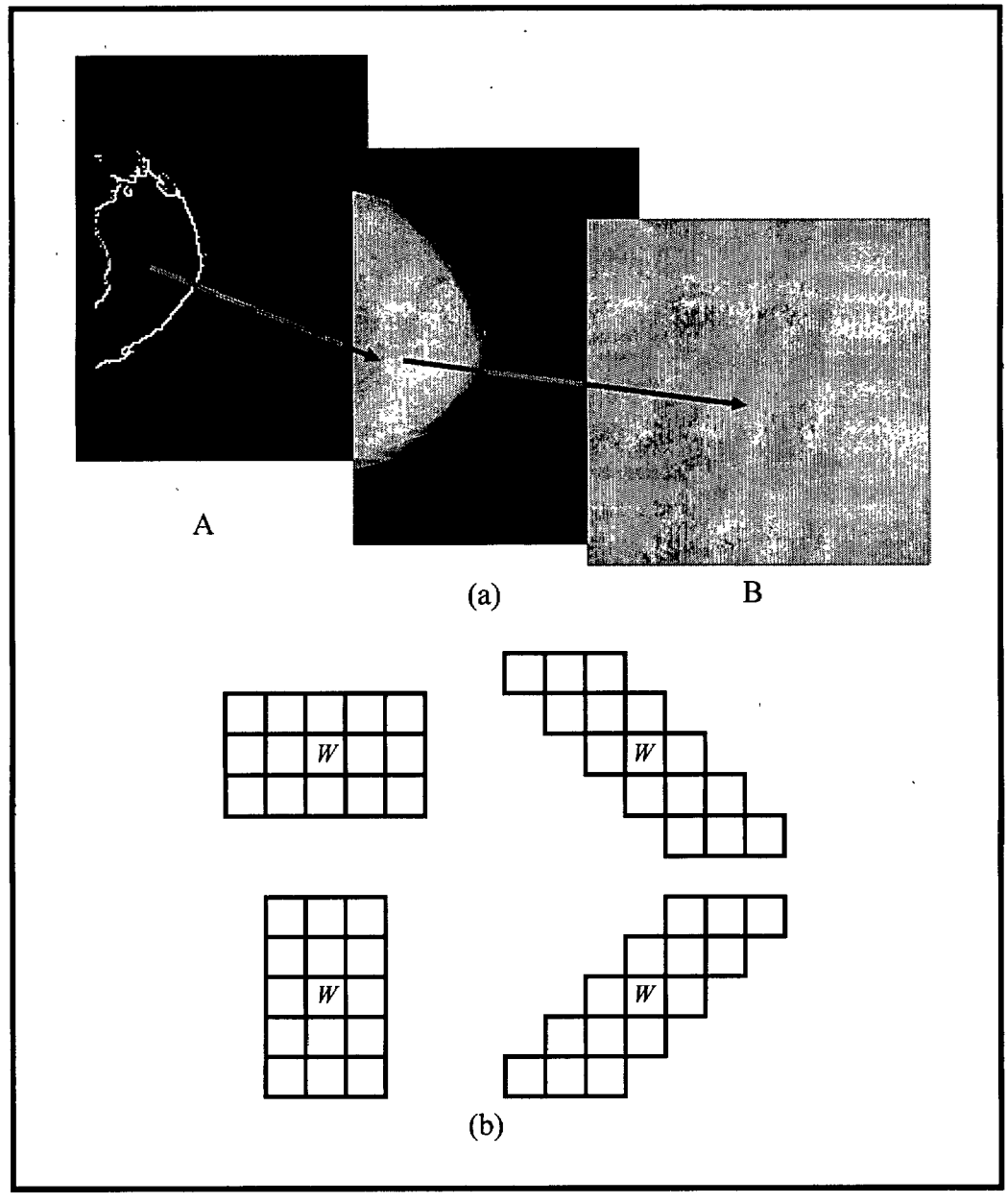


FIGURE 3.31 (a) Each white pixel *W* in boundary image *B(i,j)* shown in A corresponded to a region B that was perceived to be a transition point between radiodense and radiolucent. (b) The different collections of pixels that were obtained to be analyzed by the local contrast estimation.

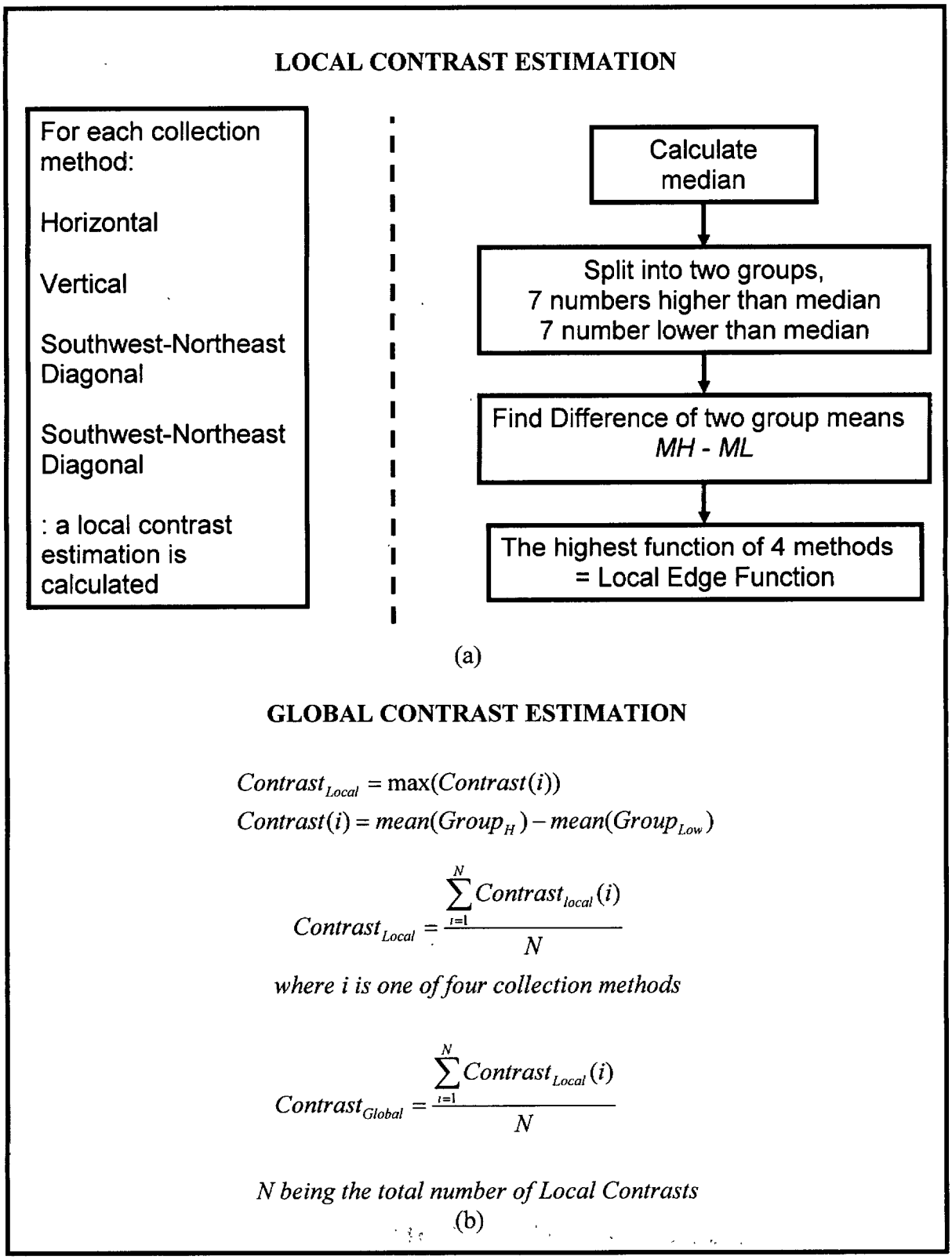


FIGURE 3.32 (a) Procedure for Local Contrast Estimation. (b) The procedure for Global Contrast Estimation.

The main theory behind the LCE is that if threshold T is the correct estimate, the global contrast estimate would be the highest compared to other thresholds. Figure 3.33 shows two regions, one where the image is segmented by the correct threshold (b) and one where the image is segmented by the incorrect threshold (a). If an incorrect threshold is chosen for segmentation, the relative contrast at a perceived edge is not that high, indicating that it is a false boundary. When the correct threshold is chosen, the contrast between the two regions is much higher.

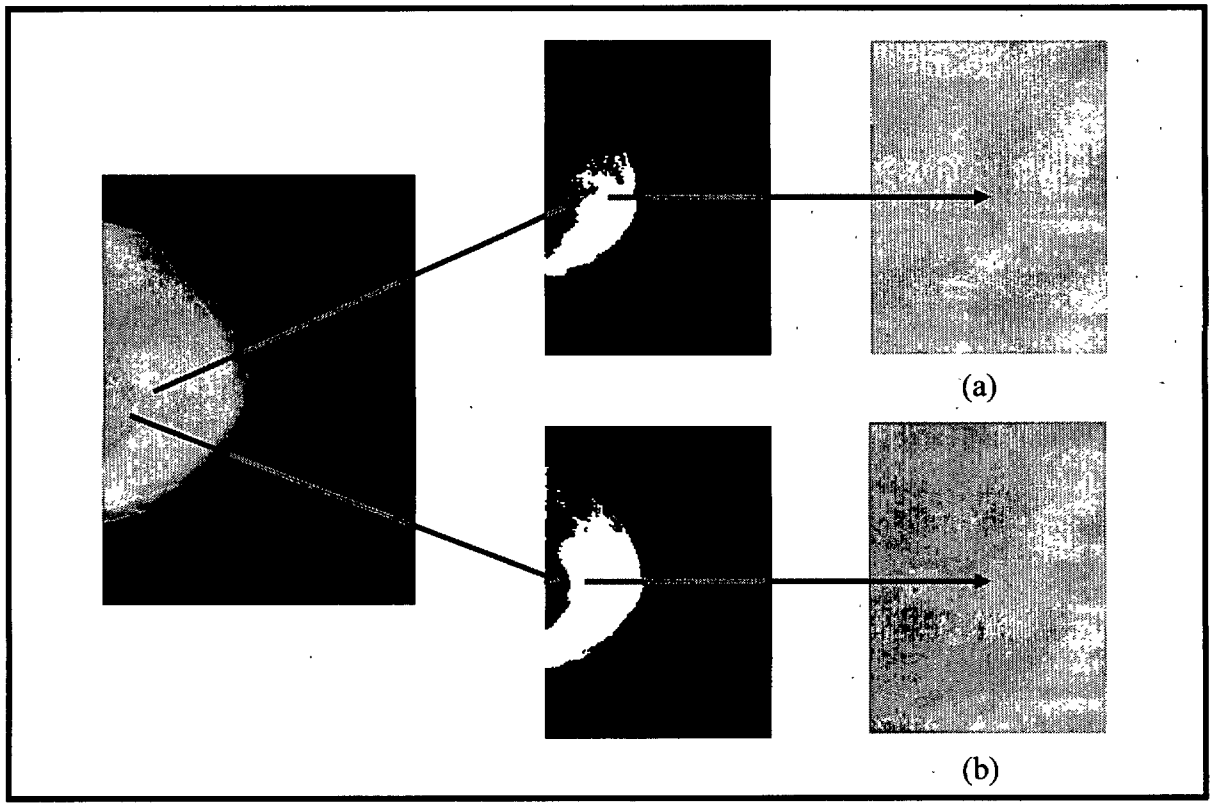


FIGURE 3.33 (a) Region defined as the edge when segmented with incorrect threshold. (b) Region defined as the edge when segmented with correct threshold.

So finally, once the global contrast estimates are obtained for every threshold defined by ranges obtained previously, a function of thresholds to global contrast

estimates is obtained, as shown in Figure 3.34. Based on the function, the threshold with the highest threshold is chosen as the optimal threshold to use for segmentation.

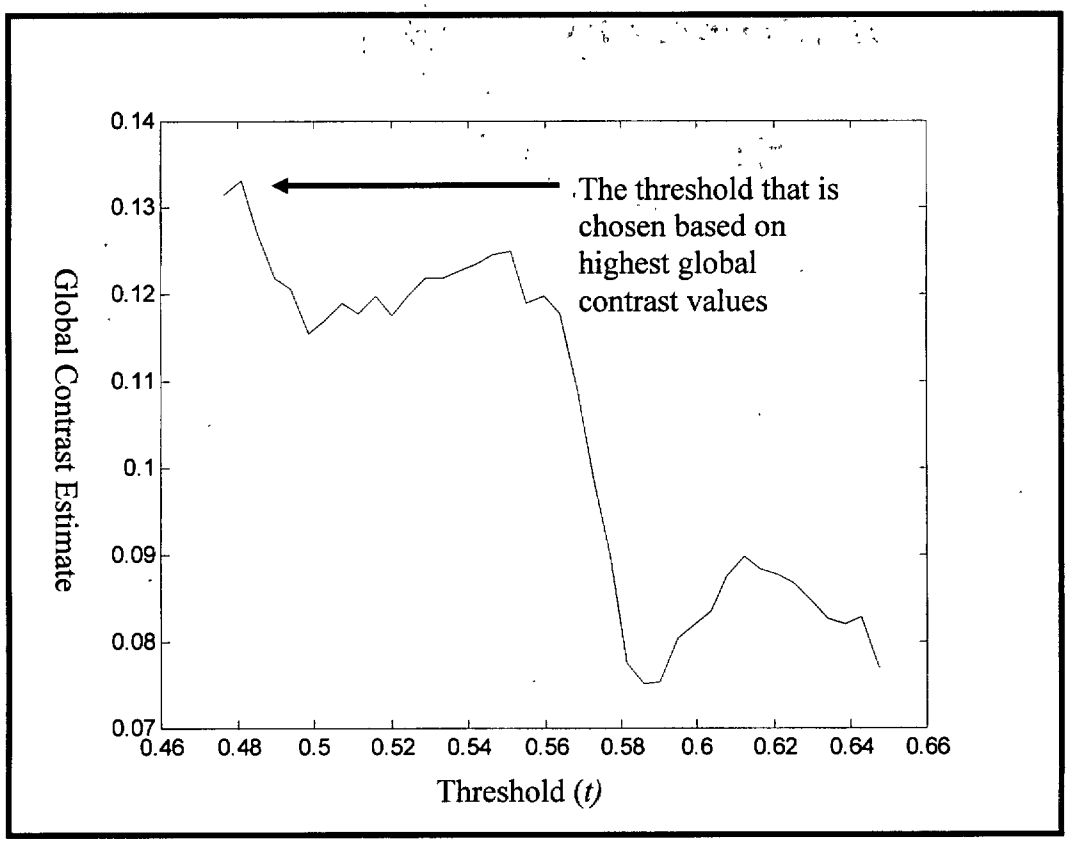


FIGURE 3.34 Once a global estimate is obtained for all thresholds, the highest estimate with its corresponding threshold is chosen as the optimal one to use for segmentation.

3.3.6 Assumptions

For the different portions of the algorithm to function properly, there are several assumptions that are made. For the border removal function:

1. If there is a border in the X-ray, it will have the highest pixel value.
2. The border will always be the outermost portion of the x-ray. For instance, the mass of white pixels will not be followed by a mass of black pixels. In this case,

the white mass of pixels no longer becomes a border but rather just a white noisy stripe.

For masking:

1. Each image has a row length of approximately 1000 pixels. If not, the masking algorithm will change the size accordingly. All parameters were set based on this pixels size.
2. It is assumed that the breast will always be semi-circular in nature on the mammogram. From the test mammograms, all images fit this assumption.

For compression:

1. The compression decay model follows a homotopy pattern.
2. The function for decay is similar to a guassian.

For the Local Contrast Estimation:

1. The two breast regions are best analyzed as layers rather than a point processing approach.
2. The threshold will never be below or above certain values in the overall distribution.

Overall:

1. Any new mammograms introduced into the system will be similar to any noise patterns that were previously observed. This is based on the assumption that any major noise will be noticed by the person who performs the mammograms on the person. (This of course will most likely not be true and will require the addition of other measures to prevent incorrect results due to noise.)
2. The estimates by the trained radiologist are correct and can be reproduced.

Many of these assumptions are based on a limited number of test inputs for validation, and therefore can be better assessed once more training data is obtained.

3.4 Summary

To summarize this approach, there are three major ideas in this chapter:

1. An investigation of texture based methods for segmentation.
2. An investigation of various image processing and enhancement techniques for preprocessing as well as segmentation.
3. Introduction of the Local Contrast Estimation algorithm that provides a more accurate estimate compared to the Toronto method.

Investigating textures and differences associated with both regions help in understanding which type of methods can be used in segmentation. Both imaging processing techniques introduced in Section 3.2 helped in the theoretical implementation of the LCE. The final implementation of the LCE was created to be resistant to noise in addition to being able to span multiple databases without problems. All methods were tested on the Harvard and FCCC databases. The results are shown in Chapter 4.

CHAPTER 4: RESULTS

All algorithms were tested using two sets of data: 10 mammograms from the Channing Laboratory, Brigham and Women's Hospital, Harvard School of Medicine (Harvard) as shown in Figure 4.1 and 34 mammograms from Fox Chase Cancer Center (FCCC) as shown in Figure 4.2. Both of these sets were analyzed by Dr. Celia Byrne using the Toronto method for use as a standard for all algorithms compared in this thesis. The percentages obtained by Dr. Byrne are shown in Table 4.1.

The Harvard and FCCC mammograms are very different from each other and accurately demonstrate the need of an algorithm that is able to span multiple datasets without problems. The Harvard dataset is characterized by an overall high tissue mean compared to the rest of the X-ray, leading to an easier localization of breast edge. The FCCC dataset has the problem of having a very low contrast between all tissue types.

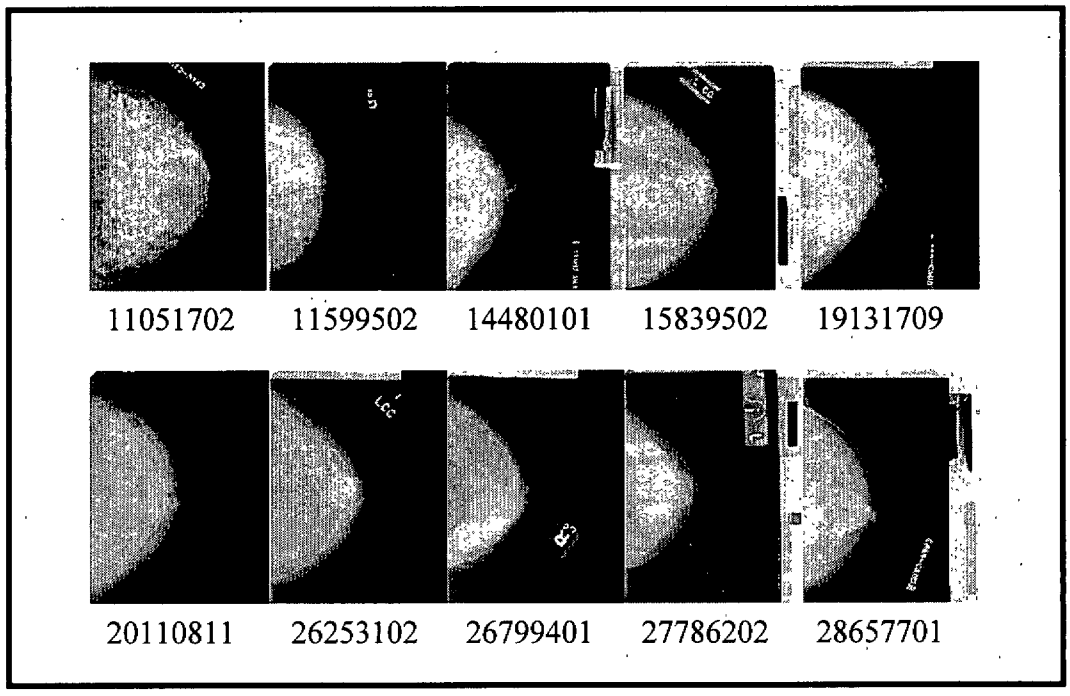


FIGURE 4.1 The 10 Harvard images used in testing the algorithms.

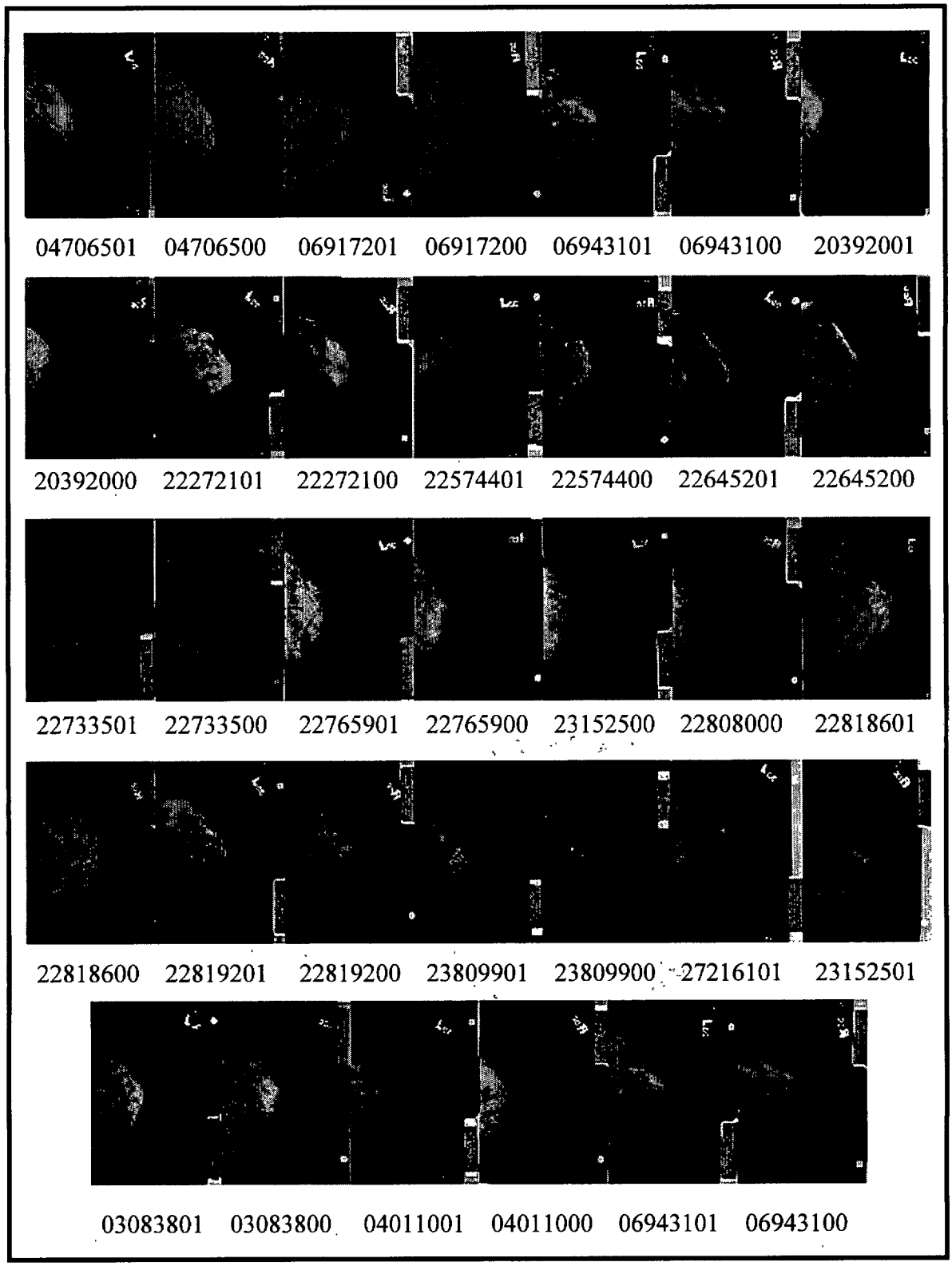


FIGURE 4.2 The 34 FCCC images used in testing the algorithms.

TABLE 4.1 Percentages obtained by Dr. Celia Byrne using the Toronto method.

Image Number	Radiodensity %	Image Number	Radiodensity %
19131709	1.50%	22819201	16.20%
20110811	2.50%	03083800	19.10%
11599502	12.90%	03083801	15.20%
15839502	13.30%	06917200	6.50%
11051702	21.40%	06917201	13.30%
26253102	22.50%	06943100	16.80%
28657701	33.60%	06943101	17.40%
14480101	40.80%	20180800	39.10%
27786202	50.10%	20180801	35.70%
26799401	55.30%	20392000	56.00%
04706500	17.80%	20392001	21.70%
04706501	15.60%	22574400	38.80%
22272100	19.10%	22574401	35.00%
22272101	18.60%	22645200	22.90%
22733500	16.50%	22645201	14.20%
23809900	21.50%	22733501	26.00%
23809901	30.70%	22765900	30.10%
04011000	53.30%	22765901	31.80%
04011001	52.20%	22808000	40.40%
22818600	11.10%	23152500	23.40%
22818601	12.70%	23152501	34.10%
22819200	14.70%	27216101	9.90%
Harvard Dataset			
FCCC FRAP Dataset			

To make a good estimation of general performance, results from both datasets were combined.

For all methods that were evaluated but were not included in the final algorithm, only a subset of the results is shown. This subset includes five images from the Harvard database and five images from the FCCC database, as shown in Figure 4.3. These images

were selected due to their wide ranges in radiodensity percentage as estimated by the radiologist. For all algorithms that were included in the final implementation, the results for all 44 images from the database are shown along with a qualitative analysis of their performance compared with previous methods. For some methods, a demonstration done on canonical images obtained from the SIPI (Signal and Image Processing Institute) database are also shown. The results from the canonical images are meant to show accuracy of code as well as an example of the approach. The images are shown in Figure 4.4.

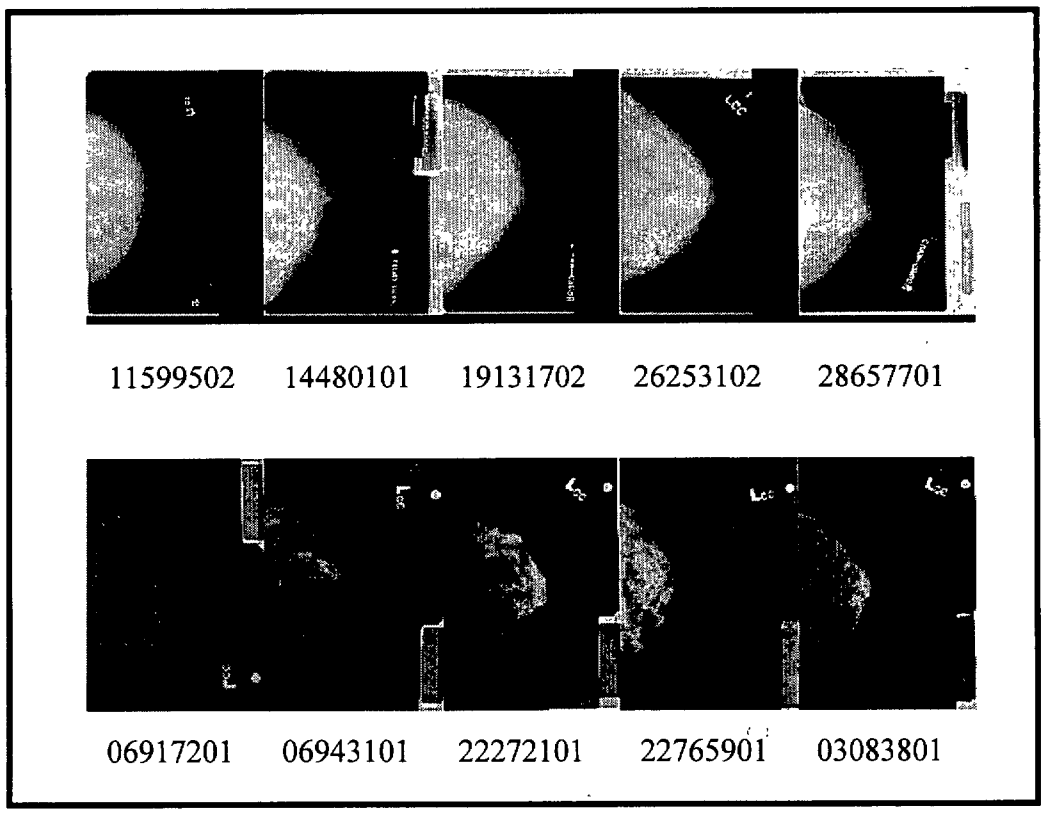


FIGURE 4.3 Subset of images chosen for displaying intermediate methods that are not used in the final implementation of radiodense tissue segmentation.

