



National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service

Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

University of Alberta

“Breast Cancer Diagnosis by Computer”

by

Shun Leung Ng (Anthony)

A thesis
submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree
of Master of Science

Department of Computing Science

Edmonton, Alberta
Spring 1990



National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service

Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

ISBN 0-315-60285-6

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Shun Leung Ng (Anthony)

TITLE OF THESIS: Breast Cancer Diagnosis by Computer

DEGREE: Master of Science

YEAR THIS DEGREE GRANTED: 1990

Permission is hereby granted to THE UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

(Signed) *Ng Shun Leung*

Permanent Address:
Flat C, 9th floor,
1148 Canton Road,
Mongkok, Kowloon,
Hong Kong

Date: *March 9, 1990*

**“Trust in the Lord with all your heart
and lean not on your own understanding;**

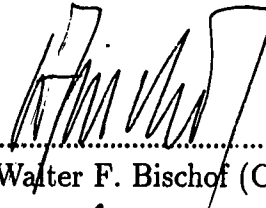
**in all your ways acknowledge him,
and he will make your paths straight.”**

Proverbs 3:5-6

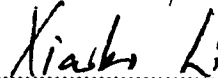
UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled **Breast Cancer Diagnosis by Computer** submitted by **Shun Leung Ng (Anthony)** in partial fulfillment of the requirements for the degree of Master of Science.



.....
Dr. Walter F. Bischof (Co-supervisor)



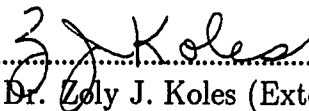
.....
Dr. Xiaobo Li (Co-supervisor)



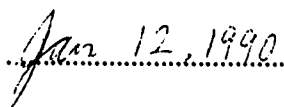
.....
Dr. Wayne A. Davis



.....
Dr. Barry Joe



.....
Dr. Zoly J. Koles (External)

Date: 

Dedicated to

my parents and my family

for their patience, love and encouragement

Abstract

The objective of this research is to construct an automated system for detecting stellate tumors, the most common type of breast cancer, in X-ray mammograms.

Methods are developed to isolate the breast region from the background and to locate and diagnose a stellate tumor. Specifically, a method which combines the simple thresholding and blob coloring techniques is developed for breast region isolation. Stellate tumors have a star-like shape consisting of two components: a surrounding radiating structure and a bright central tumor mass. Three approaches to radial structure recognition (RSR) are proposed and tested, namely edge-oriented RSR, field-oriented RSR and spine-oriented RSR. In addition, a modified directional selective median filter is proposed for removing the radiating structure of a stellate tumor so that template matching can be used to detect the central tumor mass of a stellate tumor. Finally, a classification process is employed to examine each detected suspicious area and to determine the tumor type.

A review on mammography and other tumor detection techniques in literature is included.

Acknowledgements

I would like to express my deepest thanks to my supervisors, Dr. Walter F. Bischof and Dr. Xiaobo Li, for their guidance, support and encouragement.

Thanks are also due to the members of my examining committee, Dr. Wayne A. Davis, Dr. Barry Joe and Dr. Zoly J. Koles, for their helpful comments and suggestions, Dr. Bill Castor, Director of the Radiology Department of the Cross Cancer Institute of the University of Alberta, for providing the mammograms.

Special thanks to my family and each individual in Paul Fellowship, Edmonton Chinese Alliance Church, for their love and encouragement throughout the process of writing this thesis.

Thanks should also go to Mr. Gordon Fong, Mr. Patrick Lau, Mr. Kris Ng, Ms. Shuk-Mei Lai and Ms. Alice Yu. Without their help, the completion of this thesis would not have been possible.

This research was financially supported by the Department of Computing Science, University of Alberta, and the National Sciences and Engineering Research Council of Canada.

Table of Contents

1	Introduction	1
1.1	Computer Automation	2
1.2	Objective of Thesis	3
1.3	Overview of Thesis	4
2	Breast Cancer and Mammography	5
2.1	Introduction to Breast Cancer	5
2.2	Management of Breast Cancer	6
2.2.1	Mass Screening with Mammography	7
2.3	Mammographic techniques	8
2.3.1	Mammographic Imaging system	9
2.3.2	Effectiveness of Mammography	10
2.4	Mammographic Classification of Breast Tumors	11
2.4.1	Stellate Tumors	13
2.4.2	Circumscribed Tumors	16
2.4.3	Calcifications	17
2.4.4	Thickened Skin Syndrome	20
3	Developments on Automated Breast Tumor Detection	21
3.1	Introduction	21
3.2	A Survey on Automated Breast Tumor Detection	23
3.2.1	Feasibility Demonstration	23
3.2.2	Tumor Classification	25
3.2.2.1	General Malignancy Differentiation	26
3.2.2.2	Specific Malignancy Differentiation	29
3.2.3	Detection of Abnormalities	31
3.2.3.1	Detection using Textural Statistics	31
3.2.3.2	Detection using Textural and Shape Measures	32