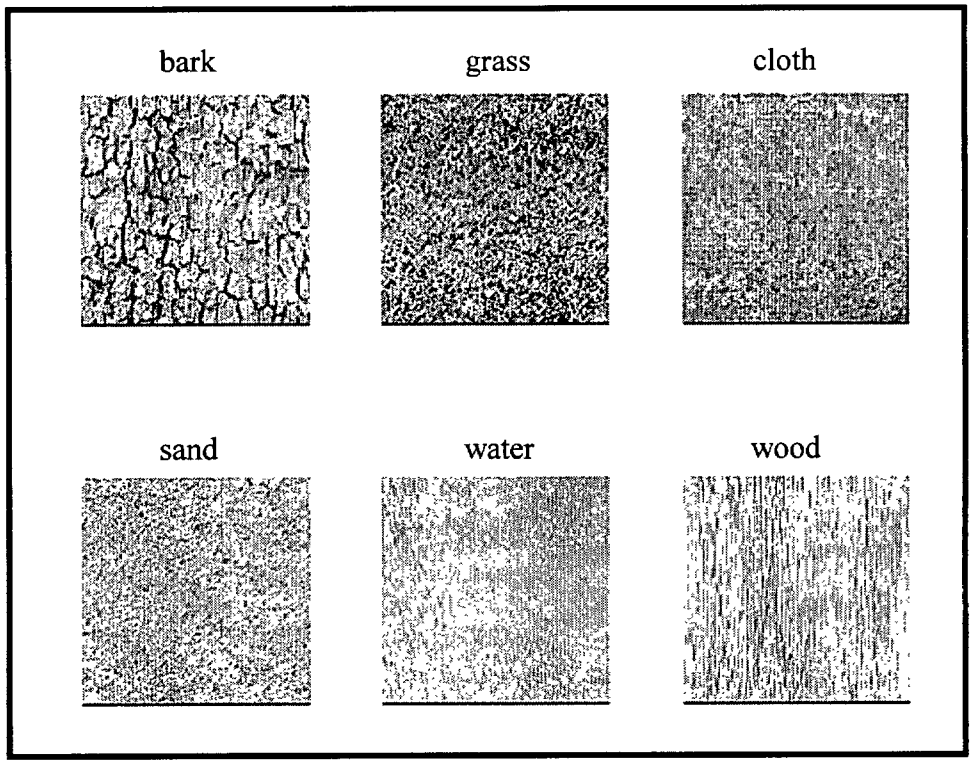


FIGURE 4.4 The canonical images obtained from the SIPI database that will be used throughout Chapter 4.

4.1 Rowan University Digitized Mammogram Database

During the past few years, an extensive database of digitized mammograms has been collected at Rowan University. This database represents an excellent opportunity for studies based on race, location, diet, and age. The database consists of two major groups:

1. Mammograms from Chinese American (CA) patients.
2. Mammograms from Family Risk Analysis Program (FRAP) patients.

All mammograms were scanned using two types of scanners. The Agfa scanner was used originally and produced joint photographic experts group (jpeg) images. Because the Agfa only has an optical density of 3.0 and is not considered a medical grade scanner, the images from the Agfa are characterized by low contrast and high amounts of noise as

shown in Figure 4.5(a). Because the images are in jpeg format, there are only 256 (8-bit) gray levels in each image. The Lumisys is currently the scanner being used for all the digitization of all mammograms. This scanner has an optical density of 3.85 and is considered to be a medical grade scanner. The Lumisys scans all mammograms into 4096 (12-bit) gray level DICOM (Digital Imaging and Communications in Medicine) format and does a very good job of not introducing noise into the scanned images, as shown in Figure 4.5(b). The Lumisys also produces higher contrast images making any analysis easier.

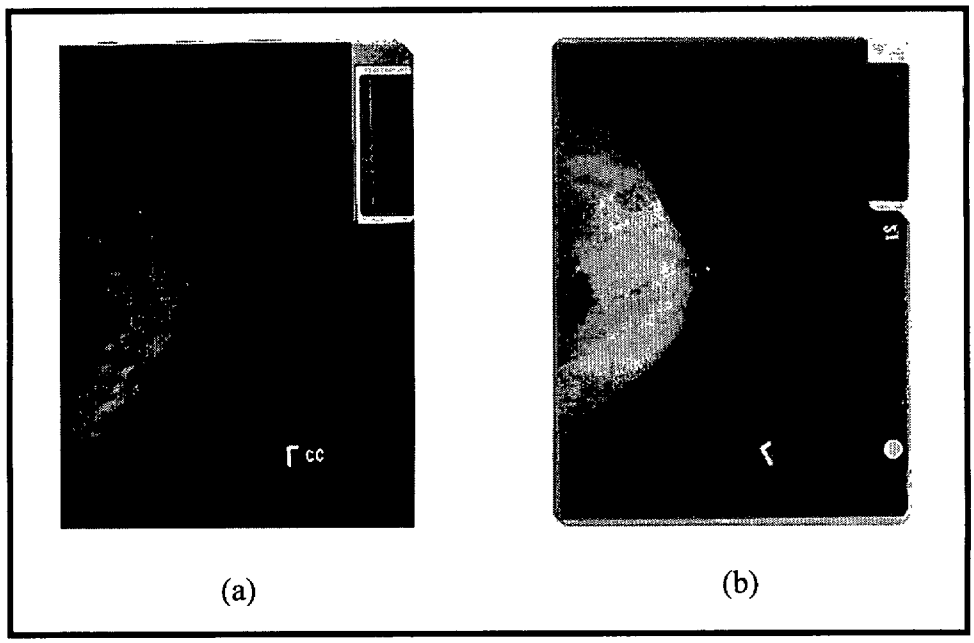


FIGURE 4.5 (a) Mammogram from CA database scanned using the Agfa. (b) Mammogram from CA database scanned using Lumisys. They are not from the same patient.

The Chinese American database only contains RCC and LCC images. It contains 55 pairs (LCC and RCC) of mammograms that have been scanned using the Agfa images. There are 134 pairs (LCC and RCC) of mammograms that have been scanned

using the Lumisys. In total, there are 189 pairs of mammograms, which equates to 378 total mammograms in the Chinese American database. The FRAP database on the other hand contains many image groups that contain RMLO and LMLO images as well as the RCC and LCC. The FRAP database contains 339 images that have all been scanned by the Agfa in jpeg format. In total, there are 717 mammograms in the database. Table 4.2 shows the number of mammograms in Rowan database.

TABLE 4.2 The Rowan University digitized mammogram database.

	# of Agfa images	# of Lumisys images	Total
Chinese American	110	268	378
FRAP	339	0	339
Total	449	268	

4.2 Regional Variance Statistics

As was stated in Chapter 3, the main focus of the variance imaging was to obtain a preliminary indication of how well texture based methods would work on the breast tissue. The results of variance imaging are shown in Figure 4.6. For the Harvard images, variance imaging creates almost no separation of gray levels in comparing the radiodense and radiolucent regions. The imaging method suppresses the tissue region and only enhances the regions near the edges. This was to be expected due to high contrast located within the edges. The FCCC images show some difference between the radiodense and radiolucent regions. This indicates that even though textures for segmentation may be

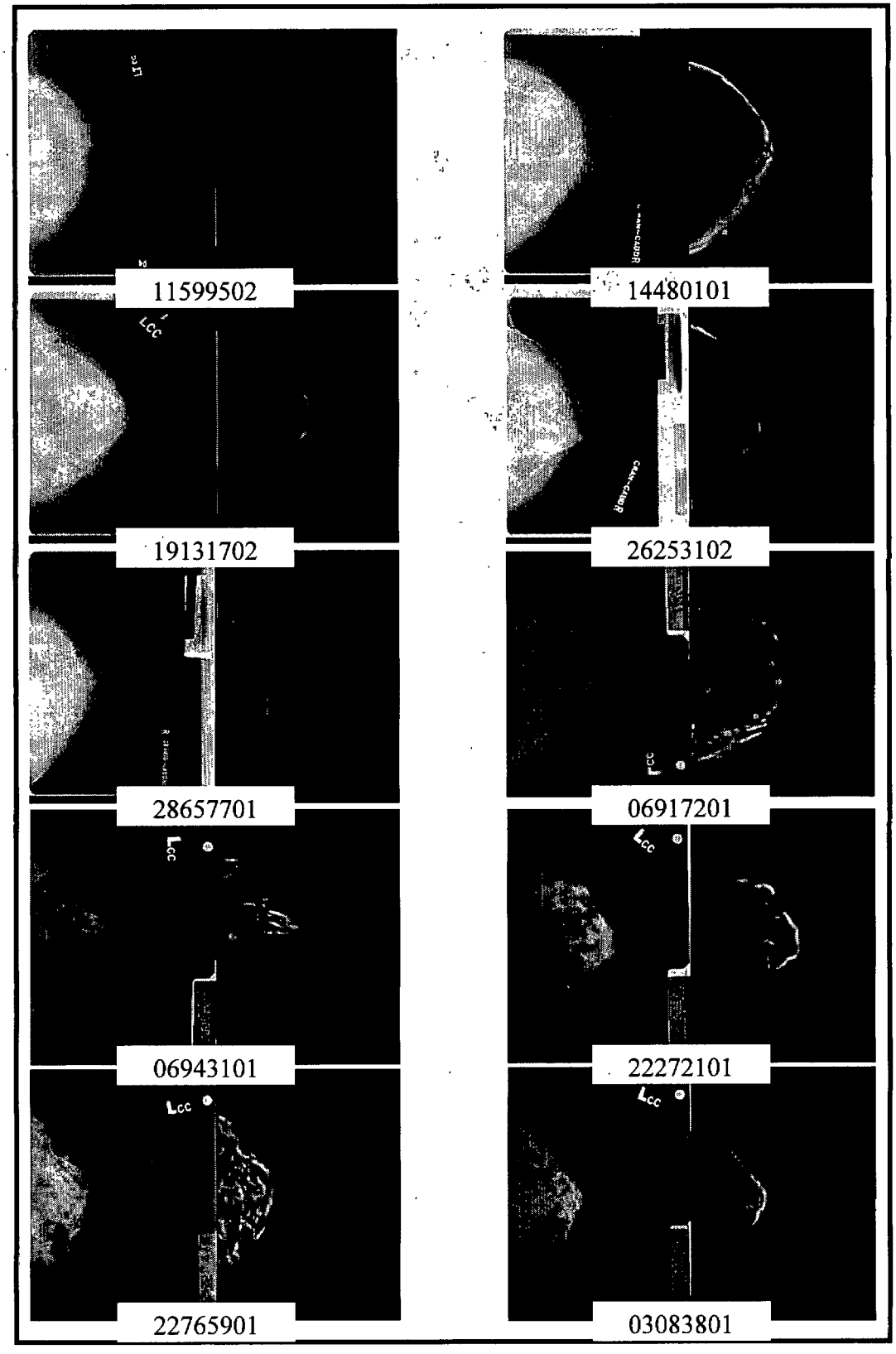


FIGURE 4.6 Variance imaging results.

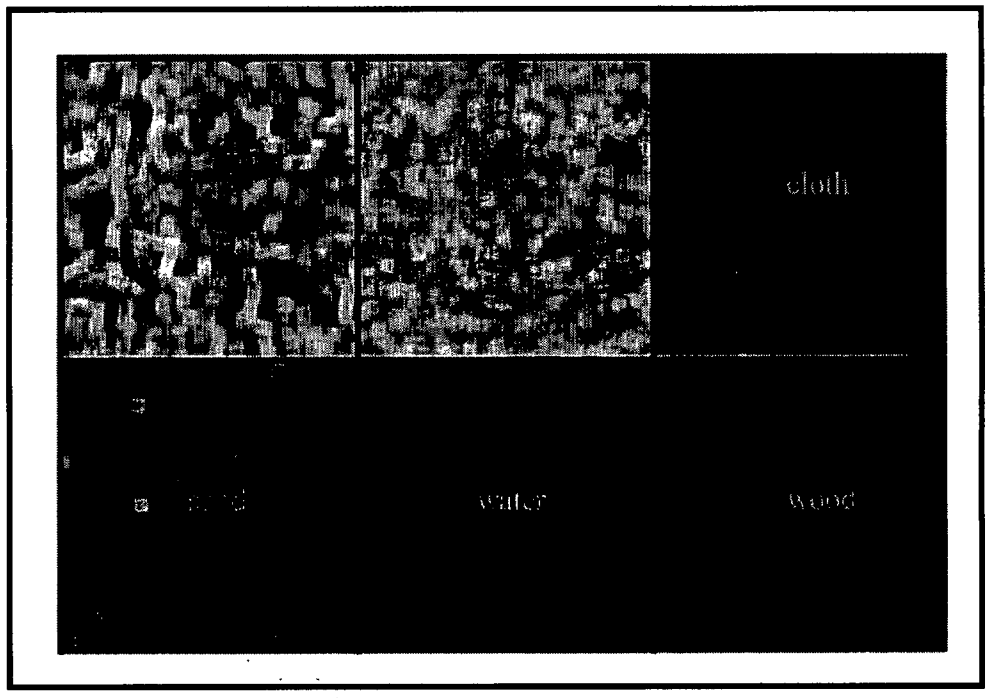


FIGURE 4.7 Variance imaging results on canonical images.

viable, the two different databases may behave differently depending on what procedure is used. When the technique is used on canonical images, it performs relatively well, as shown in Figure 4.7. Even though there are many variations within each type of texture, the overall mean between textures are distinct.

For overall variance comparisons, the statistics correlate well with what was seen in the images. The radiodense and radiolucent variances throughout the ten images are vastly different from each other, as shown in Figure 4.8 and Figure 4.9. The average variance value for the radiodense regions was 0.021, with a standard deviation of 0.018. The radiolucent regions also had similar results where the average variance of 0.0029 had a standard deviation that was 0.0026. However, the results do show that there is a relative difference between the radiodense and radiolucent regions. For all statistics, the

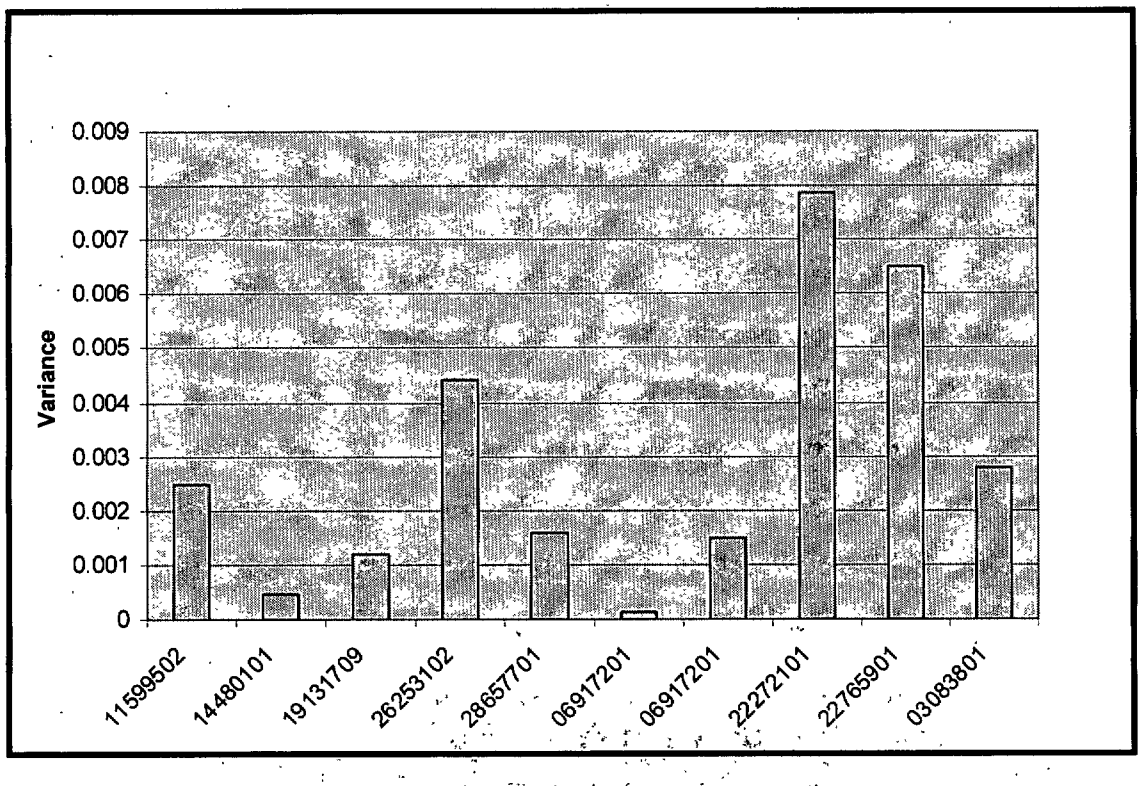


FIGURE 4.8 Variances of radiodense regions throughout the 10 images.

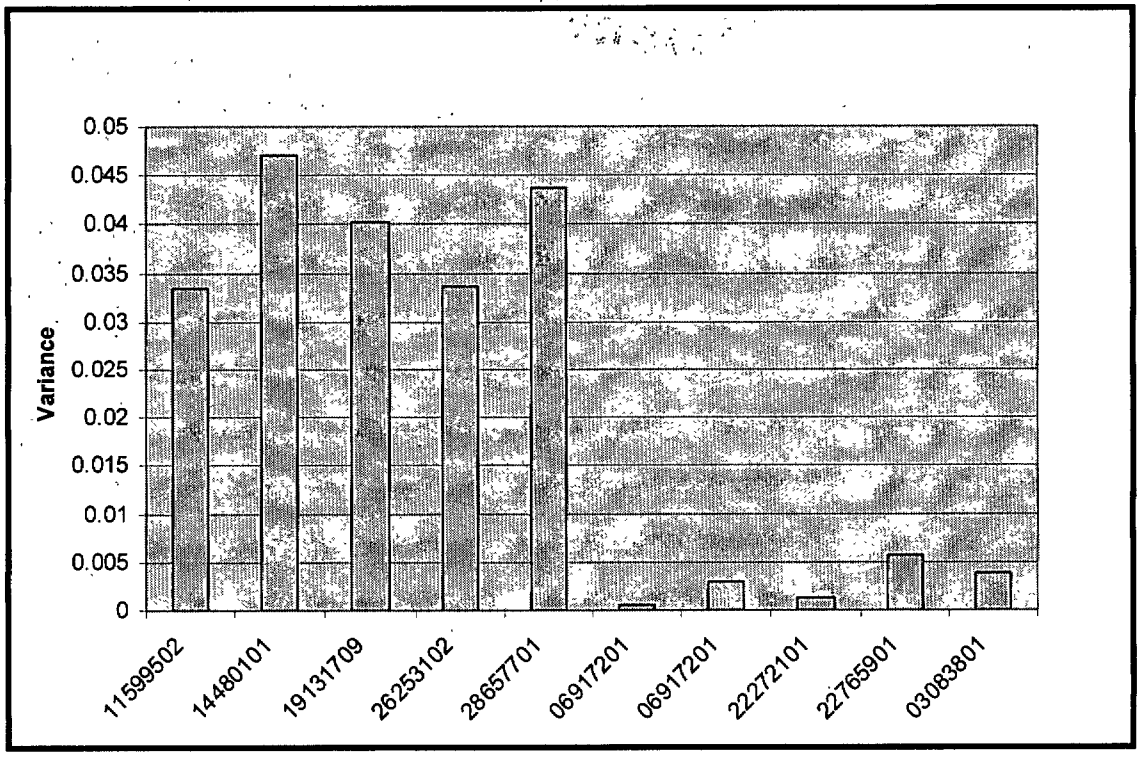


FIGURE 4.9 Variances of radiolucent regions throughout the 10 images.

relative difference was 0.0183. Within the same image, this difference was more drastic; therefore allowing for an initial assumption based on the results. The relative difference between the two regions, radiodense and radiolucent, indicate that a texture segmentation may be possible, but the vast changes between images also indicates that the procedure must be adaptive.

4.3 Gabor Filtering

Figure 4.10 shows the collection of all known 20 x 20 radiodense samples, based on Dr. Celia Byrne's estimate, that were used to do a detailed frequency analysis for Gabor filtering. The 20 x 20 radiolucent samples are shown in Figure 4.11. These samples were selected from the subset of images shown in Figure 4.4. For the analysis of the frequency domains, the mean of each sample was normalized before the two dimensional Fast Fourier Transform was evaluated. Figure 4.12 shows the corresponding Fourier Transform shown in log view for the radiodense tissue. Figure 4.13 shows the corresponding Fourier Transform for the radiolucent tissue.

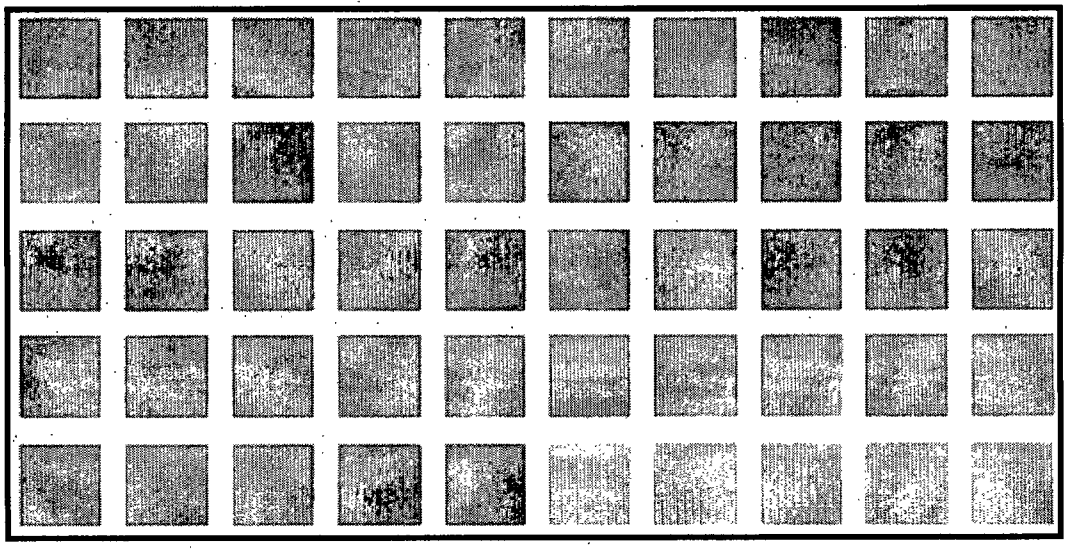


FIGURE 4.10 Known radiodense samples from subset of images shown in Figure 4.3.

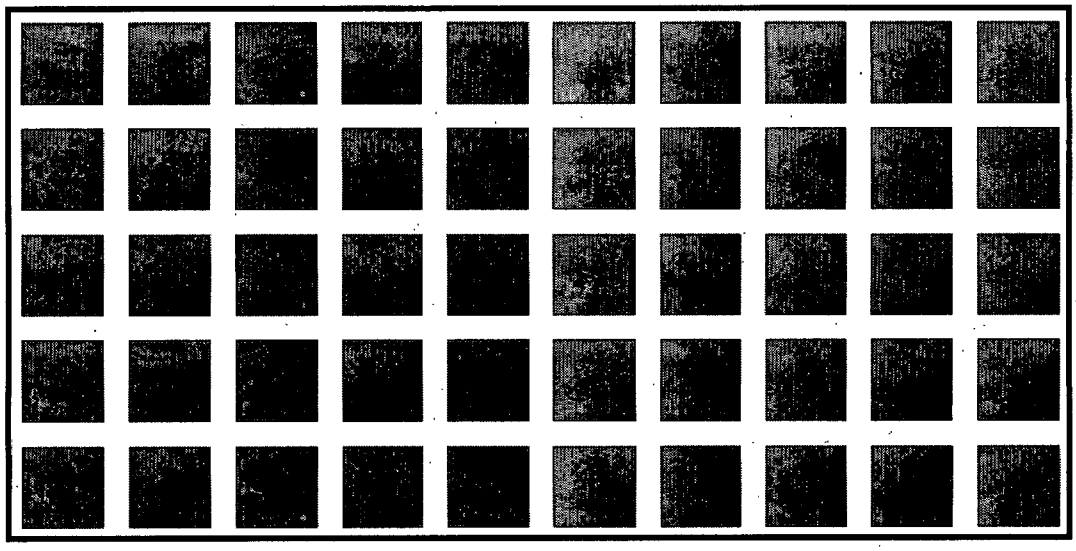


FIGURE 4.11 Known radiolucent samples from subset of images shown in Figure 4.3.

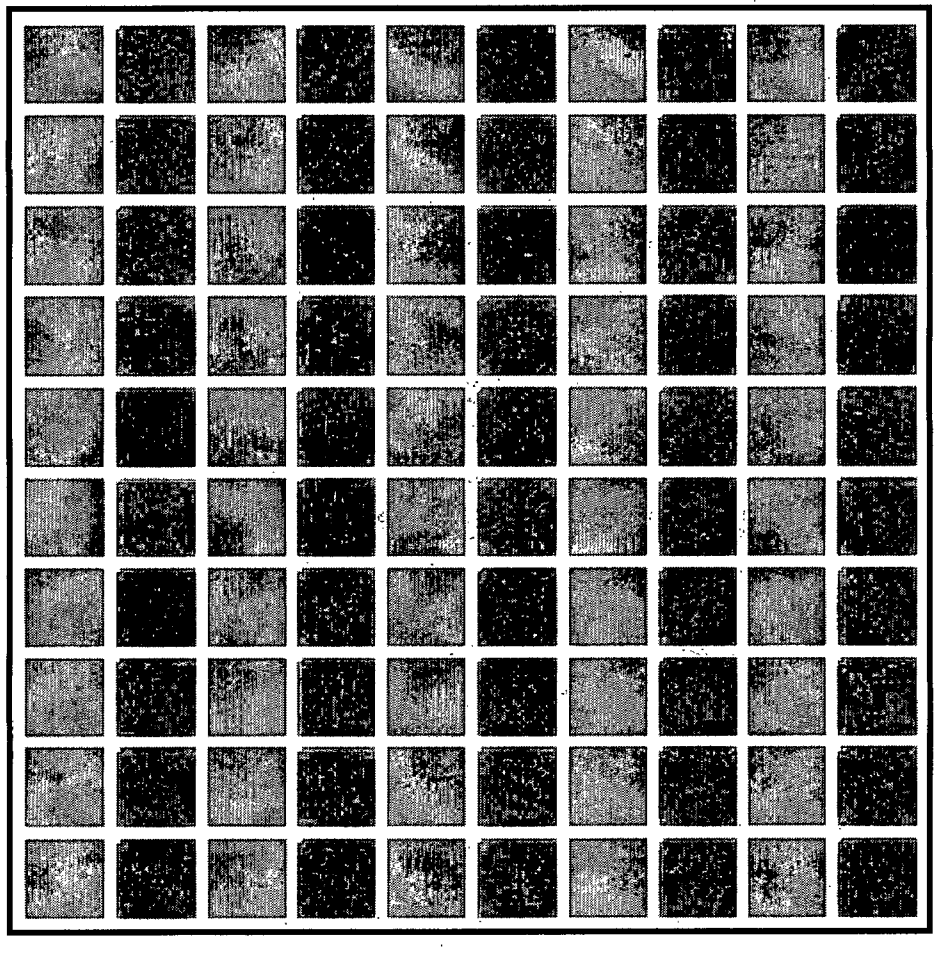


FIGURE 4.12 The mean normalized radiodense samples with their corresponding Fourier Transforms.

FIGURE 4.13 The mean normalized radiodense samples with their corresponding Fourier Transforms.

The theory behind the frequency analysis was to find a range of frequencies that were more prevalent in radiodense tissue than in radiolucent tissue. To accomplish this task, all frequency profiles were averaged into two matrices, one representing the radiodense tissue and another representing the radiolucent tissue. The ratios of these two images were then obtained, as shown in Figure 4.14. It was expected that regions of bright patches were going to appear upon analysis, which would have indicated the vast frequency differences in the two tissue types. Unfortunately, the information only shows very small regions where the ratio between radiodense and radiolucent regions are high

enough for Gabor passband frequency selection. The frequency analysis was also restricted to the lower spatial frequency domains because overall energy of the high frequency regions was insignificant. As a result, there were only a few sets of frequencies that could be chosen for Gabor filtering testing. These frequencies correspond to the blocks location in Figure 4.14 of: (3,17), (5,3), (19,5), and (9,2). Therefore, the results shown in Figure 4.15, which are a collection of results based on using Gabor window sizes of variance, σ_x = .5, .75, 1, 1.5 in the 4.1, just became a form of low pass blurring. The Gabor variances were kept symmetric, therefore $\sigma_x = \sigma_y$.

$$H(u,v) = \exp\{-2\pi^2[\sigma_x^2(u-u_0)+\sigma_y^2 v^2\} + \exp\{-2\pi^2[\sigma_x^2(u+u_0)+\sigma_y^2 v^2\} \qquad (4.1)$$

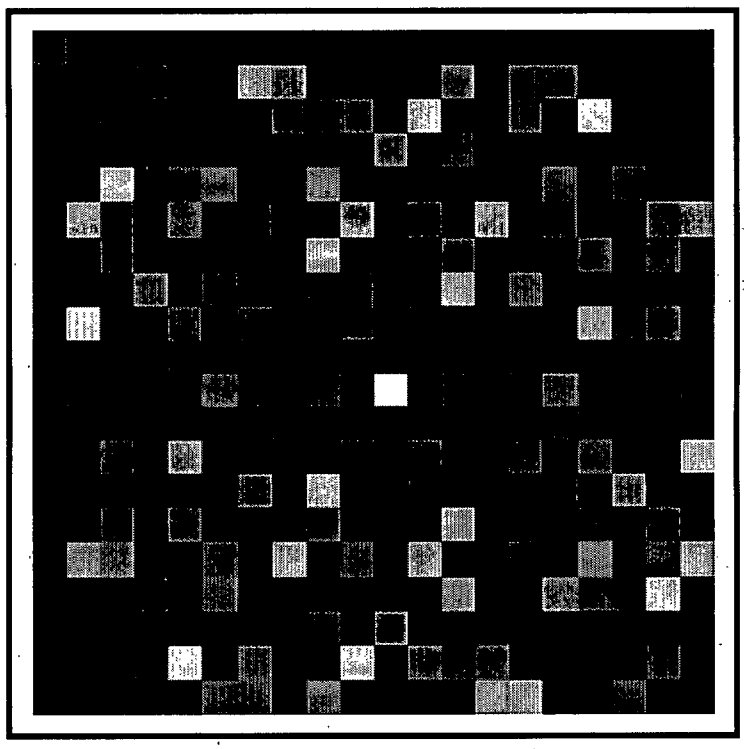


FIGURE 4.14 The ratio of the average radiodense and radiolucent frequency regions.

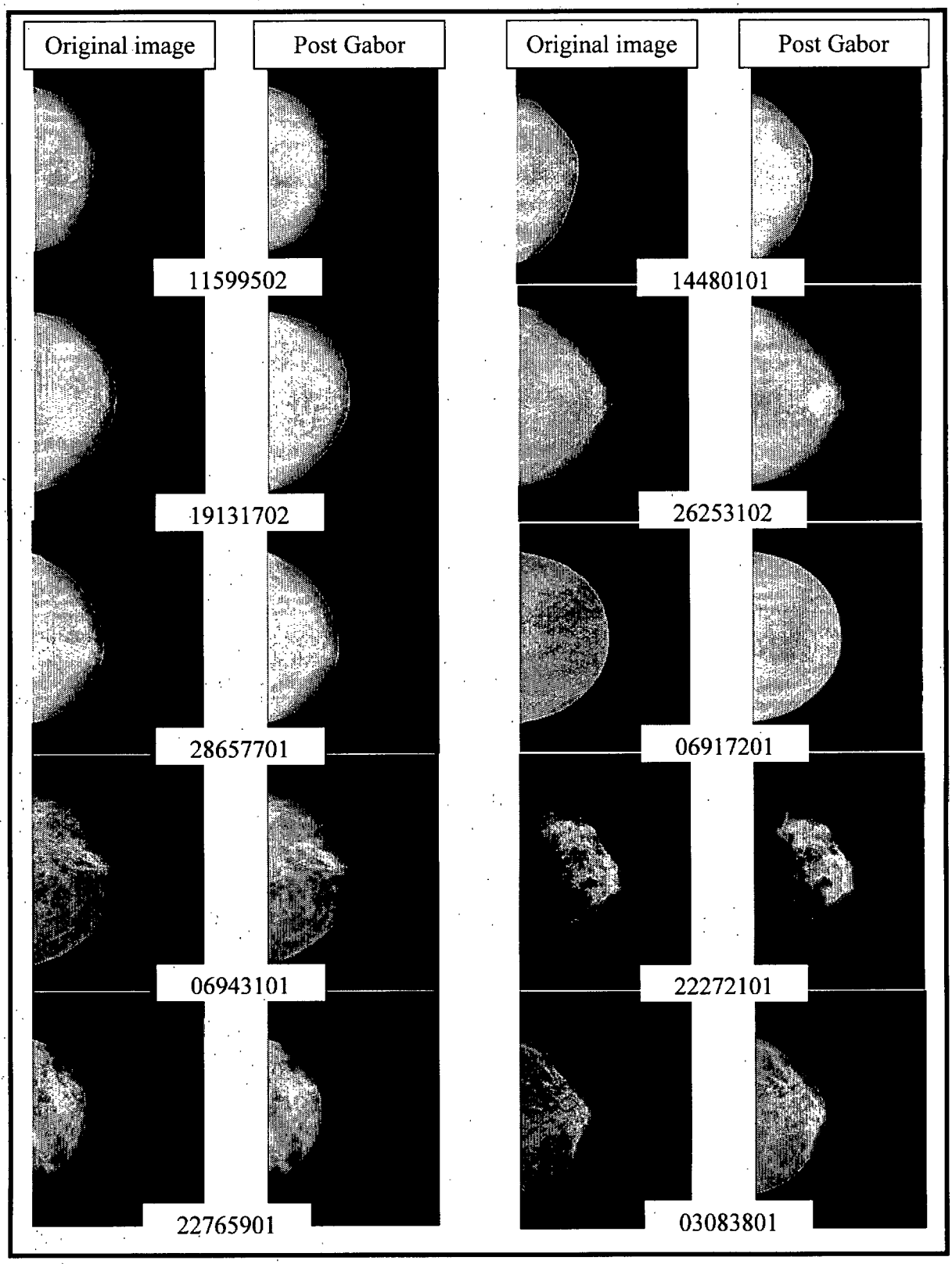


FIGURE 4.15 A collection of results for the Gabor Filter based on σ_x values of .5, .75, 1, 1.5 as well as passband frequency locations of (3,17), (5,3), (19,5), and (9,2).

4.4 Co-occurrence Matrix

Figure 4.16 represents typical results obtained by using co-occurrence matrices for the Harvard images. These results were generated from using various co-occurrence comparisons such as (0,1) and (0,2). Other co-occurrence values were used and

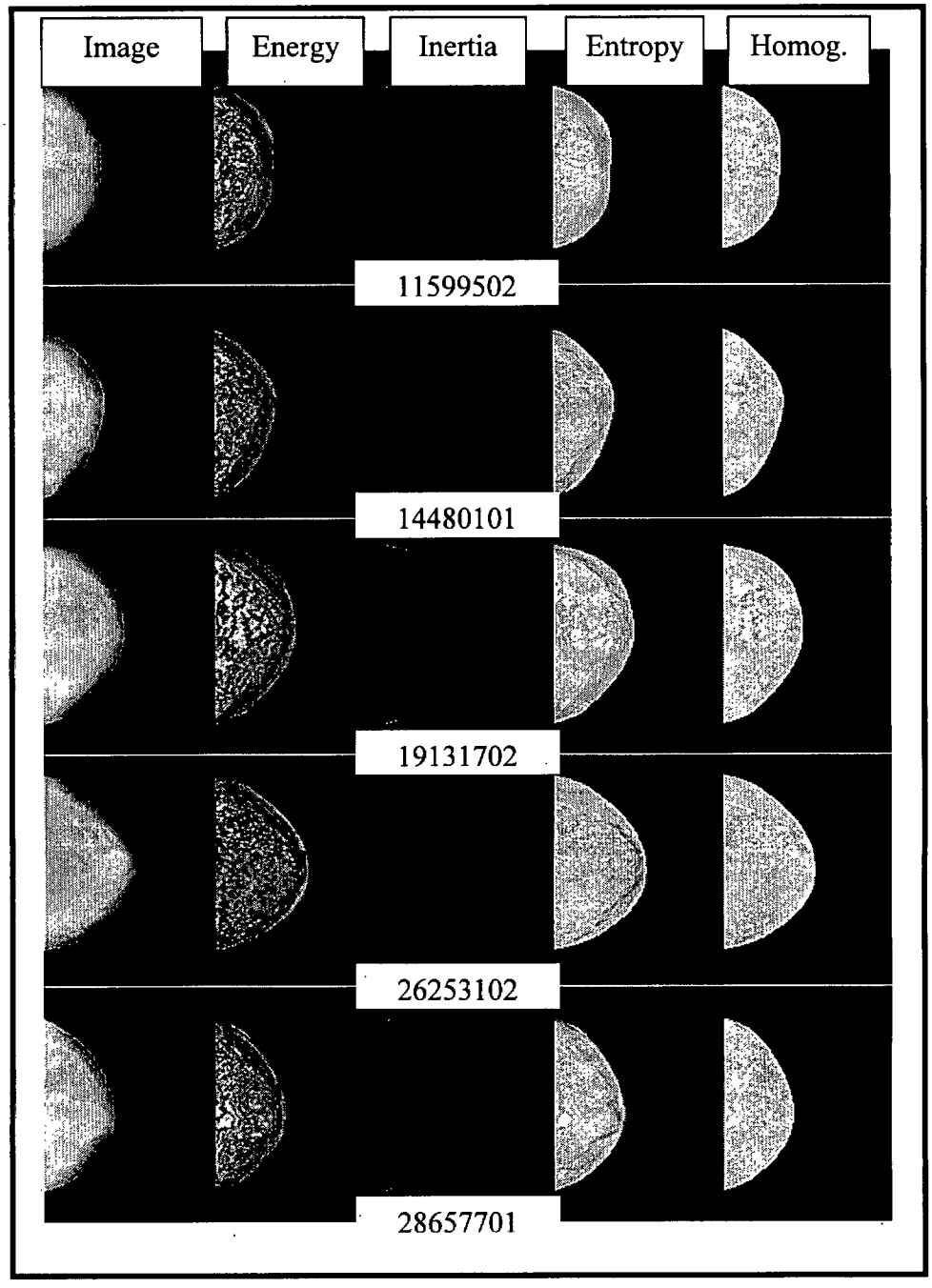


FIGURE 4.16 Results of co-occurrence matrices for some Harvard images.

produced similar results as the ones shown in Figure 4.16. A mask size of 9 x 9 was used because it was shown have the best tradeoff between resolution and consistency. The results for the energy, entropy, moments, and inverse moments, respectively, do not even remotely correspond to the actual regions. This implies that for the Harvard images, the texture qualities for both regions are very similar. As a result, when using a unsupervised clustering method such as K-means, the segmentation result does not classify the correct parts of the breast tissue, as shown in Figure 4.17.

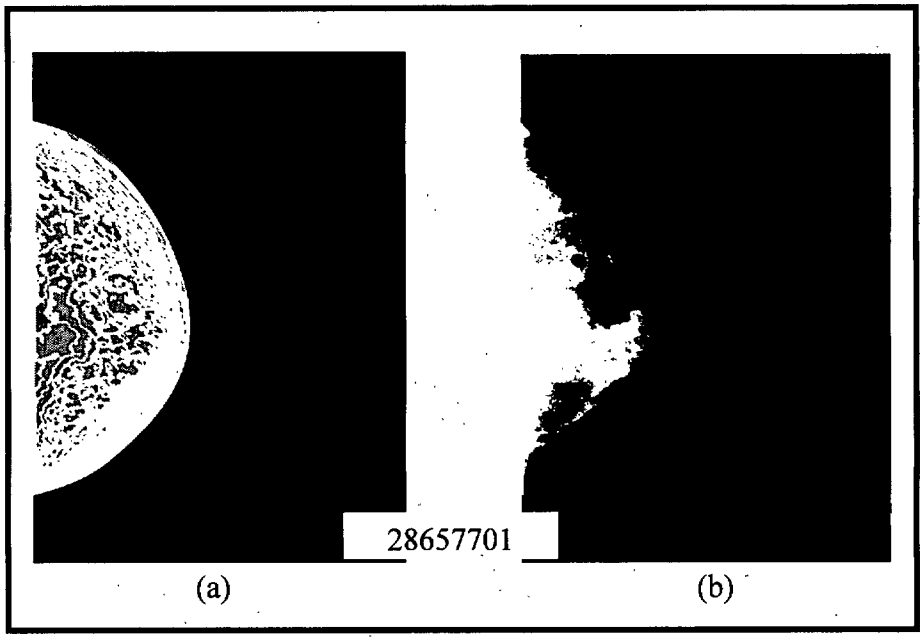


FIGURE 4.17 (a) Result of a K-mean clustering of co-occurrence features for a Harvard image. (b) Desired segmentation results based on Toronto method.

When the same technique is used for the FCCC images, the results obtained appear to be more promising. The energy, entropy, and homogeneity statistics provide some separation between the radiodense and radiolucent tissue, as shown in Figure 4.18. As a result, the unsupervised clustering image shown in Figure 4.19 is better correlated

with the actual profile created from the Toronto technique. However, the results are still not concise enough to be used for biomedical applications.

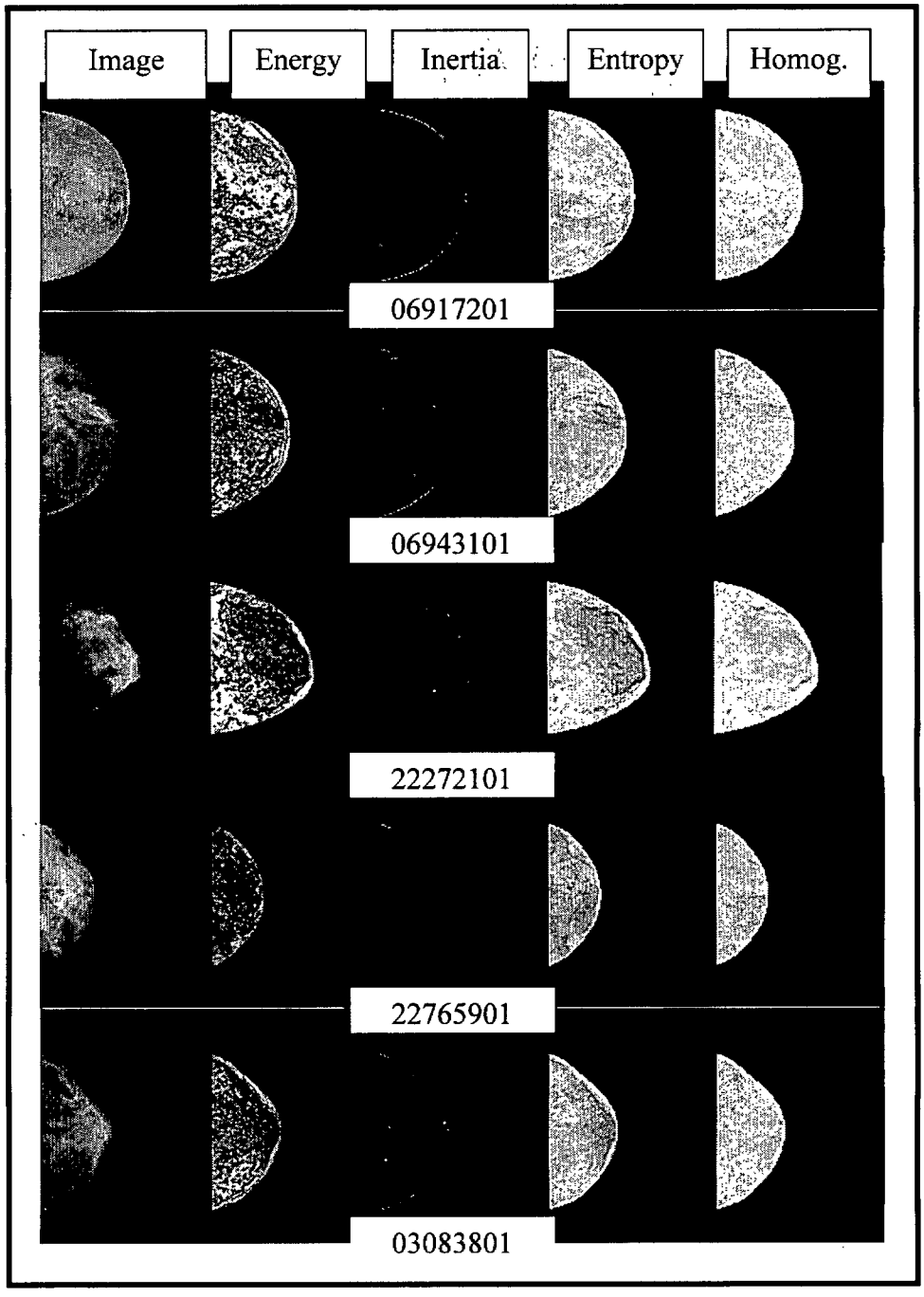


FIGURE 4.18 Results of co-occurrence matrices using the FCCC images.

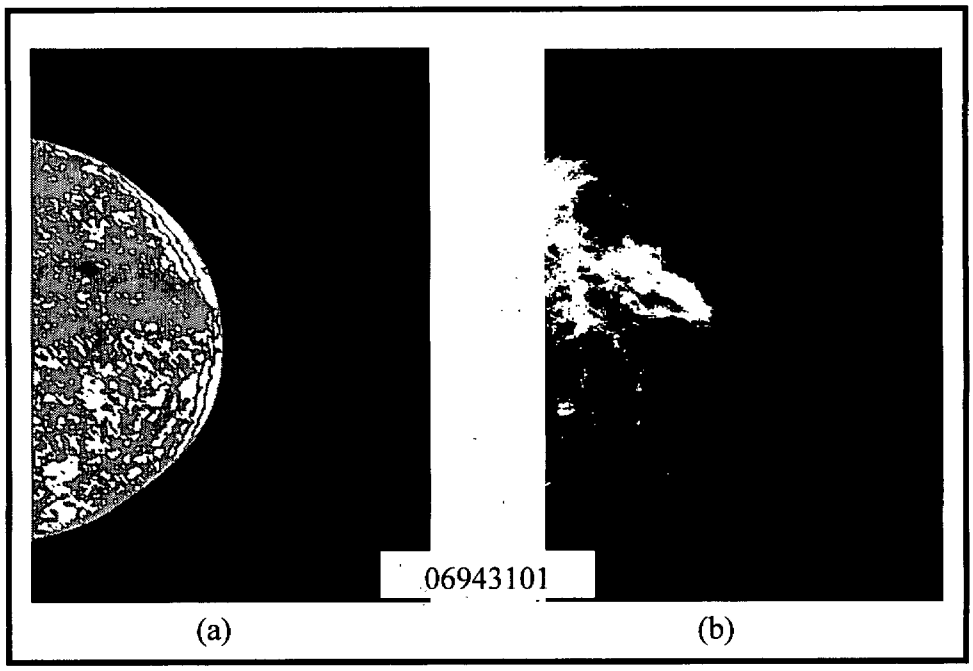


FIGURE 4.19 (a) Result of a K-mean clustering of co-occurrence features for a FCCC image. (b) Desired segmentation results based on Toronto method.

4.5 Law's Texture Mask

Law's texture energy measures were used in the evaluation because of the high number of features that could be obtained. For each image, a set of 15 different features was obtained. Of course, a greater number of features do not necessarily mean better performance, but it provides another insight to the viability of textures in this application.

Figure 4.20 demonstrates a typical suite of results obtained from using Law's texture energy measure on one of the Harvard images. The results obtained from the technique do not add any additional information to the segmentation process; the additional images actually provide a worse separation of the two tissue types than using just the gray level comparisons by themselves. The FCCC images shown in Figure 4.21 are also similar to the Harvard images, providing no additional information.

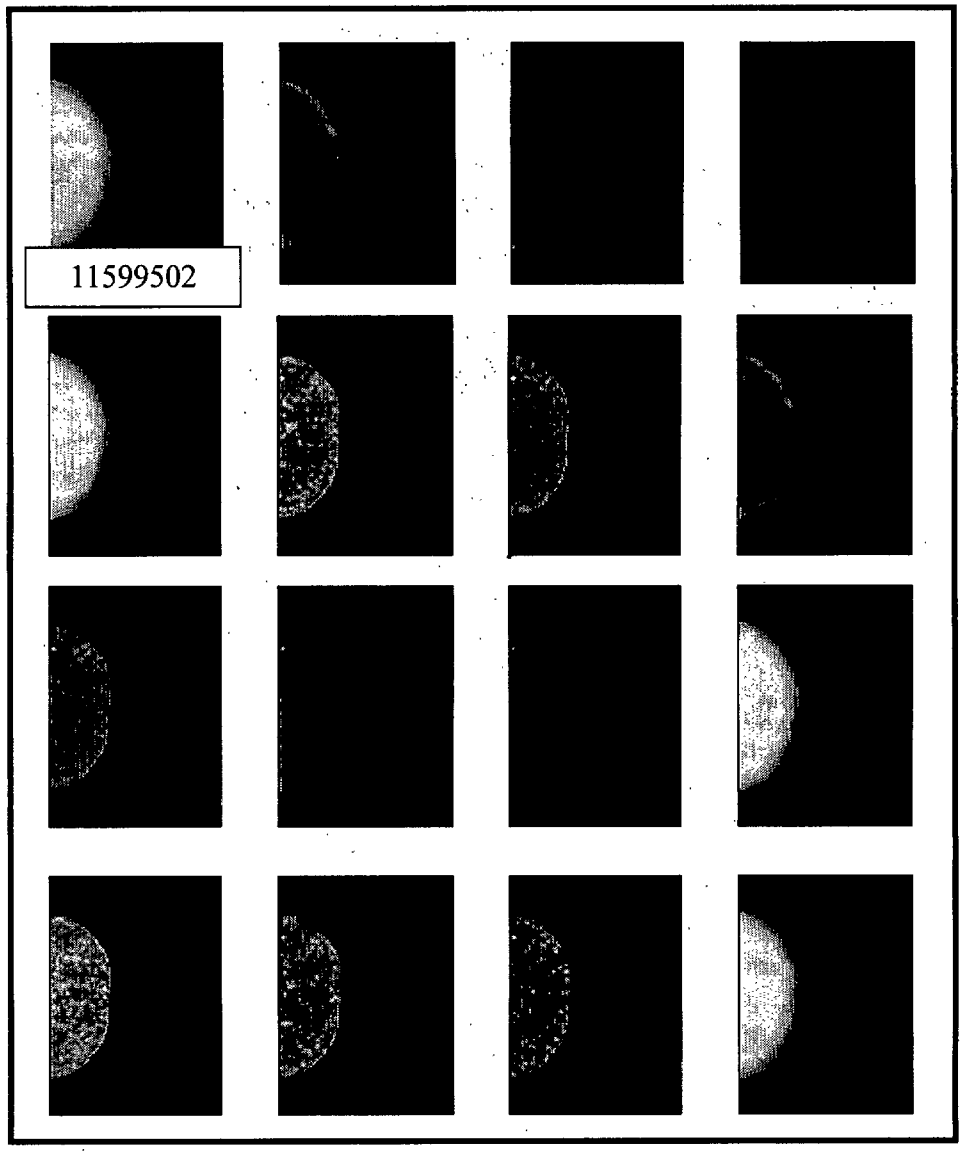


FIGURE 4.20 Typical results for the Harvard images using Law's energy texture measures.

The results from the texture segmentation methods show that the radiodense and radiolucent regions do not exhibit much difference in their texture properties. Even the methods that work for one database do not work well when evaluated using another database. There are just too many variations from mammograms obtained from different areas and databases. Because of this, a method that requires the statistics to be defined from image to image, like textures, is not viable.

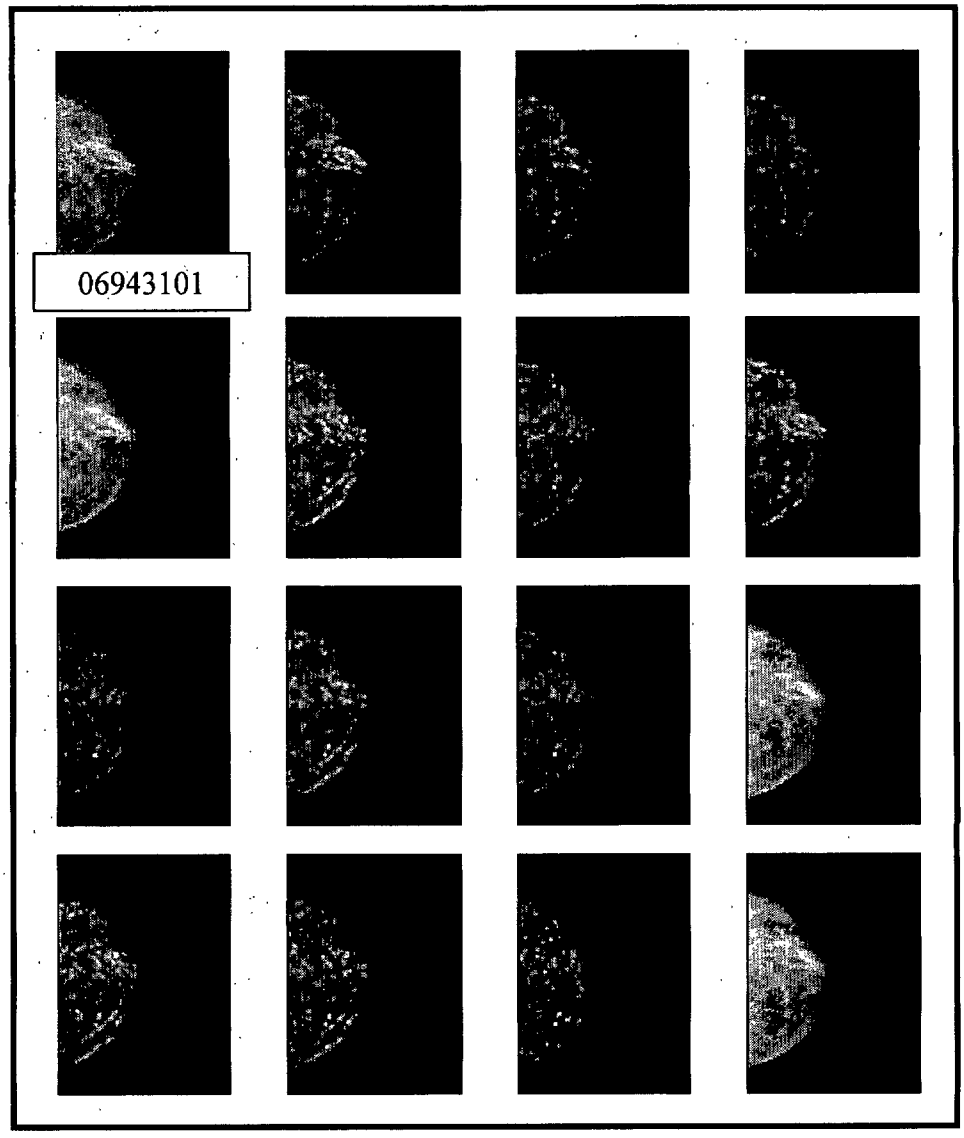


FIGURE 4.21 Typical results for the FCCC images using Law's energy texture measures.

4.6 Non-linear Transformations

Results for non-linear transformations, both pixel-based and regional-based for the Harvard images are shown in Figure 4.22. The regional-based method performs better than the pixel-based method because a single noisy pixel will not affect the result in a negative way. Using non-linear transformations allows for a better separation of radiodense and radiolucent tissue; unfortunately, this procedure is entirely subjective

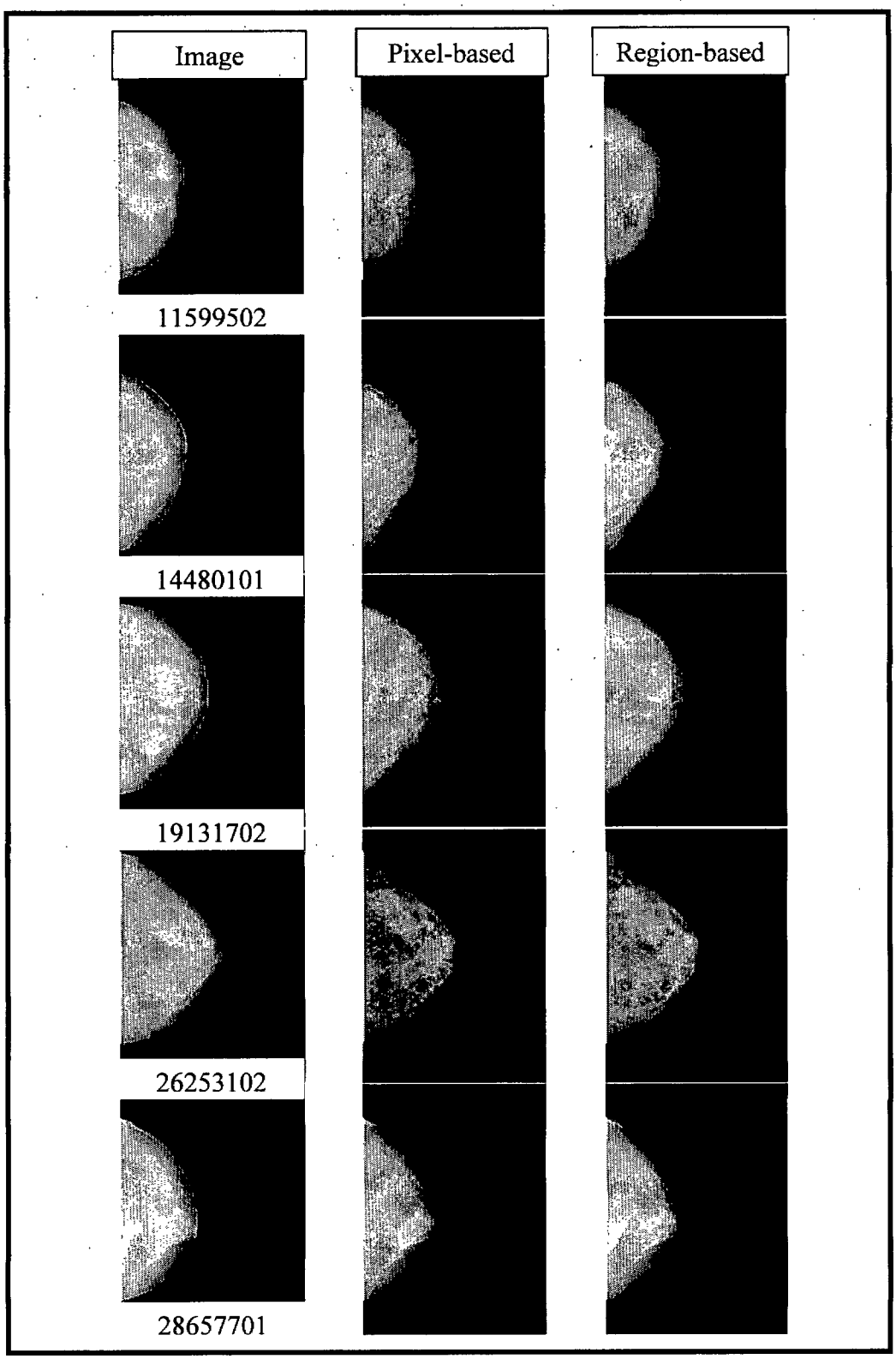


FIGURE 4.22 Results of non-linear transformation using both pixel based and regional based for the Harvard images.

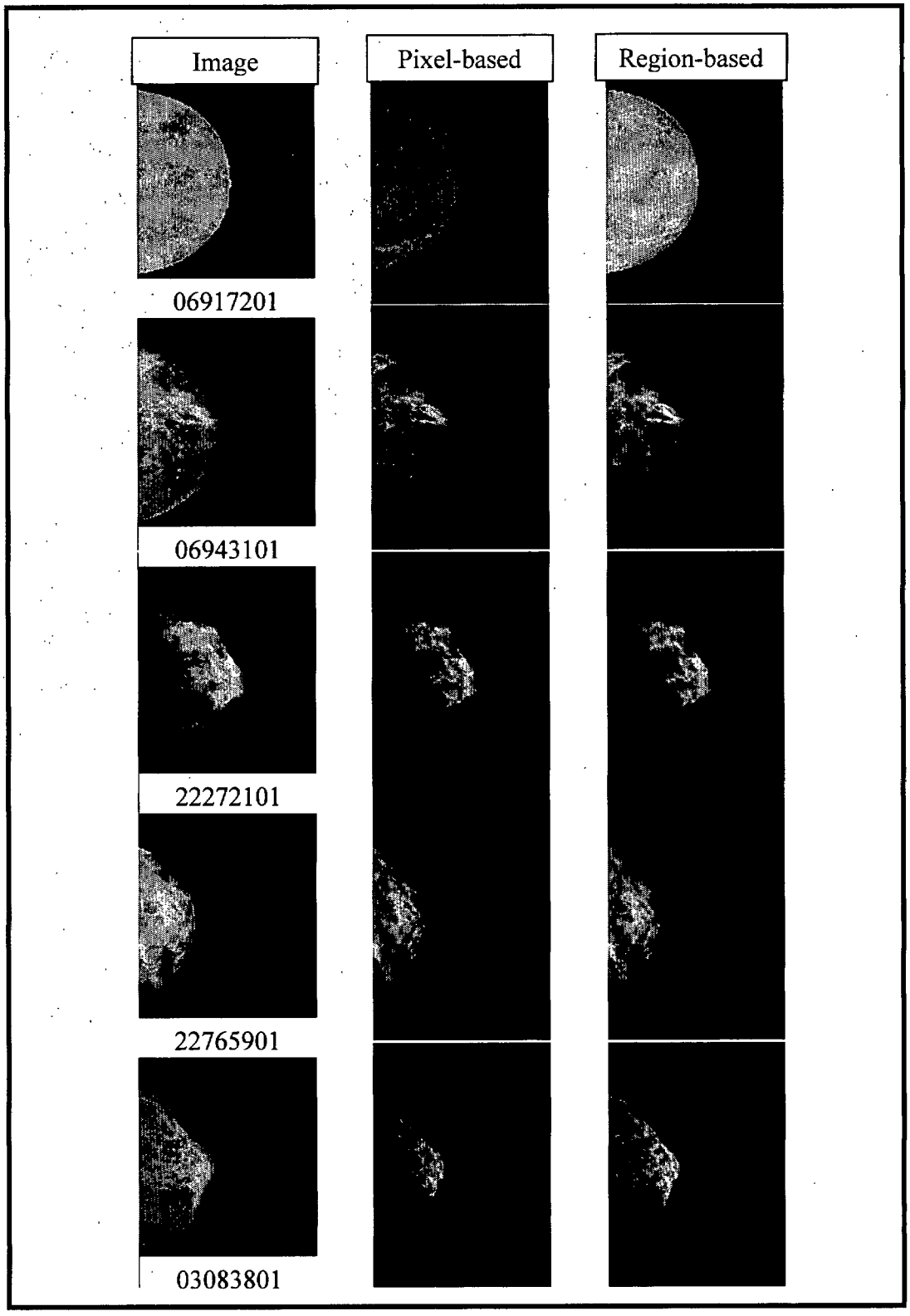


FIGURE 4.23 Results of non-linear transformation using both pixel based and regional based for the FCCC images.

and does not base the transformation on any statistic. Still, using this type of method for preprocessing images that have similar tissue characteristics, like the Harvard set does, is promising. However, using non-linear transformation for FCCC images results in poor outcomes, as shown in Figure 4.23. If the results are compared with what Dr. Celia Byrne obtained using the Toronto method, putting the images though a non-linear transformation will cause the images to always have a lower estimate than the one she obtained. Because the technique does not work well with both databases, it is not a viable method.

4.7 Gray Level Connectivity

To obtain the connected components for the mammograms, the images were first evaluated for the lowest gray level valued connected component. In this evaluation, it was assumed that the lowest connected component was radiolucent and anything with higher values was radiodense. This is not always the case for the images, but as the results show, this evaluation demonstrates how most radiodense regions are grouped up into large regions. This characteristic allows the use of regional estimation methods, such as the Local Contrast Estimation introduced in this thesis.

Once the lowest connected level was evaluated, additional regions were obtained by a random selection of a pixel that was not in a previous region. From this pixel, connectivity was established as long as the connected pixel was within a certain preset range. Because of this preset value, this method is not completely automated because it has to be changed for each image. Regardless, the method shows why evaluating radiodensity as regions rather than pixels can produce more accurate results. Figure 4.24

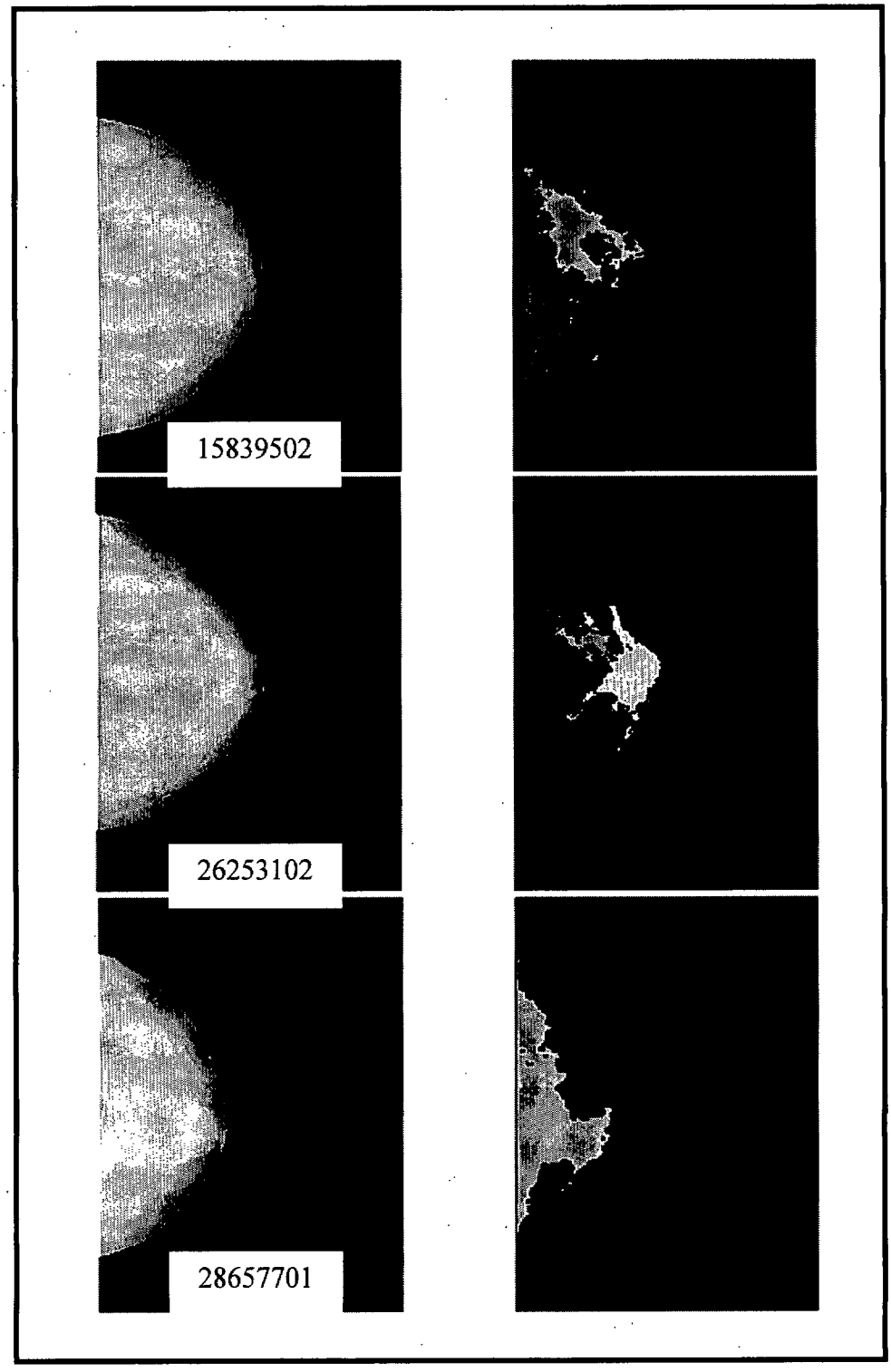


FIGURE 4.24 Connectivity results for three Harvard images.

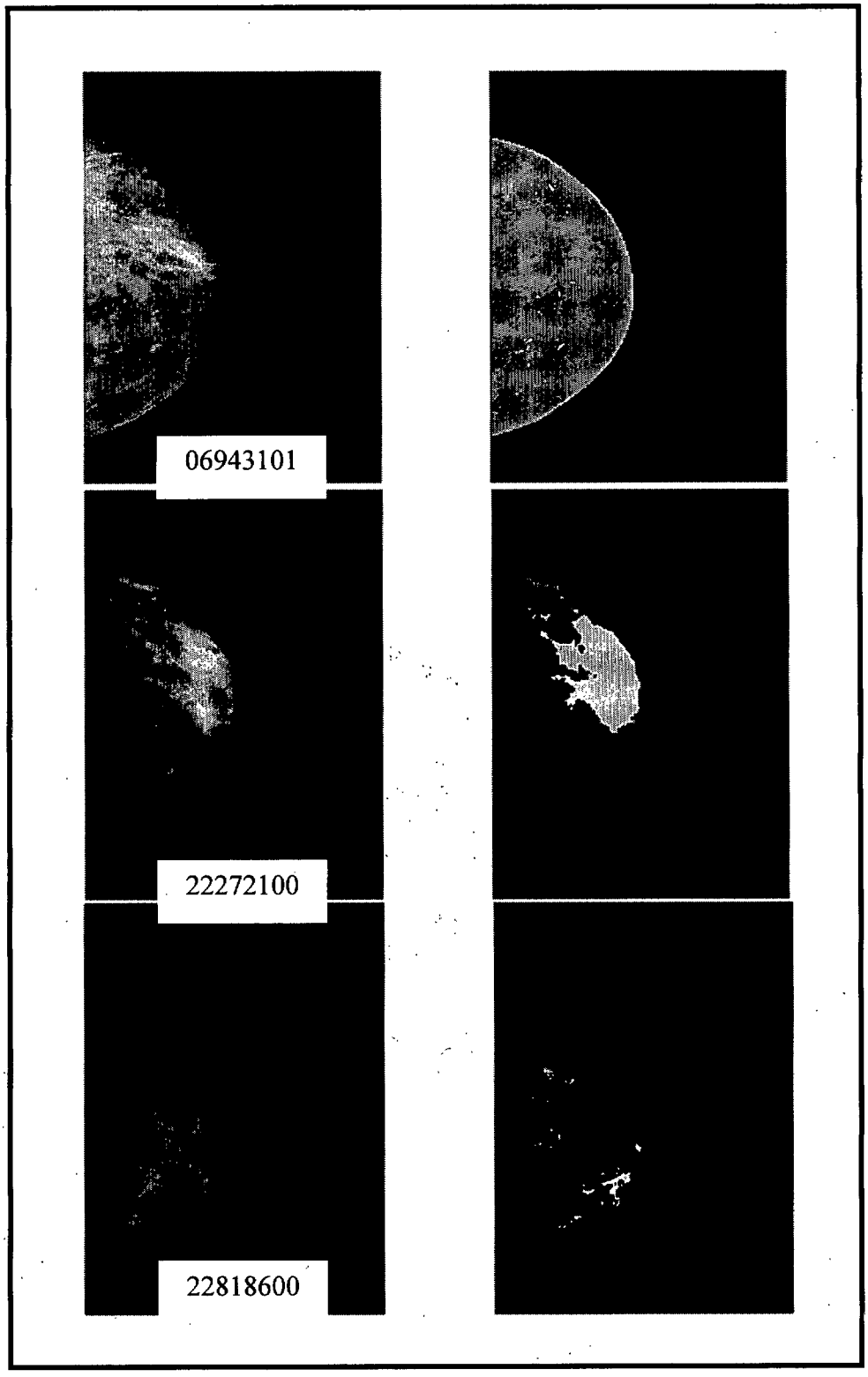


FIGURE 4.25 Connectivity results for three FCCC images.

shows results obtained from three Harvard images. Even though they are not completely accurate, the regions demonstrate the connective nature of radiodense and radiolucent tissue. Figure 4.25 shows typical results from the FCCC database.

4.8 Image Preprocessing

Figure 4.26 demonstrates some results of the stripe removal procedure from the Harvard set. The stripes are not completely removed, but do not pose much of a problem because results from the masking algorithm show that the complete algorithm is very noise resistant. The results for the FCCC images are shown in Figure 4.24.

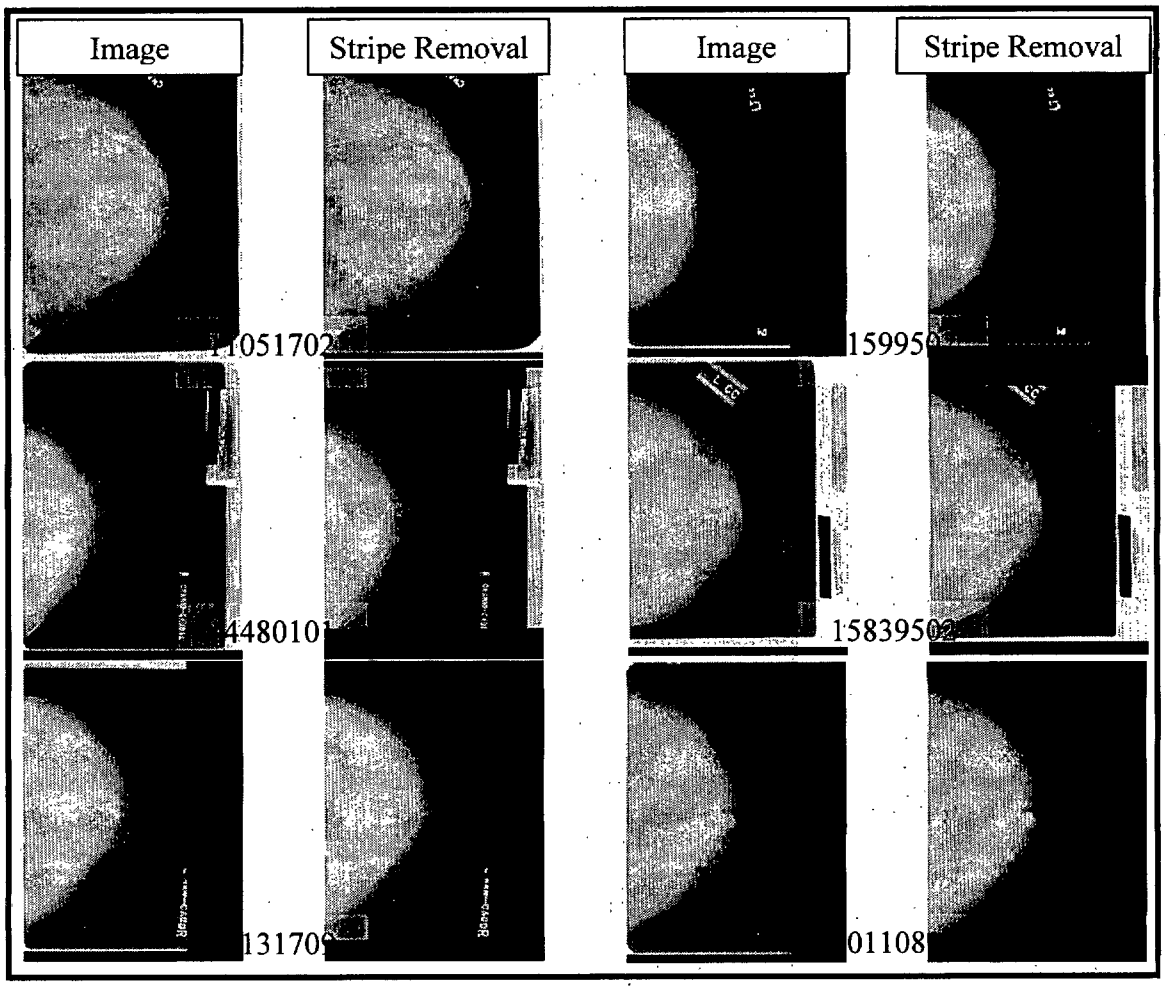


FIGURE 4.26 Stripe removal results for some of the Harvard images.

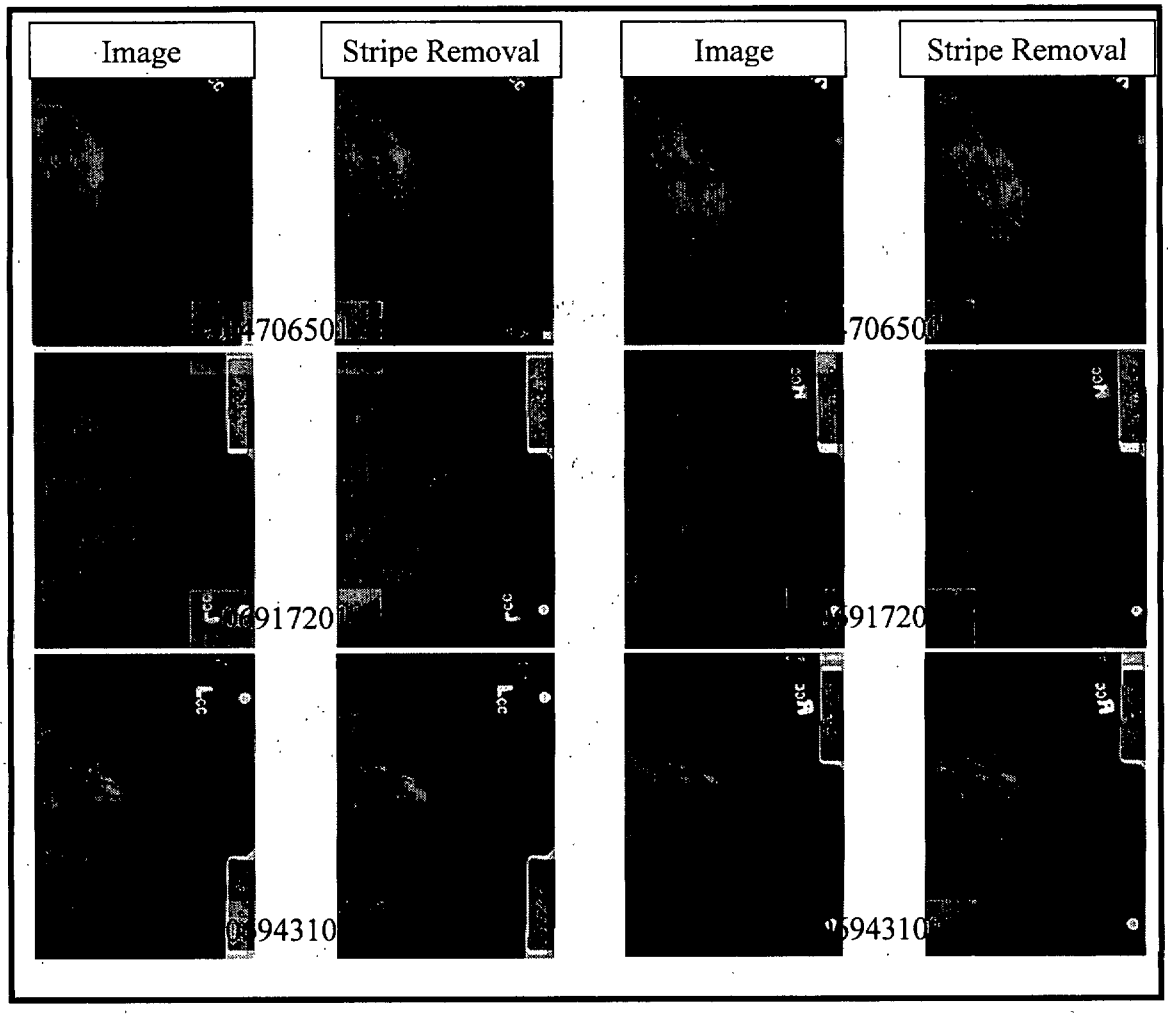


FIGURE 4.27 Stripe removal results for some of the FCCC images.

4.9 Tissue Segmentation

This section is the first portion of the results that can be compared with previous methods that have been created at Rowan University. One of the main problems with the previous method for tissue segmentation is that it was too generalized and very susceptible to noise. Figure 4.28 shows a comparison of the mask created by the method introduced in this thesis and the method created previously at Rowan University [1] for some of the Harvard methods. The green lines represent the edge detected by the method in this thesis. The yellow lines represent the edge created by the previous method. The results

for the method introduced in this thesis are more precise than the previous method. The results of the FCCC images, compared in Figure 4.29, again demonstrate the accuracy of the method introduced in this thesis.

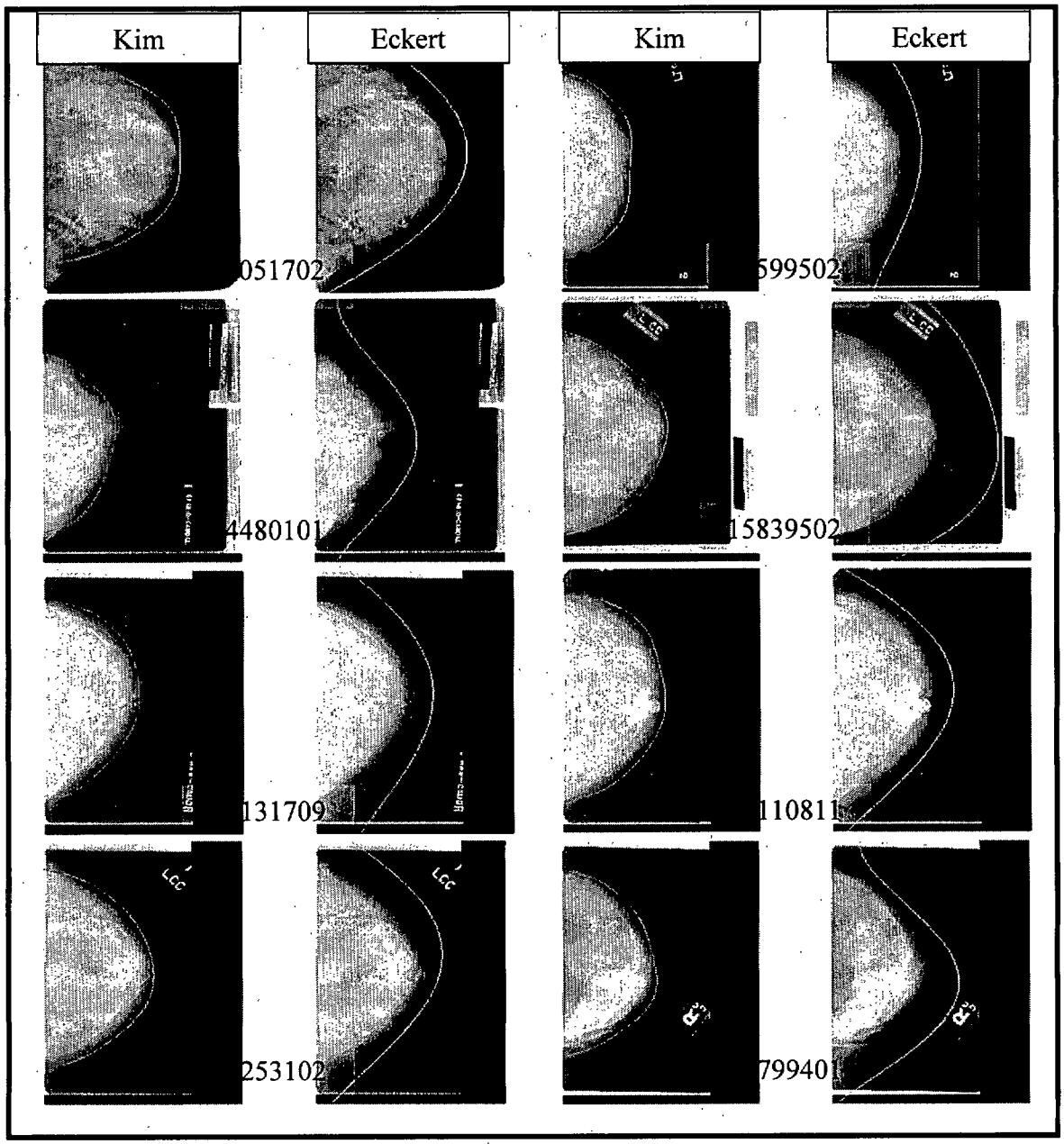


FIGURE 4.28 A comparison of the tissue segmentation method introduced in this thesis (green) and the tissue segmentation method created by Eckert (yellow) for some Harvard images.

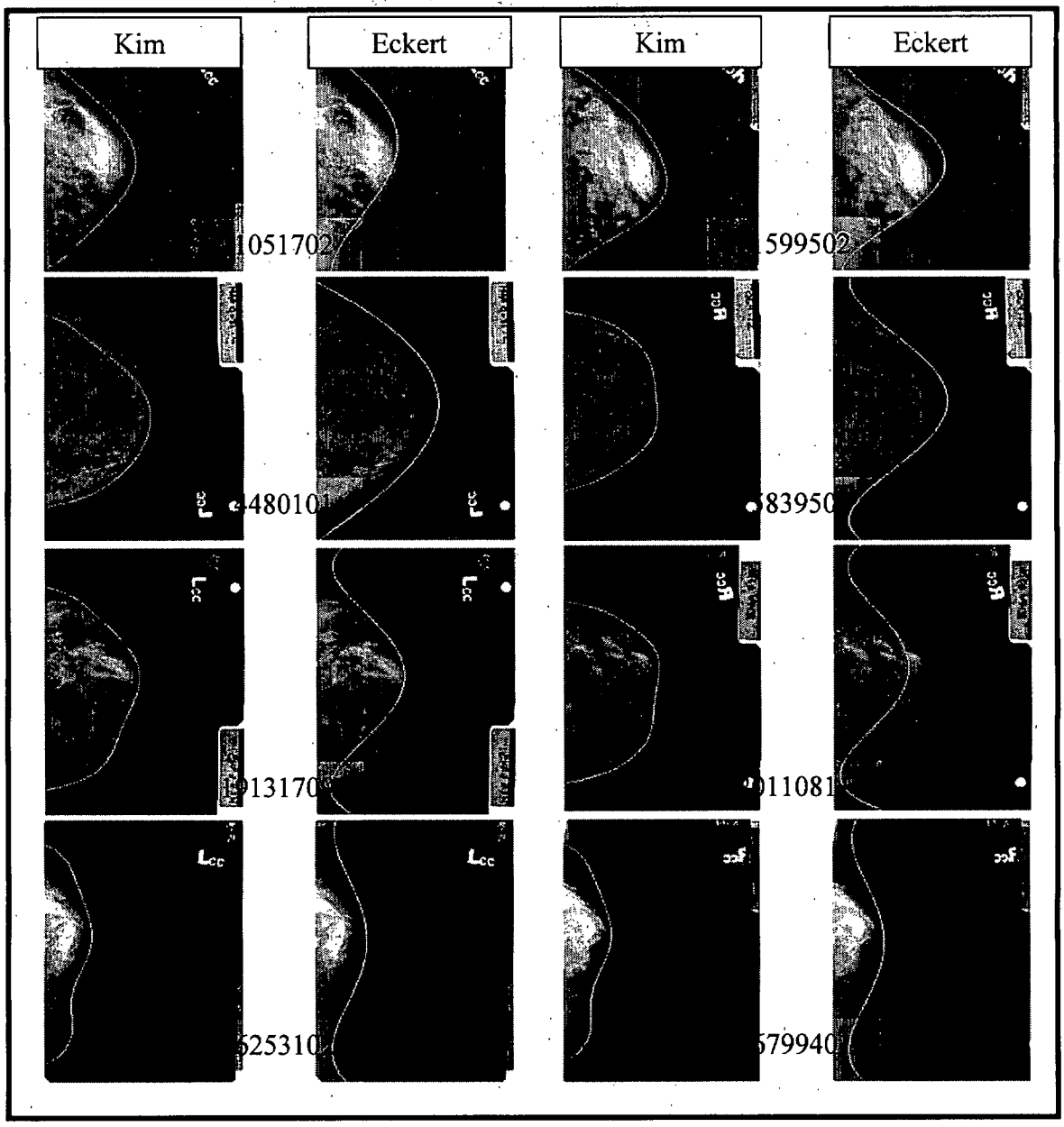


FIGURE 4.29 A comparison of the tissue segmentation method introduced in this thesis (green) and the tissue segmentation method created by Eckert (yellow) for some FCCC images.

The images also play an important role in the comparison of final results. The radiodensity results from Eckert's method are based on masks that were created manually

rather than through the method that was introduced in his thesis. Therefore, the final results also show an accurate indication of true automated performance.

4.10 Compression

Figure 4.30 shows an example of the compression masks that were created by the Eckert method. There are several major problems with these masks. The maximum and minimum value of compression applied is based on the breast length, where A is the maximum value and B is the minimum value. Equations 4.2 and 4.3 represent the relationship between the A and B values:

$$T_v(x_e) = B = kT_{CNP}e^{-\frac{x_e^2}{2\sigma^2}} \tag{4.2}$$

$$\sigma^2 = \frac{-x_e^2}{2\log(\frac{B}{A})} \tag{4.3}$$

The Eckert procedure assumes that the function of compression is represented by a Gaussian curve. This assumption is reasonable, but is made even more subjective by the amount of compression applied. For some images, the Eckert method assumes that there is more than 200% compression in various parts of the breast. This assumption is not based on any mathematical principles but rather experimental values that may have been a result of over fitting. As the final results of Eckert's method show, many of the final processed images do not correlate well with what is considered to be the connective nature of radiodense tissue. Because of the uncertainty of the actual compression amount and the function that it follows, this thesis only assumes that at most, compression

increases the radiodensity by 20%. The compression function was also left as a Gaussian curve.

In addition to some problems with the theoretical concepts behind the compression, there are some implementation issues with the code. Because the mask was created by fitting as many lines as possible between the breast edge and chest wall, this caused resolution and scaling problems. Figure 4.31 shows how when lines are fit in this method, the sides of the breast become saturated but the center of the breast still contains many empty spaces. Because of the empty spaces, when put through the multi-pass low-pass filtering operation that the method uses, the centers become overall lower than the sides. As a result, the Gaussian decay that is modeled is inaccurate. Also, because the

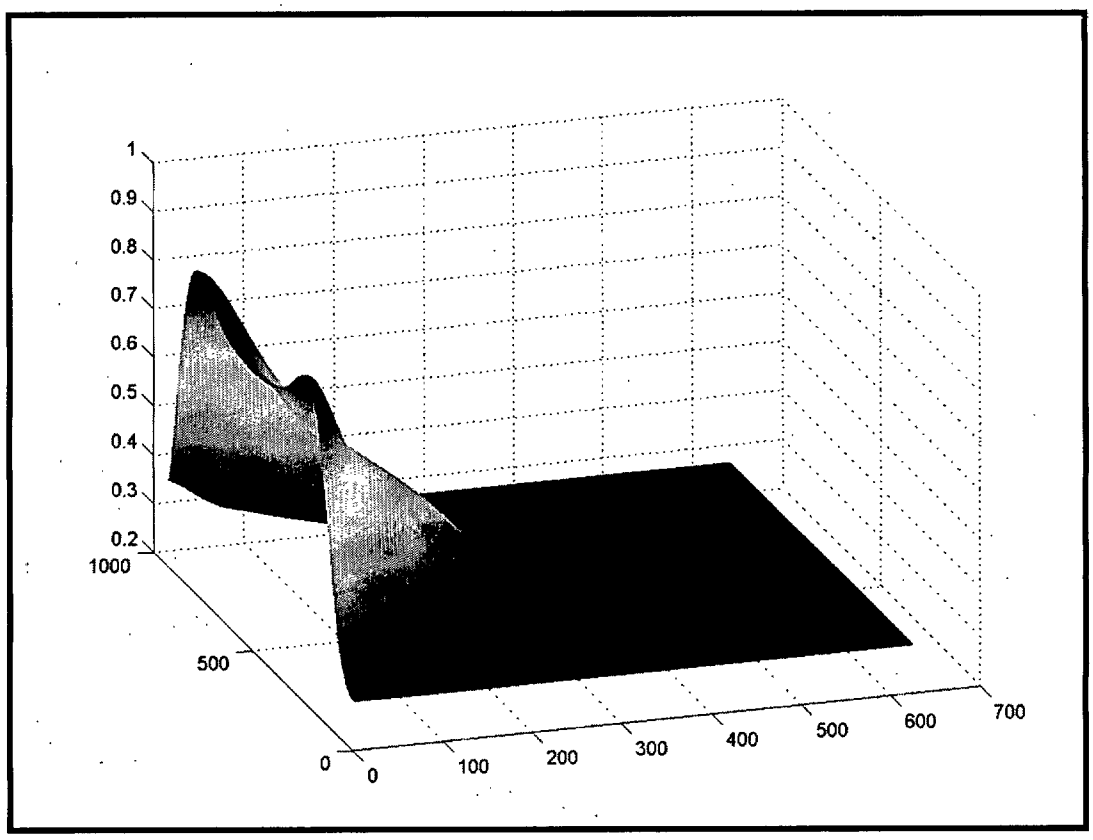


FIGURE 4.30 Compression mask created by Eckert method.

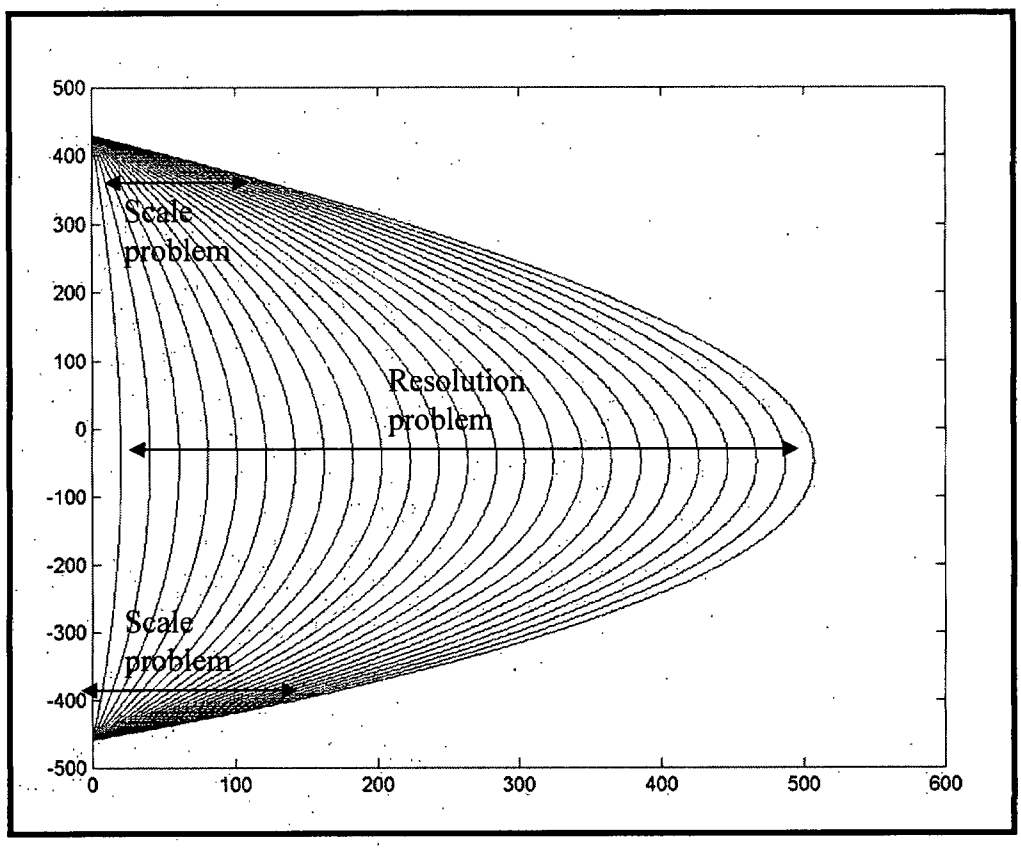


FIGURE 4.31 The problems associated with the SV-CNP compression mask.

compression in the Eckert method is based on the incorrect tissue mask, more error is introduced into the process.

Figure 4.32 shows some examples of the new compression masks created by this thesis for the Harvard database. The main aspect to observe is how the function decreases from the breast edge to the chest wall. At all points along the chest wall, the values should be the same. The breast edge should also have the same values. Figure 4.33 shows 10 different FCCC images that have been put through the same procedure.

Because of the combination of this process along with other aspects of the Eckert approach, some of the final quantities that Eckert obtained are a result of multiple

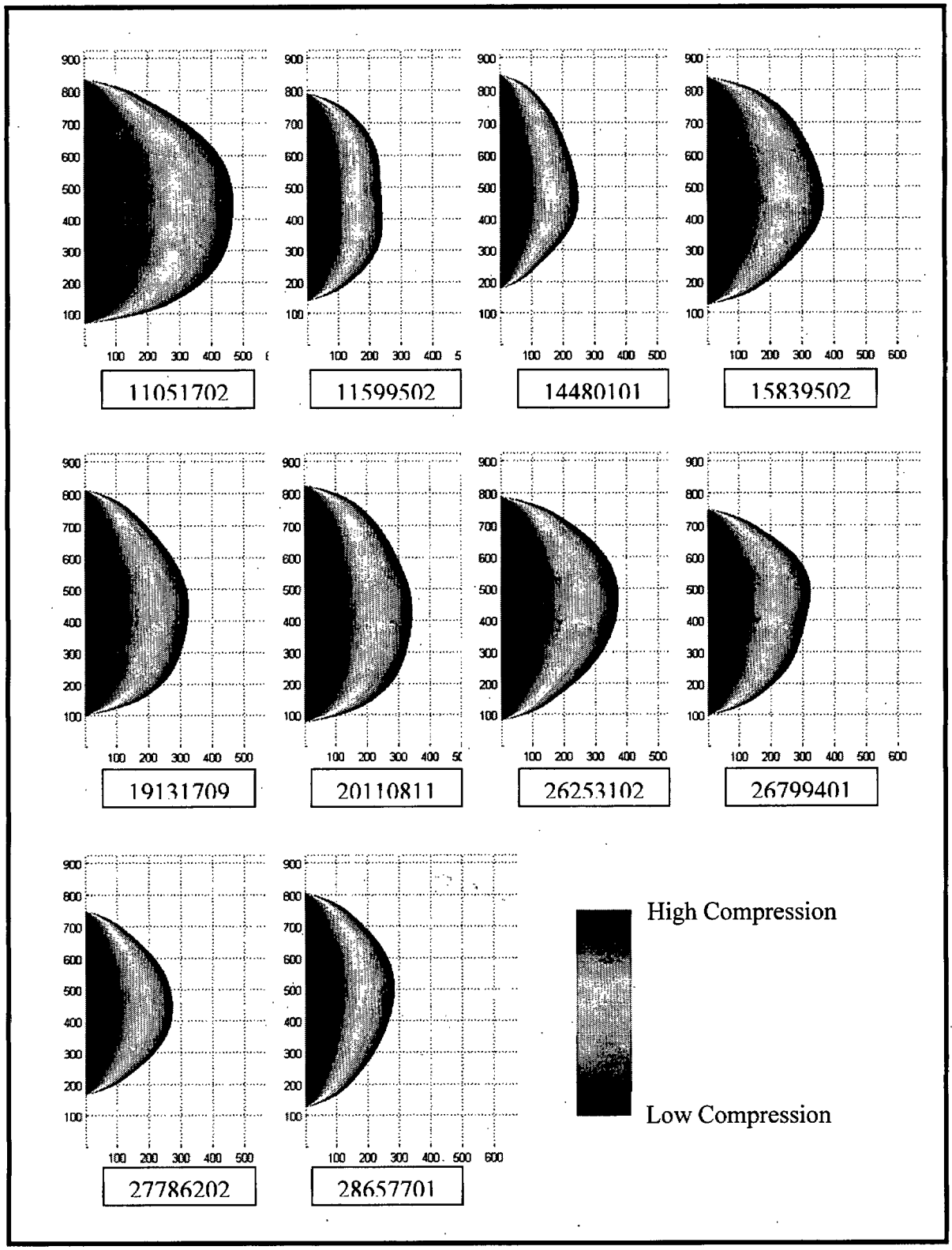


FIGURE 4.32 Compression masks created by the method introduced in this thesis for the 10 Harvard images.

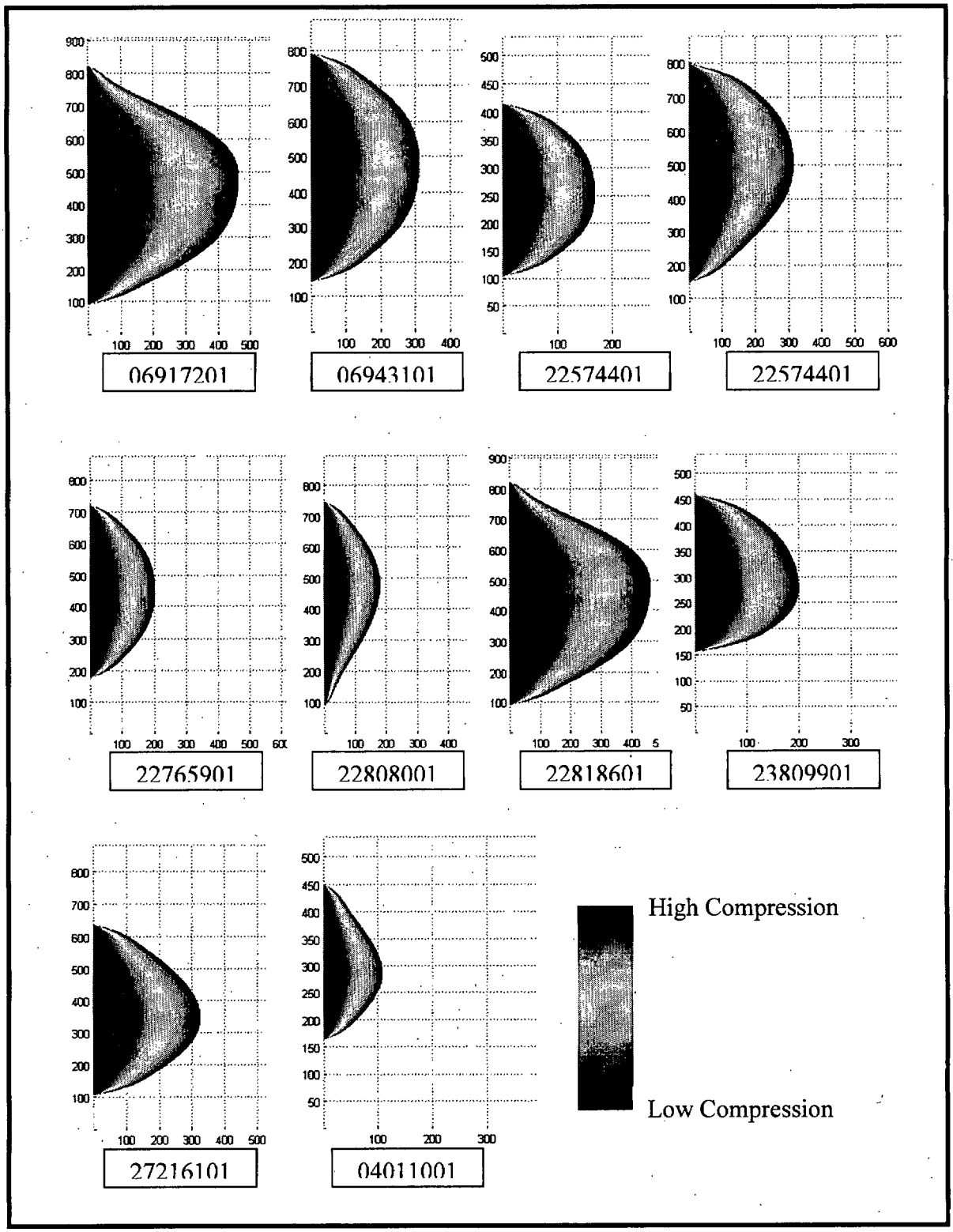


FIGURE 4.33 Compression masks created by the method introduced in this thesis for 10 FCCC images.

incorrect procedures. Even the main focus of the Eckert approach, compression adjustment, is implemented through several fitting procedures that are questionable. The results shown in the next section compare the outcomes both quantitatively (percentages) and qualitatively (images).

4.11 Local Contrast Estimation vs. Previous Methods

The section first shows the results of each algorithm, the CNP, SV-CNP, and the LCE separately. The Harvard dataset only has one set of results. For the FCCC database, there are two different sets: one containing all FCCC images and one for all non-flagged images. The flagged images were all images that Dr. Celia Byrne was unconfident about the percentages that she had provided. In addition, the second half of this section shows a comparison of the three methods for the two databases combined.

Figure 4.34 and Figure 4.35 show results obtained for the Harvard images using the CNP method that has been previously created at Rowan University. The final segmentation results look well defined and tend to correlate with what are the high radiodense regions. The percentages obtained from this method, as shown in Table 4.3, perform very well compared to Dr. Celia Byrne and show that it is correlated at a value of 0.911 and has a MSE value of 1091.2 (Table 4.4, Figure 4.36). Unfortunately, there is one crucial problem with this method. If we observe Equation 4.4, used to obtain the thresholds used for segmentation:

$$T_{CNP} = \frac{\mu_1 + \mu_2}{2} + \left(\frac{\alpha - \sigma^2}{\alpha}\right)\left(\frac{\mu_2 - \mu_1}{2}\right) \tag{4.4}$$

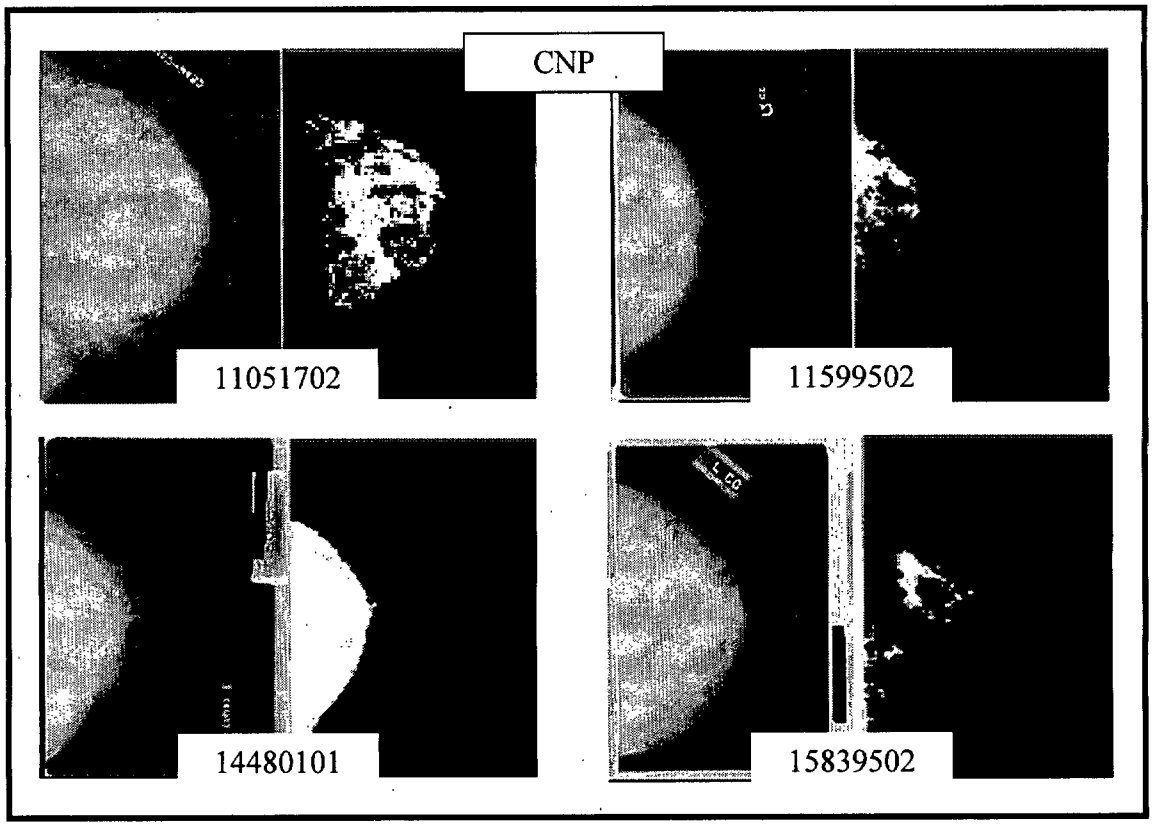


FIGURE 4.34 Visual results of first four Harvard images using CNP method.

TABLE 4.3 Percentage comparisons of the CNP method to the estimates given by Dr. Celia Byrne using the Toronto method.

Image Number	Radiodensity Estimate of Toronto Method	Radiodensity Estimate of CNP	MSE of CNP
11051702	21.4	21.07	0.11
11599502	12.9	14.95	4.20
14480101	40.8	71.00	912.04
15839502	13.3	7.50	33.64
19131709	1.5	0.34	1.34
20110811	2.5	1.65	0.72
26253102	22.5	23.70	1.44
26799401	55.3	47.82	55.95
2778620	50.1	58.00	62.41
28657701	33.6	38.00	19.36

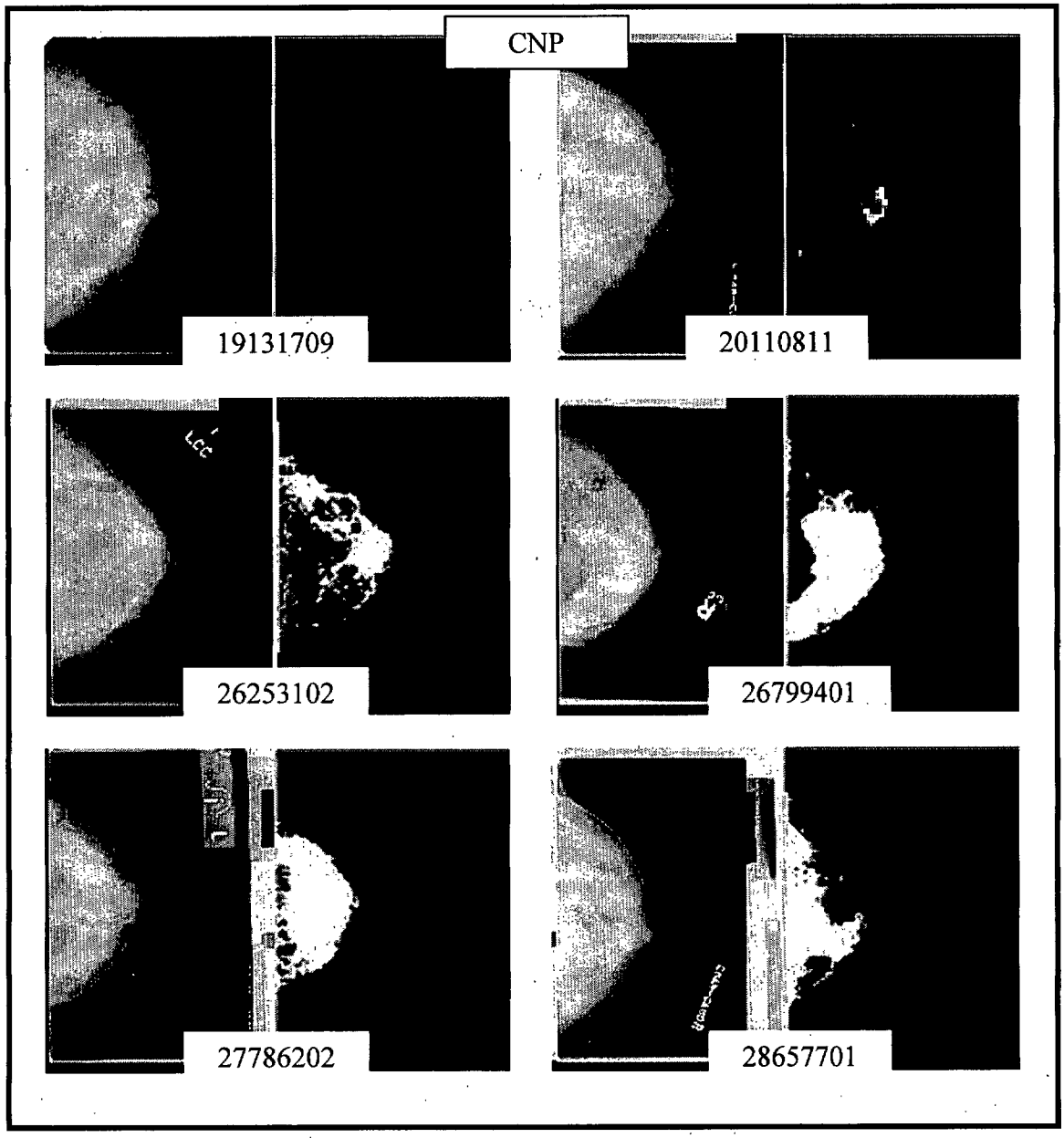


FIGURE 4.35 Visual results of last six Harvard images using CNP method.

it is seen that even though μ_1, μ_2, σ are based on image statistics, the α parameter is based only on experimental analysis of what α would give the best results. Because it is not based on any image statistic, the results obtained for Harvard database were produced

using three α parameters for only ten images. This essentially makes this algorithm a semi-automated procedure where a new α parameter is chosen for each new group of

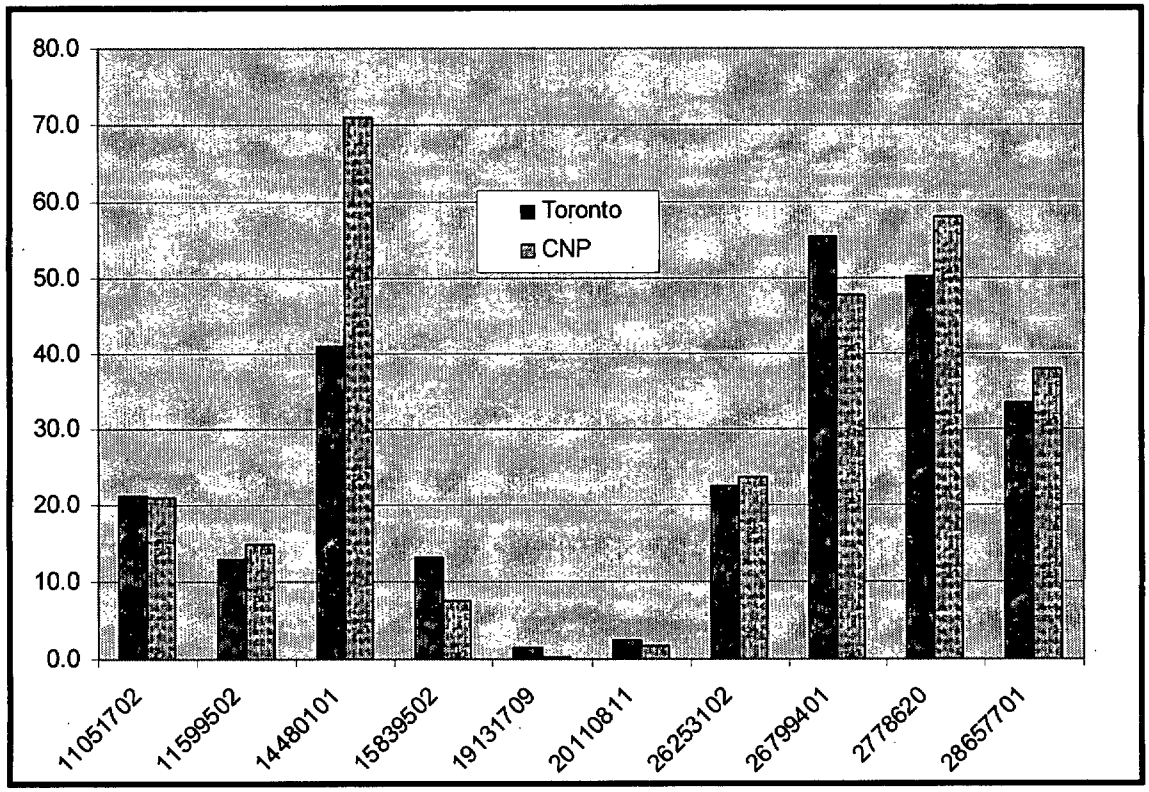


FIGURE 4.36 Comparison of percentages between Toronto method and CNP for Harvard images.

TABLE 4.4 Correlation and MSE of the CNP to Toronto for all Harvard images.

Correlation of CNP and Toronto for Harvard images
0.911871657
MSE Harvard images
1091.21

images. Therefore, when the algorithm is exercised on a new database where the α parameter has not been previously chosen for it, it does not perform well. The results in Table 4.5 show the performance that the CNP achieved for FCCC images. The CNP

TABLE 4.5 The percentages obtained for FCCC database for the CNP. The Correlation and MSE are compared with the Toronto method.

Image Number	Radiodensity Estimate of Toronto Method	Radiodensity Estimate of CNP	MSE of CNP
04706501	51.7	24.90	718.2
04706500	47.4	27.10	412.1
06917201	6.7	49.90	1866.2
06917200	7.3	47.30	1600.0
06943101	20.0	43.60	557.0
06943100	20.5	37.50	289.0
20392001	54.3	26.00	800.9
20392000	57.8	30.30	756.3
22272101	47.7	20.40	745.3
22272100	26.1	22.60	12.3
22574401	30.2	34.90	22.1
22574400	24.0	35.30	127.7
22645201	12.1	34.80	515.3
22645200	12.3	48.30	1296.0
22733501	10.1	44.30	1169.6
22733500	25.0	37.90	166.4
22765901	63.7	38.90	615.0
22765901	73.1	42.10	961.0
23152500	37.5	23.40	198.8
22808000	48.9	33.80	228.0
22818601	27.9	26.40	2.3
22818600	27.4	26.10	1.7
22819201	42.4	20.90	462.25
22819200	81.9	21.40	3660.3
23809901	36.6	21.80	219.0
23809900	18.2	24.00	33.6
27216101	17.2	42.40	635.0
23152501	37.5	28.80	75.7
03083801	25.0	34.10	82.8
03083800	9.1	36.00	723.6
04011001	79.9	40.20	1576.1
04011000	53.0	34.10	357.2
06943101	20.0	43.60	557.0
06943100	20.5	37.50	289.0

Correlation of CNP and Toronto on FCCC images
-0.390753886
MSE FCCC images
21732.72
Correlation of CNP and Toronto on FCCC images (without flagged)
0.30616301
MSE FCCC images (without flagged)
14381.54
Images Flagged by Dr. Celia Byrne

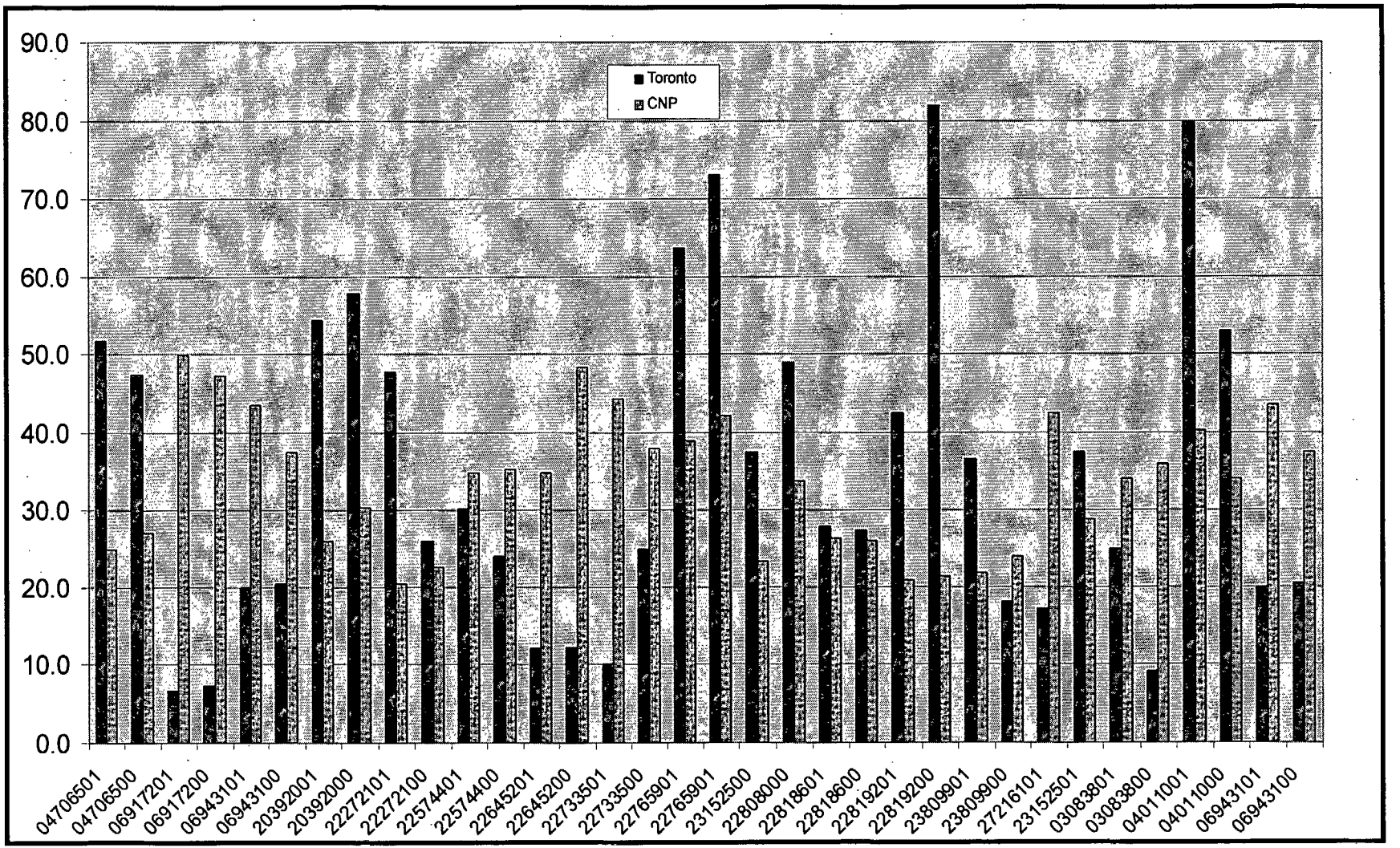


FIGURE 4.37 Comparison of CNP vs. Toronto method for FCCC images.

percentages for all FCCC images, shown in Table 4.4 (Figure 4.37), are actually negatively correlated with the estimates given by the Toronto method at a value of -0.390 with a MSE of 21732. The results without the flagged images obtain a correlation value of 0.306 and a MSE of 14381.5. This shows that the CNP is very dependent on a changing α parameter, and this essentially makes it a supervised algorithm.

Figure 4.38 and Figure 4.39 show the results obtained from the SV-CNP algorithm. If the performance is evaluated on just the percentages alone, the algorithm works well. The percentages obtained from the SV-CNP, as shown in Table 4.6 (Figure 4.40, Table 4.7) were correlated at a value of 0.92 and had a MSE of 495.1 compared to the values given by Dr. Celia Byrne using the Toronto method. Even though these numbers are better than the ones obtained by the CNP, they can be misleading. The SV-CNP has two main issues as previously discussed. First, it uses the same parameters that the CNP uses for initial threshold selection. The threshold is still based on the α parameter that is user dependent. Second, the compression masks sometimes compensate for more than 200% in some locations. Because of this adjustment, some regions are always classified as radiolucent regardless of how high the gray level values are. As a result, even though the SV-CNP uses the same α parameters as stated, the final segmentation result does not correlate very well with the actual radiodense regions. In all the Harvard images, SV-CNP assumes that there is almost no radiodense tissue close to the chest wall, while most of the time it automatically classifies tissue close to the breast edge as radiodense. For this to be the case realistically for all 10 Harvard images and all 34 FCCC images is highly unlikely.

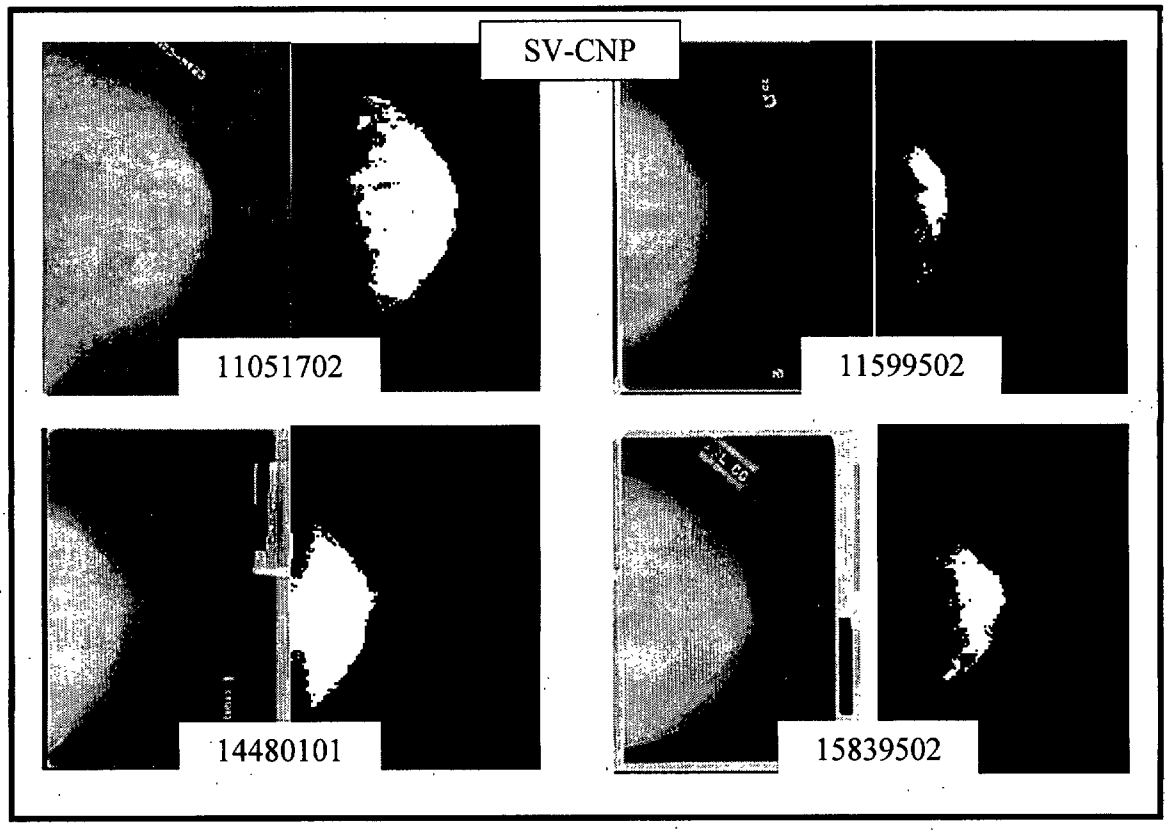


FIGURE 4.38 Results of first four Harvard images using SV-CNP.

TABLE 4.6 Percentage comparisons of the SV-CNP method to the estimates given by Dr. Celia Byrne using the Toronto method.

Image Number	Radiodensity Estimate of Toronto Method	Radiodensity Estimate of SV-CNP	MSE of SV-CNP
11051702	21.4	31.6	104.0
11599502	12.9	11.0	3.7
14480101	40.8	47.8	48.7
15839502	13.3	12.7	0.4
19131709	1.5	0.4	1.2
20110811	2.5	6.0	11.9
26253102	22.5	20.6	3.6
26799401	55.3	38.4	284.6
2778620	50.1	45.0	26.0
28657701	33.6	36.9	10.9

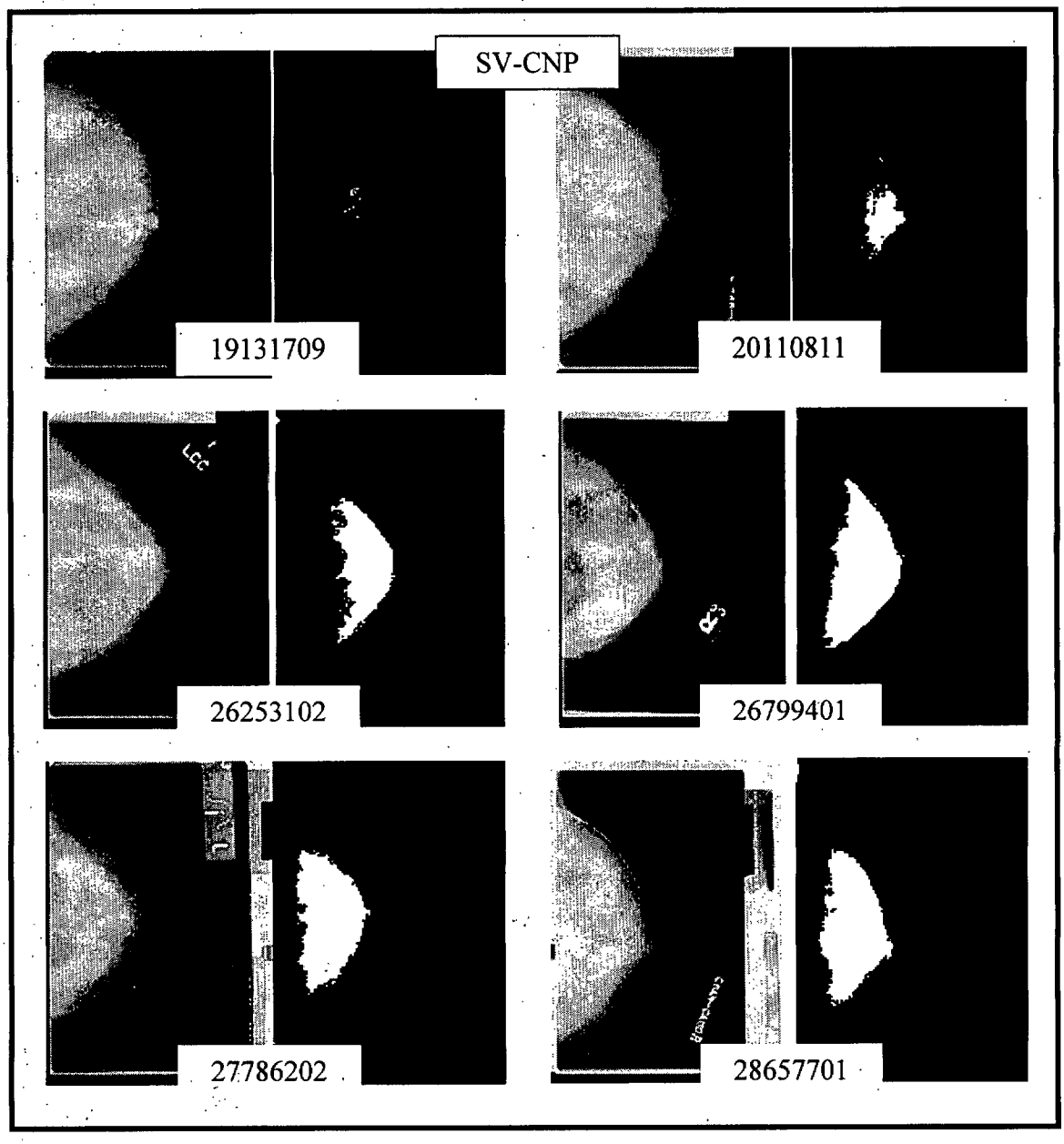


FIGURE 4.39 Visual results of last six Harvard images using SV-CNP.

Table 4.8 (Figure 3.41) shows the results obtained with the FCCC images using the SV-CNP. The performance of the algorithm decreases with a correlation of 0.48 and a MSE of 15127.58. Without the flagged images, the performance increases to a correlation of 0.733 and a MSE of 5425.79. As stated before, the SV-CNP has many issues that make one question the validity of the results. Also, the FCCC compression

adjustments have been made for the database alone. Therefore, the two separate

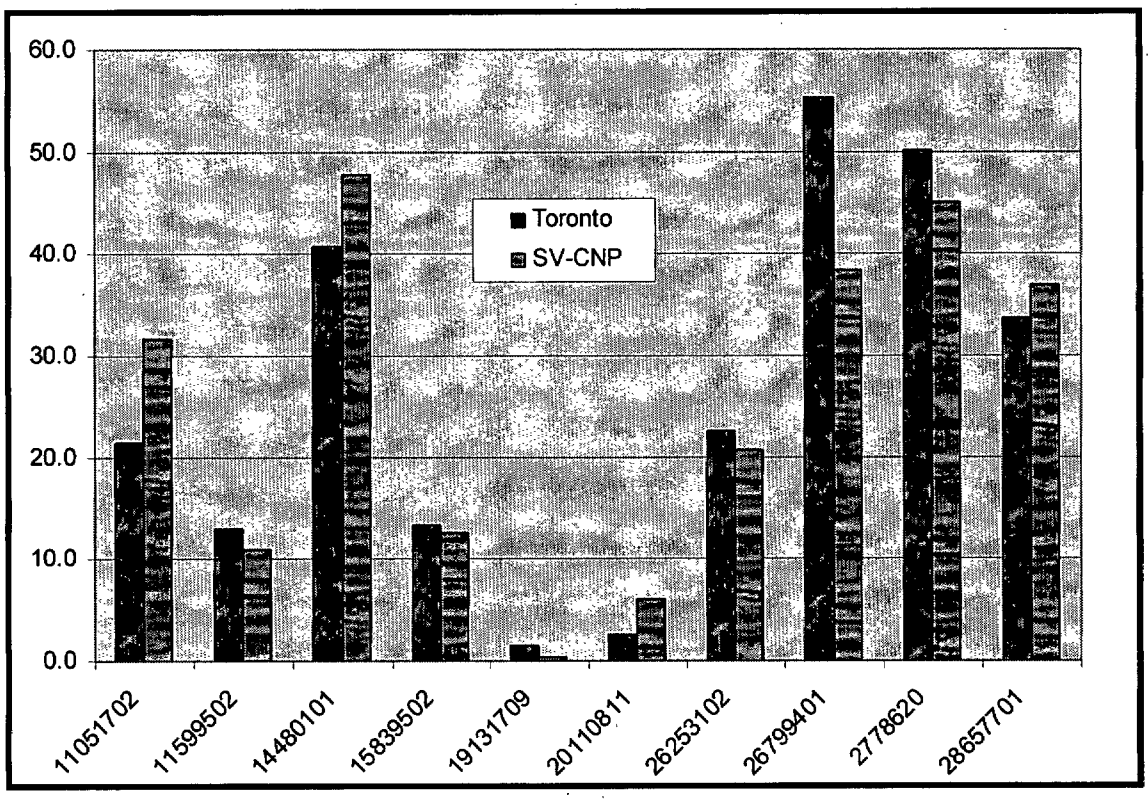


FIGURE 3.40 Comparison between the SV-CNP and Toronto method for the Harvard images.

TABLE 4.7 Correlation and MSE of the SV-CNP to Toronto for all Harvard images. Of all three methods, the SV-CNP has the highest correlation with the Harvard images.

Correlation of SV-CNP and Toronto on Harvard images
0.920639398
MSE Harvard images
495.10

TABLE 4.8 The percentages obtained for FCCC database with the SV-CNP. The Correlation and MSE are compared with the Toronto method.

Image Number	Radiodensity Estimate of Toronto Method	Radiodensity Estimate of CNP	MSE of SV-CNP
04706501	51.7	15.60	1303.2
04706500	47.4	17.80	876.2
06917201	6.7	13.30	43.6
06917200	7.3	6.50	0.6
06943101	20.0	17.40	6.8
06943100	20.5	16.80	13.7
20392001	54.3	21.70	1062.8
20392000	57.8	56.00	3.2
22272101	47.7	18.60	846.8
22272100	26.1	19.10	49.0
22574401	30.2	35.00	23.0
22574400	24.0	38.80	219.0
22645201	12.1	14.20	4.4
22645200	12.3	22.90	112.4
22733501	10.1	26.00	252.8
22733500	25.0	16.50	72.3
22765901	63.7	31.80	1017.6
22765901	73.1	30.10	1849.0
23152500	37.5	28.80	75.7
22808000	48.9	40.40	72.3
22818601	27.9	12.70	231.0
22818600	27.4	11.10	265.7
22819201	42.4	16.2	686.4
22819200	81.9	14.70	4515.8
23809901	36.6	21.50	228.0
23809900	18.2	30.70	156.3
27216101	17.2	9.90	53.3
23152501	37.5	23.40	198.8
03083801	25.0	25.30	0.1
03083800	9.1	19.10	100.0
04011001	79.9	52.20	767.3
04011000	53.0	53.30	0.1
06943101	20.0	17.40	6.8
06943100	20.5	16.80	13.7

Correlation of SV-CNP and Toronto on FCCC images
0.481121229
MSE FCCC images
15127.58
Correlation of SV-CNP and Toronto on FCCC images (without flagged)
0.7329646
MSE FCCC images (without flagged)
5425.79
Images Flagged by Dr. Celia Byrne

databases have their own polynomial fit for the compression adjustments, which do not adequately evaluate cross database performance.

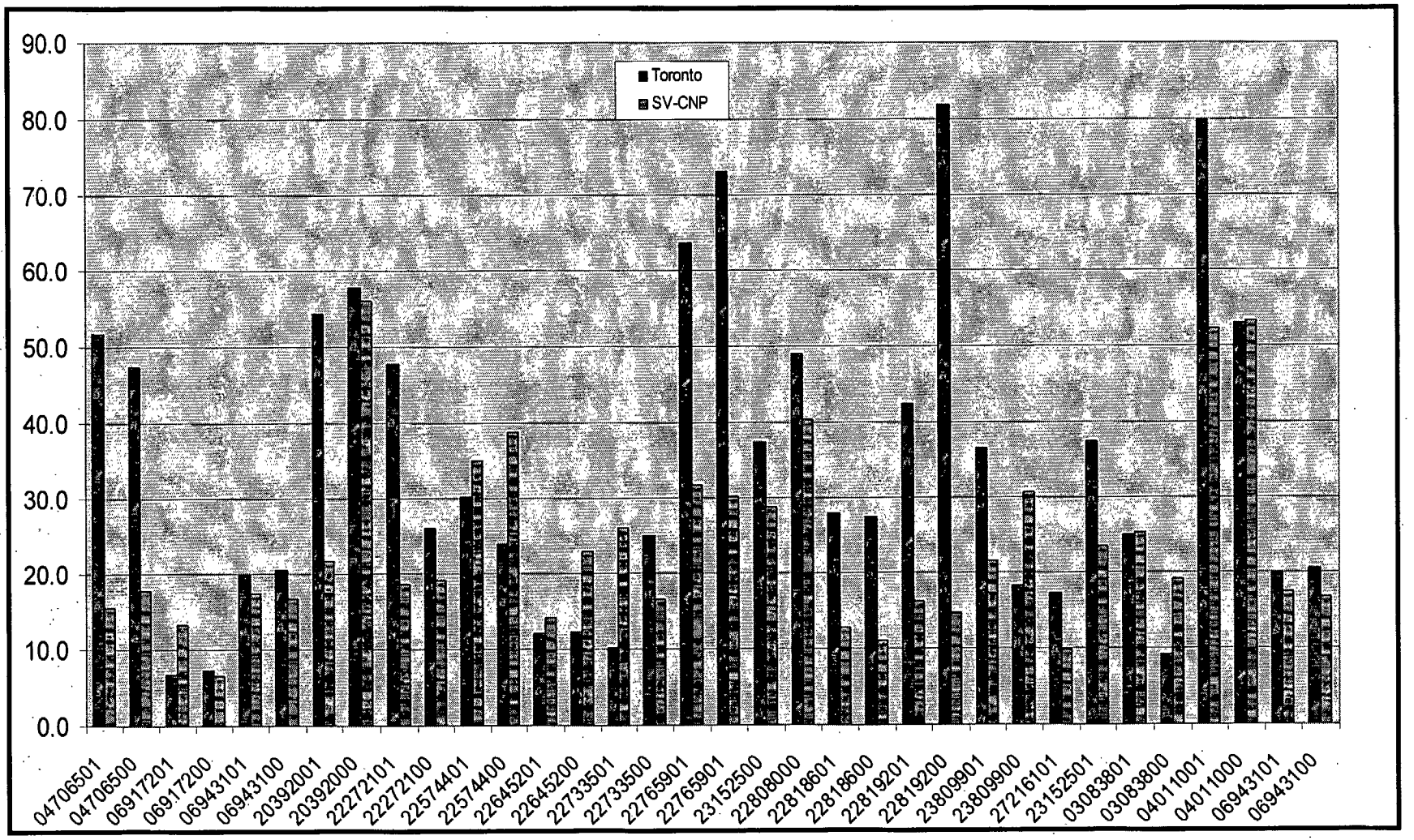


FIGURE 4.41 Comparison of SV-CNP vs. Toronto method for FCCC images.

Figure 4.42 and Figure 4.43 demonstrate the results obtained using the LCE for the Harvard images. Like the CNP, the LCE correlates very well with the regions that are considered highly radiodense. But unlike the CNP, the algorithm does not use any variables such as α to fit the segmentation based on experimental results. The segmentation results are based on image statistics alone. The percentages obtained from

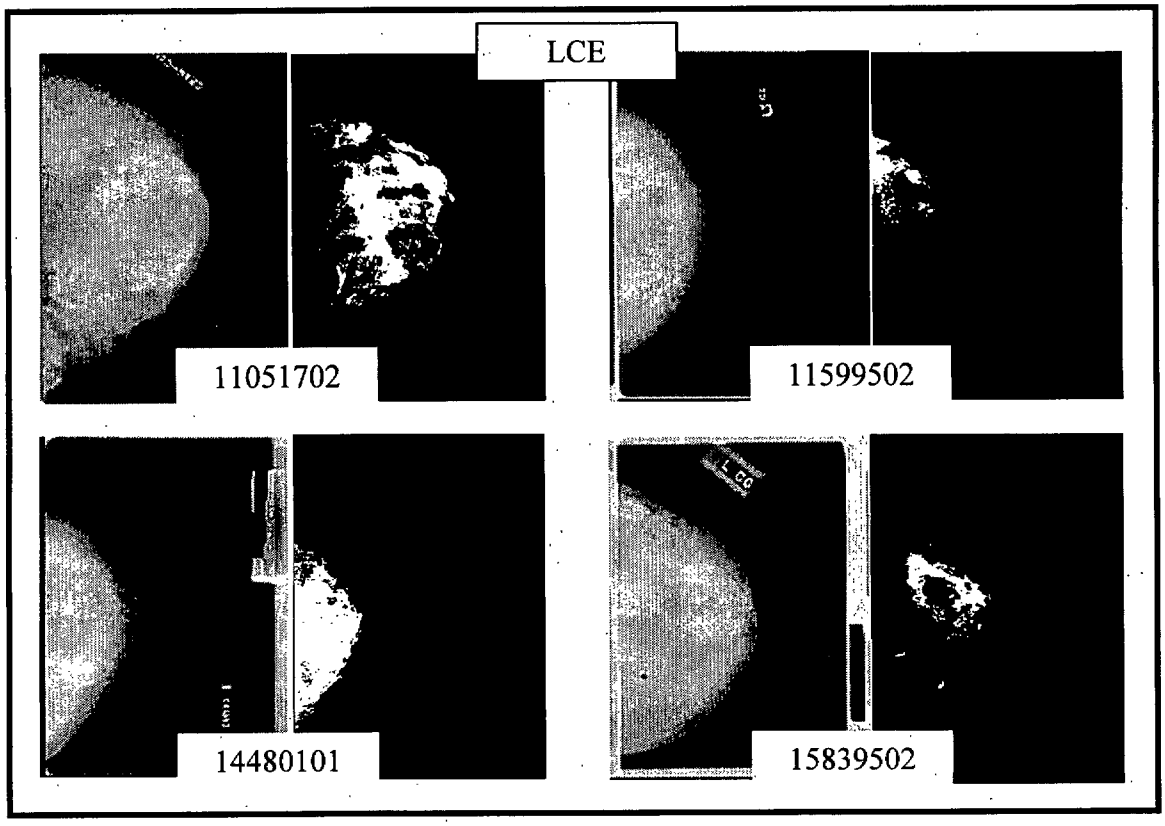


FIGURE 4.42 Results of first four Harvard images using LCE.

the LCE method for the Harvard images, as shown in Table 4.9 (Figure 4.44 Table 4.10) have a correlation of 0.868 and a MSE of 879.1. These values are lower than what both the CNP and SV-CNP achieved for the Harvard dataset.

The FCCC results demonstrate that the LCE is the better algorithm when it comes to cross-database performance. Unlike the CNP and SV-CNP, the LCE uses the same

parameters that it used for the Harvard dataset in the FCCC dataset. The percentages for the FCCC images using the LCE, as shown in Table 4.10 (Figure 4.45), has a correlation

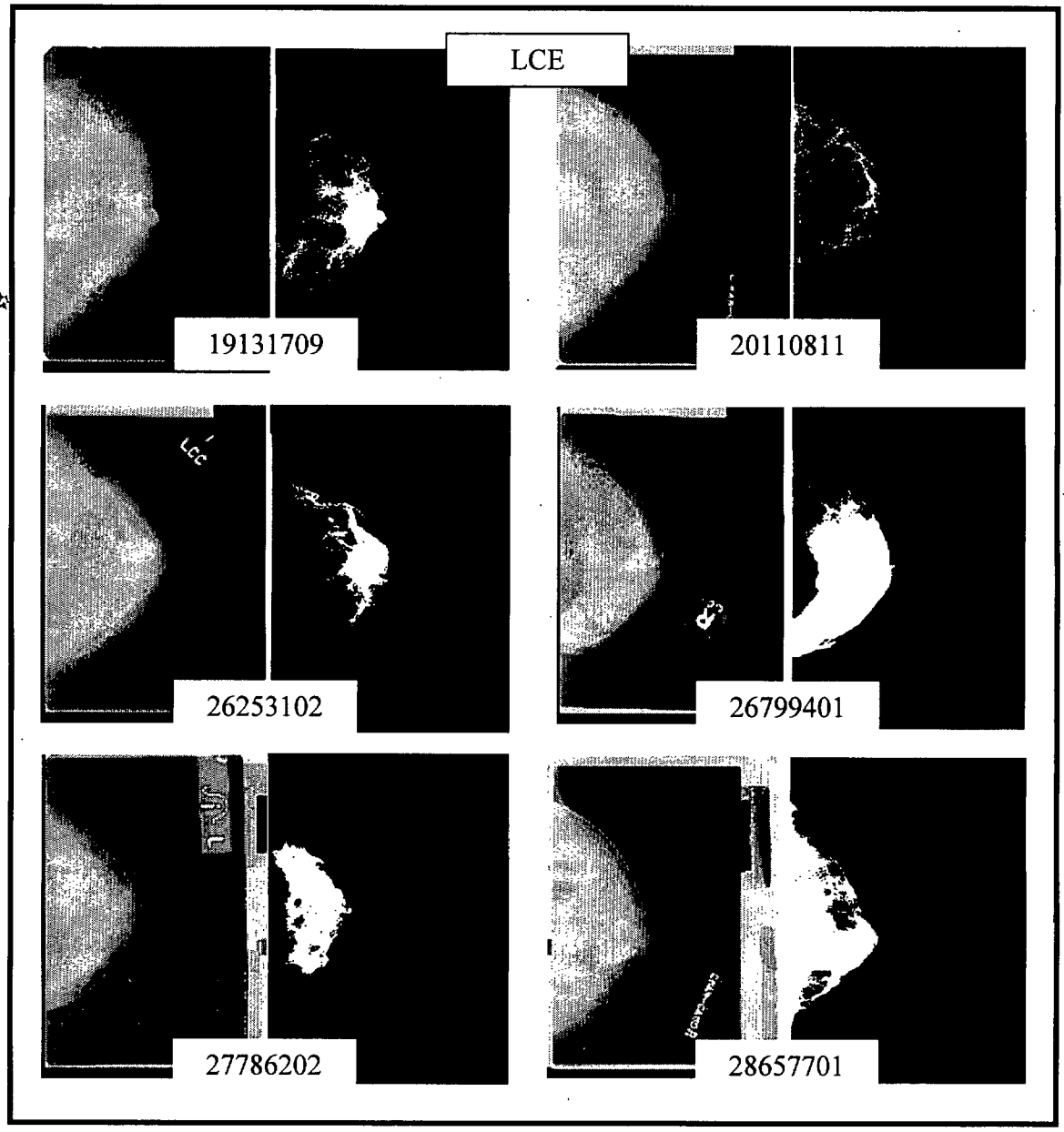


FIGURE 4.43 Results of last six Harvard images using LCE.

TABLE 4.9 Comparison of the LCE method to the Toronto method for Harvard database.

Image Number	Radiodensity Estimate of Toronto Method	Radiodensity Estimate of LCE	MSE of LCE
11051702	21.4	24.38	8.9
11599502	12.9	9.06	14.7
14480101	40.8	42.29	2.2
15839502	13.3	5.96	53.9
19131709	1.5	6.82	28.3
20110811	2.5	14.60	146.4
26253102	22.5	14.92	57.5
26799401	55.3	54.18	1.3
2778620	50.1	38.50	134.6
28657701	33.6	54.37	431.4

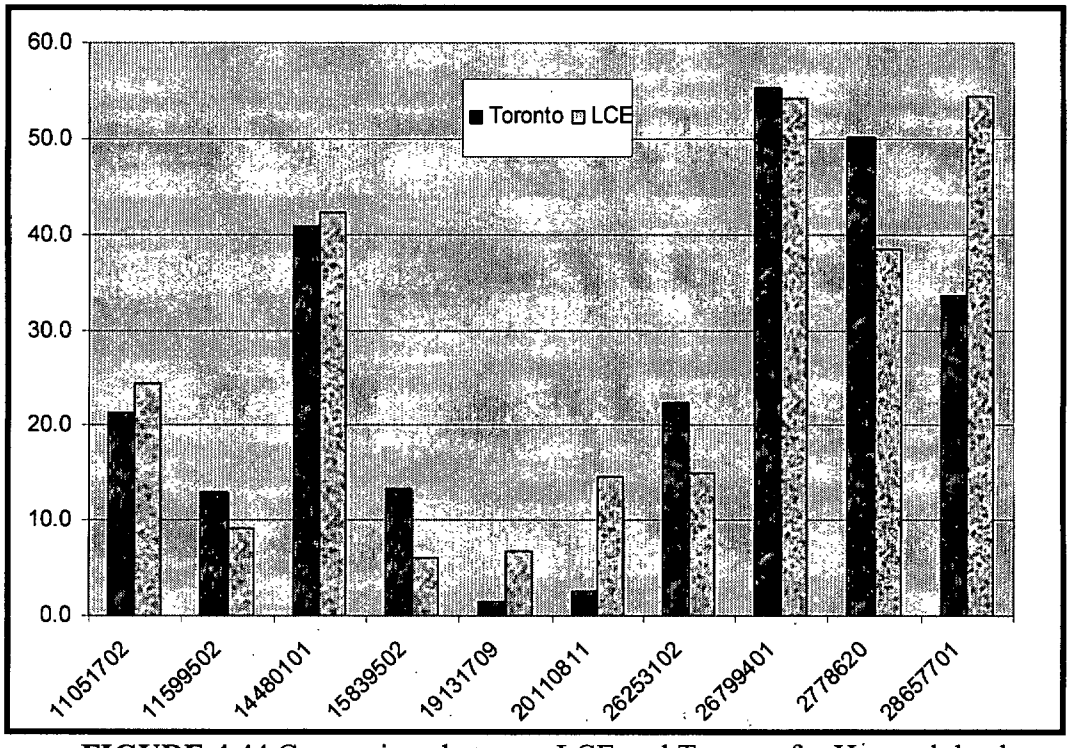


FIGURE 4.44 Comparison between LCE and Toronto for Harvard database.

TABLE 4.10 Correlation and MSE of the LCE to Toronto for all Harvard images.

Correlation of LCE and Toronto on Harvard images
0.868014326
MSE Harvard images
879.10

TABLE 4.11 The percentages obtained for FCCC database with the SV-CNP. The Correlation and MSE are compared with the Toronto method.

Image Number	Radiodensity Estimate of Toronto Method	Radiodensity Estimate of LCE	MSE of LCE
04706501	51.7	46.57	26.3
04706500	47.4	46.41	1.0
06917201	6.7	20.00	176.9
06917200	7.3	16.90	92.2
06943101	20.0	33.11	171.9
06943100	20.5	22.16	2.8
20392001	54.3	56.48	4.7
20392000	57.8	64.88	50.2
22272101	47.7	33.07	214.2
22272100	26.1	19.80	39.7
22574401	30.2	24.70	30.2
22574400	24.0	22.17	3.3
22645201	12.1	22.94	117.6
22645200	12.3	27.09	218.9
22733501	10.1	23.54	180.5
22733500	25.0	20.00	25.0
22765901	63.7	71.52	61.1
22765901	73.1	60.28	164.3
23152500	37.5	18.90	346.0
22808000	48.9	52.66	14.1
22818601	27.9	18.19	94.3
22818600	27.4	24.04	11.3
22819201	42.4	44.50	4.4
22819200	81.9	46.80	1231.8
23809901	36.6	25.56	121.8
23809900	18.2	22.40	17.7
27216101	17.2	18.23	1.1
23152501	37.5	38.72	1.5
03083801	25.0	25.50	0.3
03083800	9.1	20.14	122.0
04011001	79.9	68.27	135.3
04011000	53.0	66.98	195.5
06943101	20.0	33.11	171.9
06943100	20.5	22.16	2.8

Correlation of CNP and Toronto on FCCC images
0.848833784
MSE FCCC images
4052.28
Correlation of CNP and Toronto on FCCC images (without flagged)
0.881965271
MSE FCCC images (without flagged)
2811.69
Images Flagged by Dr. Celia Byrne

of 0.849 and a MSE of 4052.28. Unlike the other algorithms, the performance does not take a major hit when analyzing the second database. In fact, when the flagged images

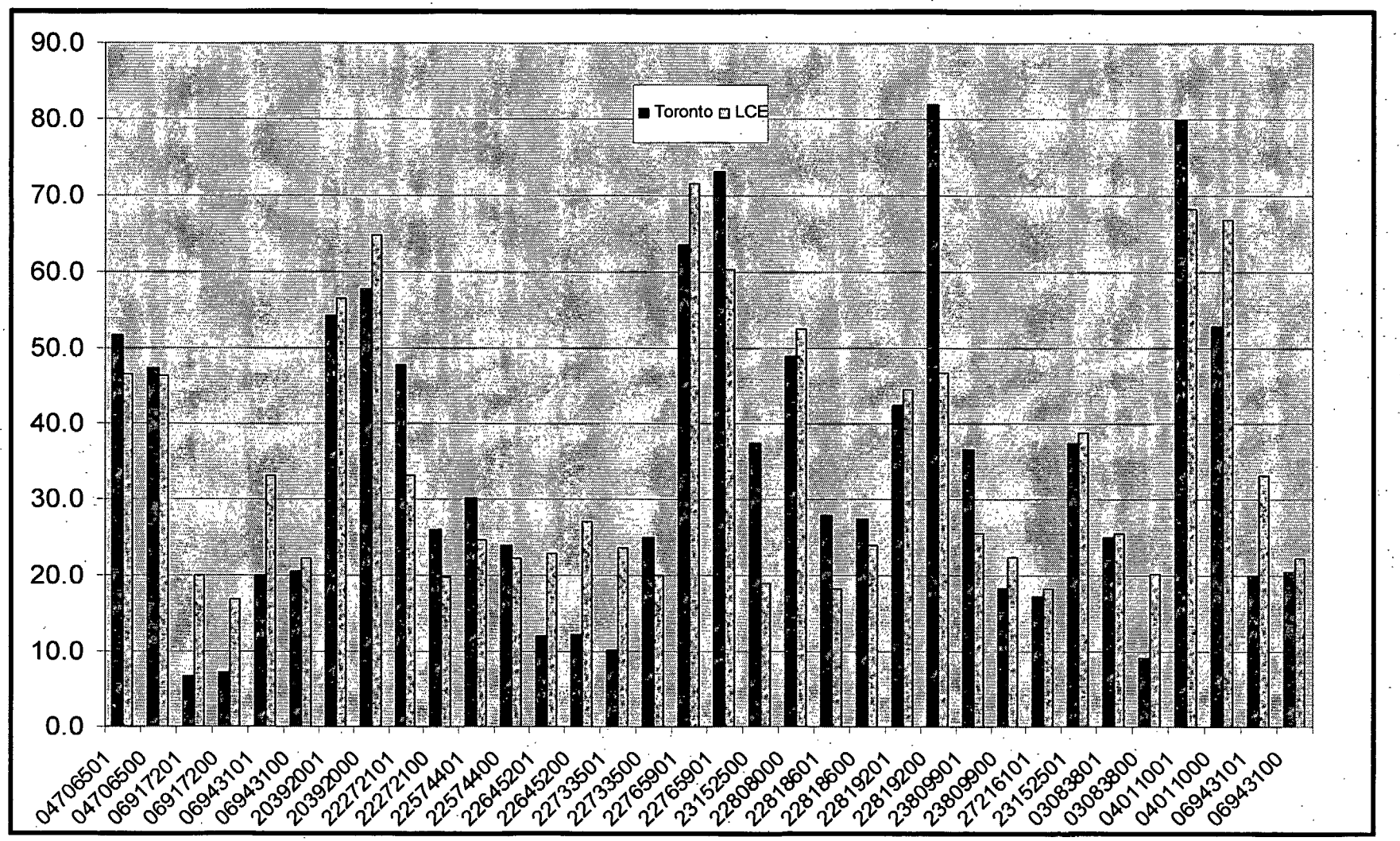


FIGURE 4.45 Comparison of LCE vs. Toronto method for FCCC images.

are taken out, the LCE performance actually increases compared to the results for the Harvard database with a correlation of 0.882 and a MSE of 2811.69.

Figure 4.46 shows a comparison of all methods using the Harvard Database. Based on these results, the SV-CNP has the best average percentage difference compared to the Toronto method with a value of 5.3%. The CNP has the second best average difference at 6.14% and the LCE has a value of 7.41%. These results were expected since both the CNP and the SV-CNP were fit to the Harvard dataset.

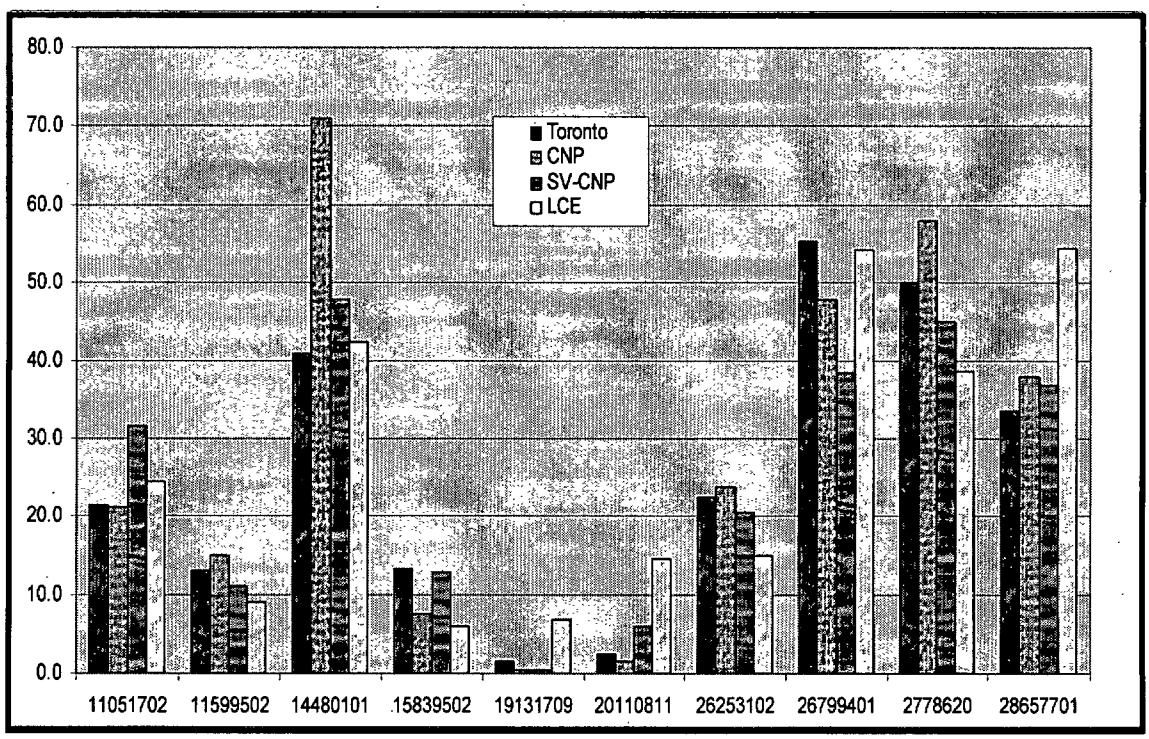


FIGURE 4.46 Comparison of all four methods on the Harvard database.

Figure 4.47 shows a comparison of all methods on the FCCC database. Based on these results, the LCE obtains the best average percentage difference with a value of 8.47%. The SV-CNP obtains a value of 15.2% and the CNP gets the worst score with a value of 21.72%.

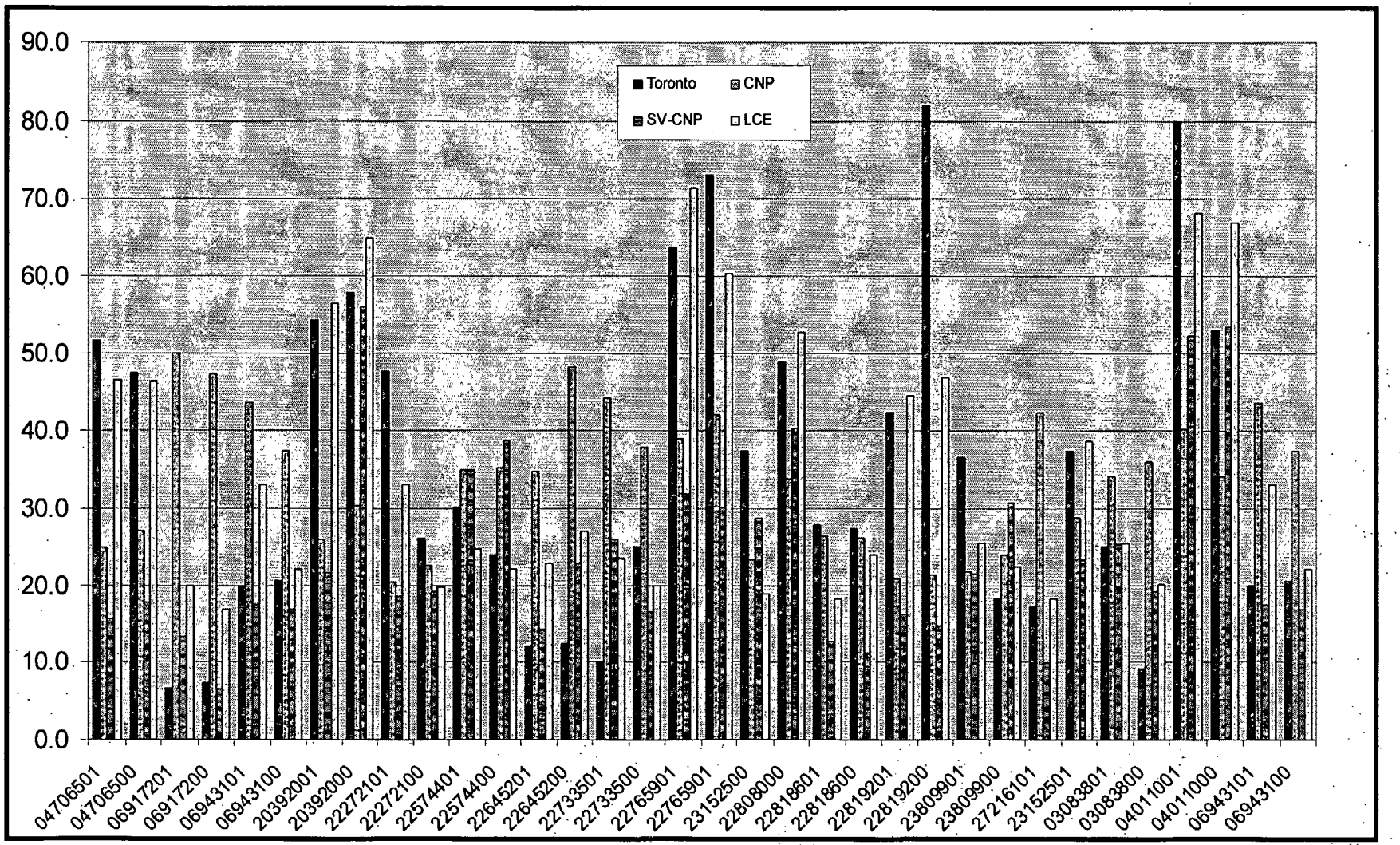


FIGURE 4.47 Comparison of all four methods on the FCCC database.

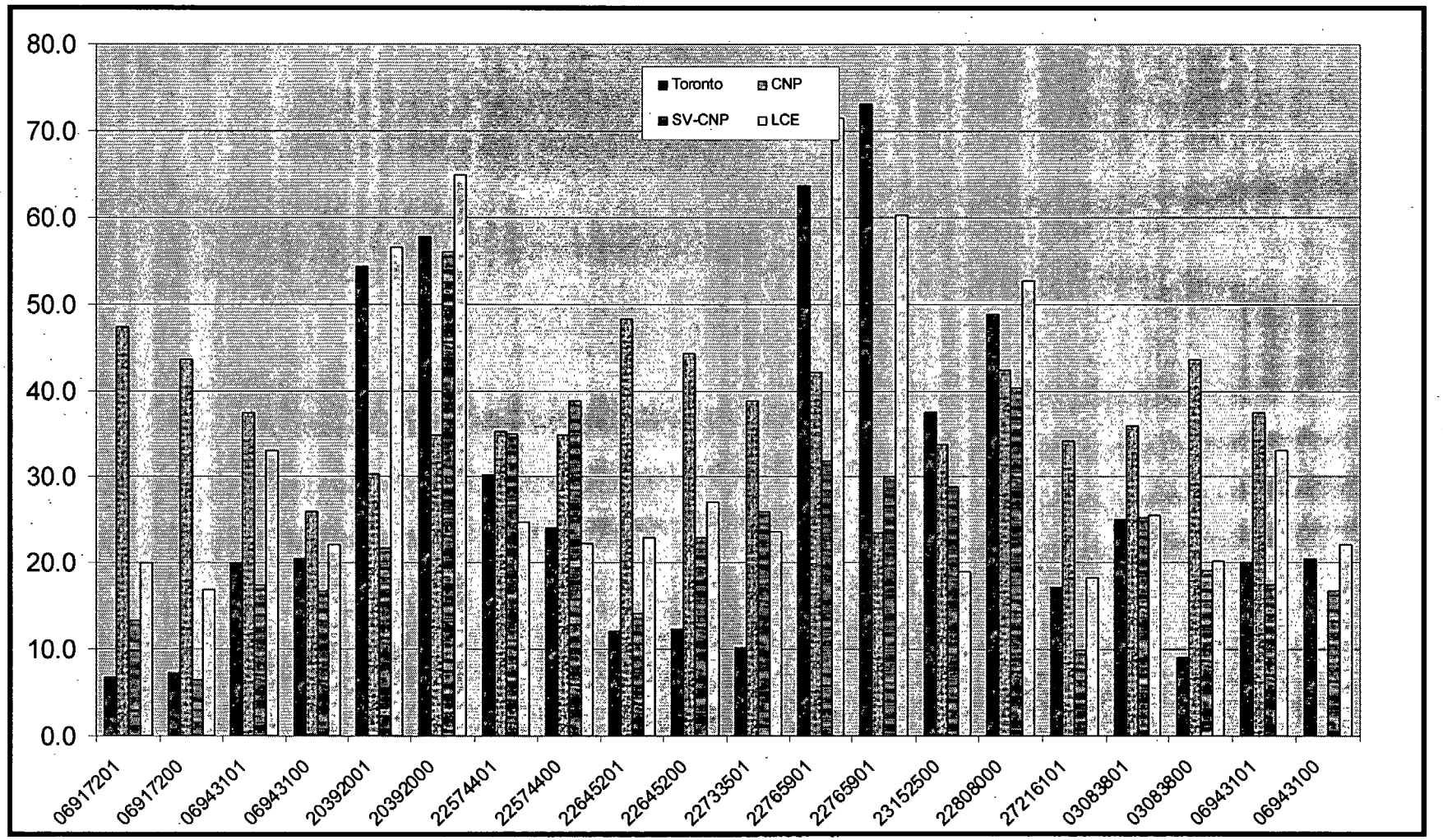


FIGURE 4.48 Comparison of all four methods on the FCCC database without flagged images.

Figure 4.48 shows a comparison of all four methods using only the unflagged FCCC images. The LCE has the best performance with an average percentage difference of 7.93%. The SV-CNP has a percentage difference of 8.79%, and the CNP has the worst performance with an average percentage difference of 17.89%.

Table 4.12 shows a comparison of various performance measures, such as the correlation, average percent difference, and the MSE of both datasets combined with and without the flagged images. The LCE performs the best in all categories. Figure 4.49 shows a bar comparison for all methods with the flagged images. Figure 4.50 shows a bar comparison without the flagged images.

TABLE 4.12 Comparison of Correlation, MSE, and average percentage difference to the Toronto method for all databases with and without flagged images.

	CNP	SV-CNP	LCE
Correlation compared to Toronto method (with flagged)	0.147	0.565	
Correlation compared to Toronto method (without flagged)	0.306	0.733	
MSE compared to Toronto method (with flagged)	22823.930	15622.682	
MSE compared to Toronto method (without flagged)	14381.540	5425.792	
Average % difference compared to Toronto method (with flagged)	18.186	12.924	
Average % difference compared to Toronto method (without flagged)	17.889	8.791	

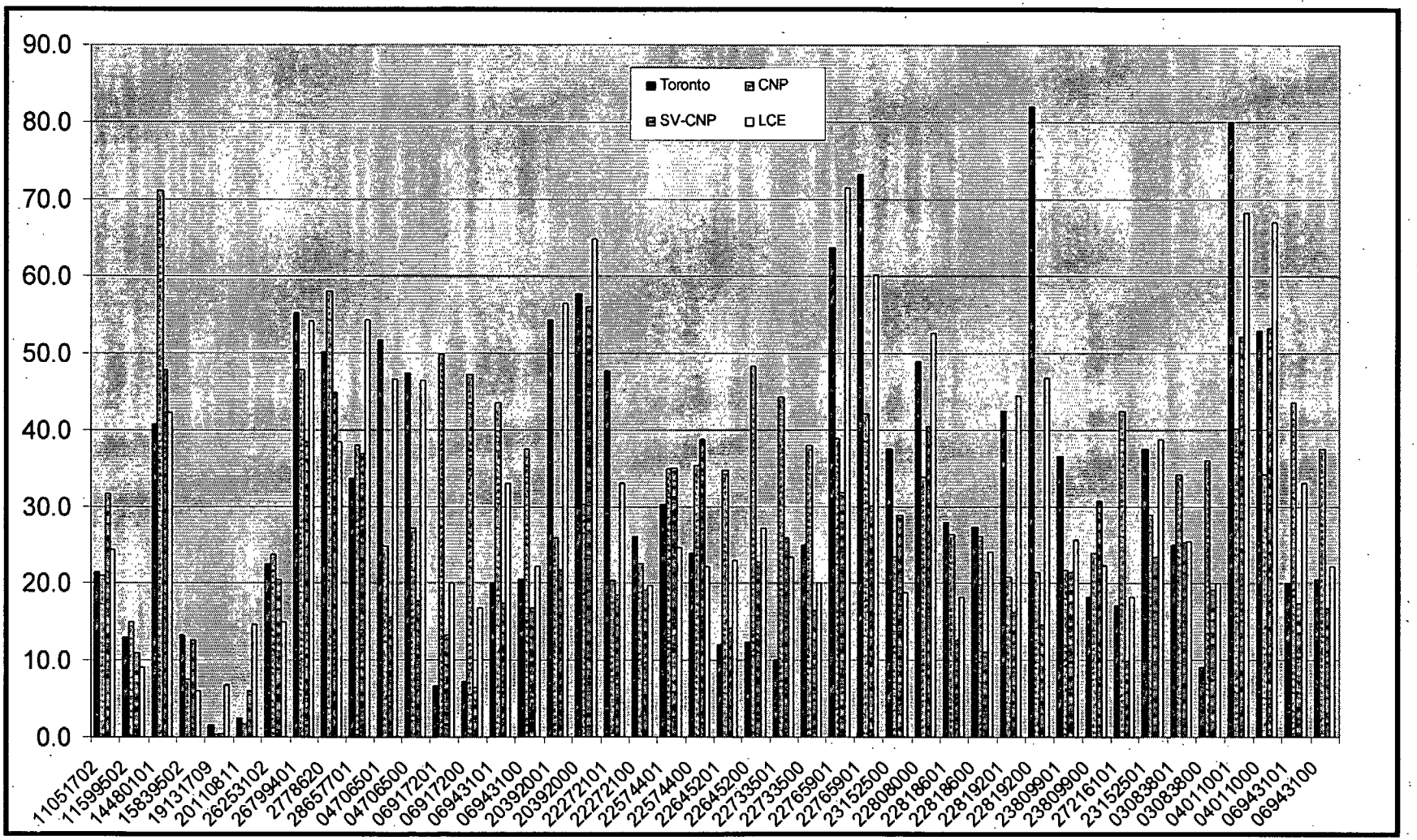


FIGURE 4.49 Comparison of all four methods on both databases without flagged images.

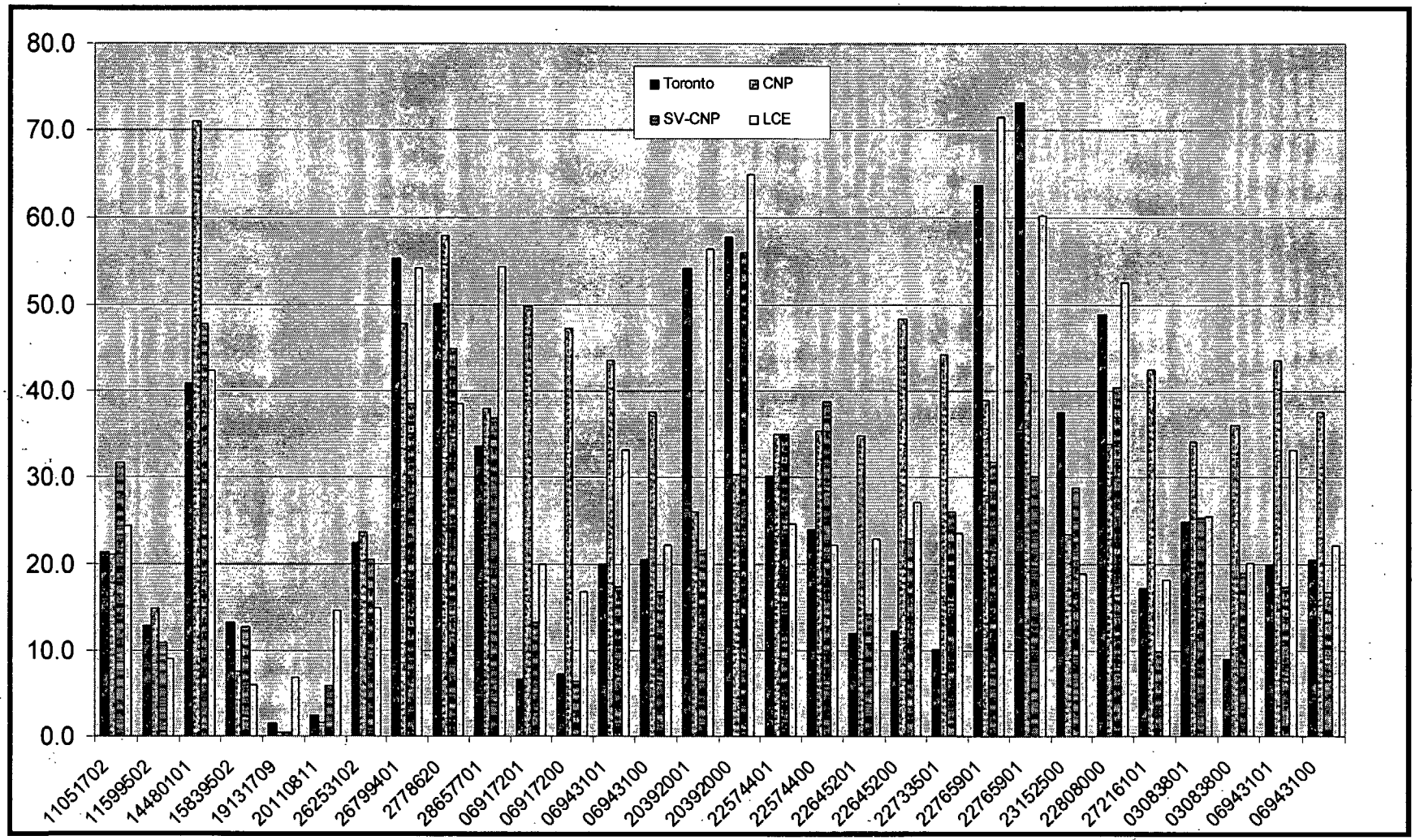


FIGURE 4.50 Comparison of all four methods on both databases without flagged images.

4.12 Summary

The results show that textures may not be suitable for the segmentation of radiodense tissue. For the FCCC dataset, some separation is achieved using texture segmentation but not enough to be viable for this process. The Harvard dataset produces no separation and indicates that any regional difference in the FCCC dataset may have been due to the high contrast between the radiodense and radiolucent regions. Perhaps texture based methods may work on other datasets, but for the two that have been studied in this thesis, textures produced no usable results. Because the main focus of this research was to create an algorithm that can span multiple databases, textures could not be used.

Fortunately, the results for the LCE algorithm look very promising. In all comparable performance statistics, the LCE performs better than the two previous algorithms. For both databases, LCE shows that it is correlated at values higher than 0.8 to the estimates given by Dr. Celia Byrne. Considering that the results are based on only image statistics rather than user input, these results are very promising. The LCE has the potential to provide cost effective and objective evaluations of mammograms at many locations and spanning different databases.

CHAPTER 5: CONCLUSIONS

Mammographic radiodensity is one of the strongest risk factors for developing breast cancer. Recent discoveries relating to the heritability of this characteristic indicate that there exists a genetic marker for breast density. This shows promise that investigations in isolating this marker may ultimately lead to understanding the causes of this debilitating disease. Previous attempts for automatically identifying and quantifying radiodense tissue in digitized mammograms have fallen short of the ideal. Many algorithms require significant heuristic parameters to be evaluated and set for predicting radiodensity. Many others have not demonstrated the efficacy of their techniques with a sufficient large and diverse patient database. This thesis has attempted to address both of these drawbacks in previous work.

5.1 Summary of Accomplishments

The major research accomplishments of this thesis work include –

1. The development of a comprehensive database of digitized mammograms. The Rowan database includes over 700 mammograms obtained from multiple age and ethnic groups and digitized using more than one type of X-ray digitizer. This database is available on the World Wide Web (password protected for security and confidentiality) for researchers from other institutions and has the potential to be a valuable resource in this area for future use.
2. The development of a completely automated technique for analyzing digitized mammograms to predict the location and quantity of radiodense tissue. The technique involves the following steps:

a. Preprocessing digitized images for noise removal using a combination of point and statistical processing techniques;
b. Breast tissue segmentation using morphological image enhancement and radial basis function artificial neural networks;
c. Compensation for X-ray artifacts of breast tissue compression that occurs as part of the mammography procedure using Gaussian interpolation models;
d. Segmentation of radiodense tissue using a novel local contrast estimation algorithm,

3. Comparison of the results obtained by exercising the automated radiodensity estimation algorithms developed in this thesis with the following previous techniques, using 44 digitized mammograms from 2 separate databases:
   a. The "Toronto" method, which is a previously established manual segmentation technique that has been used as the reference (validation) standard.
   b. The Constrained Neyman-Pearson algorithm, which is a semi-automated technique developed earlier at Rowan University.
   c. The Spatially-Varying Constrained Neyman-Pearson algorithm, which is an automated technique developed earlier at Rowan University.
4. Demonstration of the automated algorithm's ability to sift through entire databases of digitized mammograms to isolate radiodensity markers.

The automated digital image processing algorithms developed in this thesis have demonstrated the ability to rapidly sift through digitized mammogram databases for estimating radiodensity. Such algorithms are extremely useful in epidemiological studies when correlating other behavioral and genetic risk factors with mammographic radiodensity. Initial approaches towards arriving at a completely automated solution led to the investigation of image-texture methods for segmenting radiodense tissue. However, the estimation results did not appear promising. Eventually, a judicious combination of point-processing, statistical, neural and contrast enhancement techniques proved to be the optimal solution for this formidable constrained optimization problem.

The algorithm introduced in this thesis worked well with mammograms that had clearly defined radiodense and radiolucent regions. This was expected because of the fundamental theory of connectiveness that is the main concept behind the algorithm. The algorithm had difficulties in two areas:

1. With images that had radiolucent regions comparable to the X-ray region, the tissue segmentation algorithm had difficulties establishing the correct boundaries for the tissue. Many of these images that the algorithm had trouble with were the same images that the radiologist had flagged – indicating that the radiologist also had low confidence in her estimate.
2. There were several images that were similar in appearance and statistics. Because of the similar statistics of the mammograms, the algorithm estimated the radiodensities as similar quantities whereas the radiologist estimated the densities as significantly different.

Regardless of the problems, the algorithm is highly correlated with radiodensity estimates made by a radiologist – a correlation of 86% was obtained for 44 images analyzed using the automated method and the "Toronto" method. The average difference in estimates between the two methods was 8.2%.

5.2 Recommendations for Future Work

The algorithm presented in this thesis provides an option for quantifying radiodense tissue in digitized mammograms quickly, efficiently, and objectively. The results show that the performance is very promising, but there are still improvements that can be made on the algorithm and the procedure. Two main areas for improvement are: inter-observer variability in breast density estimation and obtaining accurate heuristics of breast compression.

The algorithm introduced in this thesis implemented many novel ideas to minimize the problems that can occur because of inter-database differences. Unfortunately, the algorithm in this thesis suffers from the same issue that all algorithms in this field of work do; the accuracy of the results is based on a radiologist's estimate. Regardless of how well trained the radiologist is, these estimates are still subjective. Therefore, two algorithms may be trained to different estimates of radiodensity, perform well based on these estimates, and yet still be drastically different from each other. Because there is no way to judge which of the estimates was more correct, the performance of the algorithm also becomes subjective by nature. Creating a set of canonical images with percentages that multiple radiologists agreed on would give this algorithm, as well as all future work, a better validation set to train to. This would be the

only way that a true approximation of the performance can be made.

Because the function or the amount of tissue compression that occurred was not known, the adjustment for it was kept very low in this thesis. It is still not known what role compression plays in changing the amount of radiodense tissue. Even if compression does play a role, there is still uncertainty whether or not a radiologist actually adjusts for it. The results indicate that there may not be much compression involved. An experimental analysis of the compression of breast tissue would give valuable information that would help in understanding the role of compression. As with the percentages, an inter-observer study on how each individual radiologist accounts for compression would help create a model that all future algorithms can follow.

REFERENCES

1. American Cancer Society. *Breast Cancer Facts & Figures*, 2003-2004 [Online]. http://www.cancer.org/downloads/STT/BrCaFF2003.pdf
2. L. Tabar, B. Vitak, H.H. Chen, M.F. Yen, S.W. Duffy, and R.A. Smith. "Beyond randomized controlled trials: organized mammographic screening substantially reduces breast carcinomar mortality. *Cancer 2001*, Vol. 91, No. 9, pp. 1724-1731, 2001.
3. Moore, Keith L. and Arthur F. Dalley. *Clinically Oriented Anatomy, 4th Edition.* Lippincott Williams and Wilkins: Philadelphia. C1999.
4. Ganong, William F. *Review of Medical Physiology, 20th Edition.* Lange Medical Books/McGraw-Hill: New York. C2001.
5. Ross, Michael H. et al. *Histology: A Text and Atlas, 4th Edition.* Lippincott Williams & Williams: Philadelphia. C2003.
6. J.N. Wolfe. "Breast patterns as an index of risk of developing breast cancer." *American Journal of Roentgenology*. Vol. 126, pp. 1130-1139, 1976.
7. J.N. Wolfe. "Risk of breast cancer development determined by mammographic parenchymal pattern." *Cancer*, Vol. 37, pp. 2486-2492.
8. Imaginis. 2002. Methods of Breast Biopsy. Siemens [Online]. http://www.imaginis.com/breasthealth.
9. R. A. Anderson. "Deaths: leading causes for 2000." *National Vital Statistics Reports*. Vol. 50, No.16, Centers for Disease Control, 2002 [Online].
10. C.A. Beam, E.F. Conant, E.A. Sickles. "Factors affecting radiologist inconsistency in screening mammography." *Academic Radiology*. Vol. 9, No. 5, pp. 531-540, 2002.
11. M.F. Salfity, G.H. Kaufmann, P.M. Granitto and H.A. Ceccatto. "A computer-aided diagnosis method for automated detection and classification of clustered microcalcifications in mammograms." *Proceedings of the Argentine Symposium on Healthcare Informatics*. Tandil, 41-47 (2000).
12. Byrne, C, et. al. "Mammographic features and breast cancer risk: effects with time, age, and menopause status." *Journal of the National Cancer Institute.* Vol. 87, *pp.*1622-1629, 1995.
13. Byrne. "Studying mammographic density: implications for understanding breast cancer." *Journal of the National Cancer Institute*. Vol. 89, pp. 531-533, 1997
14. Haiman, L. Bernstein, D. Van Den Berg, S. A. Ingles, M. Salane and G. Ursin. "Genetic determinants of mammographic density." *Breast Cancer. Res* 2002 4(3):R5.
15. Boyd, N.F., et al, "Heritability of mammographic density, a risk factor for breast cancer," *New England Journal of Medicine*, Volume 347(12), September 19, 2002, pp. 886-894.

16. A.F. Saftlas, M. Szklo, "Mammographic parenchymal patterns and breast cancer risk," *Epidemiol Rev*, Vol. 9, pp. 146-174, 1987.

17. A.M. Oza, N.F. Boyd, "Mammographic parenchymal patterns: a marker of breast cancer risk," *Epidemiol Rev*, Vol. 15, pp. 196-208, 1993.

18. E. Warner, G. Lockwood, D. Tritchler, N.F. Boyd, "The risk of breast cancer associated with mammographic parenchymal patterns: a meta-analysis of the published literature to examine the effect of method of classification," *Cancer Detect Prev*, Vol. 16, pp. 67-72, 1992.

19. N.F. Boyd, G.A. Lockwood, L.J. Martin, J.A. Knight, J.W. Byng, M.J. Yaffe, D.L. Tritchler, "Mammographic densities and breast cancer risk," *Breast Disease*, Vol. 10, pp. 113-126, 1998.

20. N.F. Boyd, J. Byng, R. Jong, E. Fishell, L. Little, A.B. Miller, G. Lockwood, D. Tritchler, M. Yaffe, "Quantitative classification of mammographic densities and breast cancer risks: results from the Canadian National Breast Screening Study," *Journal of the National Cancer Institute*, Vol. 87, pp. 670-675, 1995.

21. C.A. Bodian, "Benign breast diseases, carcinoma *in situ*, and breast cancer risk," *Epidemiol Rev*, Vol. 15, pp. 177-187, 1993.

22. N.F. Boyd, B. O'Sullivan, J.E. Campbell, E. Fishell, I. Simor, G. Cooke, *et al*, "Mammographic signs as risk factors for breast cancer," *British Journal of Cancer*, Vol. 45, pp. 185-193, 1982.

23. J.N. Wolfe, A.F. Saftlas, M. Salane, "Mammographic parenchymal patterns and quantitative evaluation of mammographic densities: a case-control study," *American Journal of Roentgenology*, Vol. 148, pp. 1087-1092, 1987.

24. J. Brisson, R. Verreault, A.S. Morrison, S. Tennina, F. Meyer, "Diet, mammographic features of breast tissue, and breast cancer risk," *American Journal of Epidemiology*, Vol. 130, pp. 14-24, 1989.

25. A.F. Saftlas, R.N. Hoover, L.A. Brinton, M. Szklo, D.R. Olson, M. Salane, *et al*, "Mammographic densities and risk of breast cancer," *Cancer*, Vol. 67, pp. 2833-2838, 1991.

26. S.A. Bartow, D.R. Pathak, F.A. Mettler, C.R. Key, and M.C. Pike, "Breast mammographic pattern: a concatenation of confounding and breast cancer risk factors," *American Journal of Epidemiology*, Vol. 142, pp. 813-819, 1995.

27. L. Tabar, B. Vitak, H.H. Chen, M.F. Yen, S.W. Duffy, and R.A. Smith, "Beyond randomized controlled trials: organized mammographic screening substantially reduces breast carcinomar mortality, *Cancer 2001*, Vol. 91, No. 9, pp. 1724-1731, 2001.

28. Imaginis. 2002. Methods of Breast Biopsy. Siemens [Online].

http://www.imaginis.com/breasthealth

29. American College of Radiology (ACR) Breast Imaging Reporting and Data System (BI-RADS) Third Edition. Reston, VA: American College of Radiology, 2003.

30. N.F. Boyd, H.M. Jensen, G. Cooke, H. Lee Han, "Relationship between mammographic and histological risk factors for breast cancer," *Journal of the National Cancer Institute*, Vol. 84, pp. 1170-1179, 1992.

31. M.J. Yaffe, N.F. Boyd, J.W. Byng, R.A. Jong, E. Fishell, G.A. Lockwood, L.E. Little, D. L. Tritchler, "Breast cancer risk and measured mammographic density," *European Journal of Cancer Prevention*, Vol. 7 (suppl 1), pp. S47-S55, 1998.

32. J.W. Byng, N.F. Boyd, E. Fishell, R.A. Jong, and M.J. Yaffe, "Quantitative analysis of mammographic densities," *Physics in Medicine and Biology*, Vol. 39, pp. 1629-1638, 1994.

33. P. Silgen, "Quantification of breast density using magnetic resonance imaging," *Master's thesis*, Department of Radiology, Minneapolis, MN: University of Minnesota, 1996.

34. J.C. Weinreb and G. Newstead, "MR imaging of the breast," *Radiology*, Vol. 196, pp. 593-610, 1995.

35. S.L. Lou and Y. Fan, "Automatic evaluation of breast density for full-field digital mammography," *Medical Imaging 2000: Image Processing*, Vol. 3979, pp. 1362-1369, 2000.

36. C. Zhou, H. Chan, N. Petrick, M.A. Helvie, M.M. Goodsitt, B. Sahiner, and L. Hadjiiski, "Computerized image analysis: Estimation of breast density on mammograms," *Medical Physics*, Vol. 28, pp. 1056-1069, June 2001.

37. C. Zhou, H. Chan, N. Petrick, B. Sahiner, M. Helvie, M. Roubidoux, L. Hadjiiski, M. Goodsitt, "Computerized image analysis: Estimation of breast density on mammograms," *Medical Imaging 2000: Image Processing*, Vol. 3979, pp. 1615-1624, 2000.

38. K. Bovis and S. Singh, "Classification of breast density in digital mammograms," *Submitted to Pattern Analysis and Neural Networks.*

39. P.K. Saha, J.K. Udupa, E.F. Conant, D.P. Chakraborty, and D. Sullivan, "Breast tissue density quantification via digitized mammograms," *IEEE Transaction on Medical Imaging*, Vol. 20, No. 8, 2001.

40. J.T. Neyheart, "Automated segmentation of radiodense tissue in digitized mammograms using a constrained Neyman-Pearson classifier," *Master's Thesis*, Department of Electrical and Computer Engineering, Glassboro, NJ: Rowan University, 2002.

41. R. E. Eckert, "Spatially varying threshold models for the automated segmentation of radiodense tissue in digitized mammograms," *Master's Thesis*, Department of Electrical and Computer Engineering, Glassboro, NJ: Rowan University, 2003.

42. M. Sonka, V. Hlavac, R. Boyle, *Image Processing, Analysis, and Machine Vision*, International Thomson Publishing, 1999.

43. P. Kruizinga, N. Petkov, S.E. Grigorescu, "Comparison of texture features based on Gabor filters," *Proceedings of the 10th International Conference on Image Analysis and Processing*, pp.142-147, Sept. 1999.

44. F. Argenti, L. Alparone, and G. Benelli, "Fast algorithms for texture analysis using co-occurrence matrices," *IEEE Proceedings, Part F: Radar and Signal Processing*, pp. 443-448, 1990.

45. K. I. Laws, "Texture energy measures," *DARPA Image Understanding Workshop, Los Angeles*, pp. 47-51, 1979.

46. R. C. Gonzalez, R. E. Woods, *Digital Image Processing*, Prentice Hall, New Jersey, 2001.

47. E. R. Dougherty. *An Introduction to Mathematical Morphology Processing*, SPIE Press, 1992.

48. R. O. Duda, P. E. Hart, D. G. Stork, *Pattern Classification*, John Wiley & Sons, New York, 2001.

49. Signal and Image Processing Institute, *The USC-SIPI Image Databases*

http://sipi.usc.edu/