3.2.4	Detection of a Specific Kind of Tumor	34
3.3	Related Research	35
3.3.1	Mammographic Image Enhancement	36
3.3.2	Radiographic Tumor Detection	37
3.4	Conclusion	38
4	Stellate Tumor Detection	39
4.1	Introduction	39
4.2	Overall Strategy for Detection	41
4.2.1	Overview of the Detection System	45
4.3	Preprocessing	50
4.3.1	Breast Region Isolation	52
4.3.2	Breast Region Normalization	58
4.4	Radial Structure Recognition (RSR)	61
4.4.1	Gaussian Operator	62
4.4.2	Three Different Approaches to RSR	68
4.4.2.1	Edge-oriented RSR Approach	71
4.4.2.2	Field-oriented RSR Approach	75
4.4.2.3	Spine-oriented RSR Approach	83
4.4.3	Comparison among Different RSR Approaches	101
4.4.4	Spicule Removal	105
4.4.5	Spicularity Measurement	108
4.5	Central Mass Detection	113
4.6	Classification	119
4.6.1	False Alarm Removal	119
4.6.2	Tumor Type Differentiation	121
4.7	Experimental Results and Discussion	122
5	Concluding Remarks	132
	Bibliography	135

List of Figures

2.1	Basic patterns of stellate tumors	15
4.1	Data Flow of the Detection System	46
4.2	Data Flow between RSR and CMD processes	48
4.3	Preprocessing Process	51
4.4	Breast Region Isolation	57
4.5	An Ideal Radiating Structure	61
4.6	Gaussian Distribution	63
4.7	Response in Edge Detection	66
4.8	Infinite Line Accumulation	70
4.9	ILA in Edge-oriented RSR	72
4.10	ILA in Field-oriented RSR	78
4.11	Effect of Angle Doubling	80
4.12	Ideal Radiating Structure for Spine-oriented Approach	83
4.13	Fine Structure of a Spicule	84
4.14	Examples of Acute Spicules	87
4.15	Examples of Obtuse Spicules	87
4.16	ILA in Spine-oriented RSR	89
4.17	Detection of Spinal Pixel	92
4.18	Spinal Direction Measurement	94
4.19	Consistency Test for a Potential Spinal Pixel	95
4.20	Measurement of New Spinal Direction	98
4.21	Spicule Removal	106
4.22	Spicule-smearing Filter	106
4.23	Spinal Pixel Accumulation	111
4.24	A Tumor-like Template	114

List of Plates

2.1	Image s1: a typical stellate tumor	18
2.2	Image s1: with the central mass of the stellate tumor circled .	18
2.3	Image c1: a typical circumscribed tumor	19
2.4	Image c1: with the circumscribed tumor circled	19
4.1	Effect of Gaussian Operator	65
4.2	Result of Edge-oriented RSR	76
4.3	Result of Field-oriented RSR	82
4.4	Result of Spine-oriented RSR	100
4.5	Image s1 after the preprocessing stage	123
4.6	Image c1 after the preprocessing stage	123
4.7	Image s1 overlayed with red-colored spinal pixels	124
4.8	Image c1 overlayed with red-colored spinal pixels	124
4.9	Image s1 with the suspicious areas circled	126
4.10	Image c1 with the suspicious areas circled	126
4.11	Image s1 overlayed with the red-colored spinal pixels of the suspicious areas	127
4.12	Image c1 overlayed with the red-colored spinal pixels of the suspicious areas	127
4.13	Image s1 with the detected suspicious stellate tumor marked .	128
4.14	Image c1 with the detected suspicious circumscribed tumor marked	128
4.15	Image s4 having a partially obscured suspicious stellate tumor	130
4.16	Image s9 having a suspicious stellate tumor of unusual structure	130

Chapter 1

Introduction

Breast cancer is one of the leading cancers. It develops in 9% of women born in the United States [GBKS87] [Les84] and its incidence is still increasing [TD85a] [CBRE88, page 12]. When a malignant and invasive tumor develops, the abnormal cells which form the tumor tend to infiltrate surrounding tissues, spread to other body parts and eventually lead to death.

Even though intensive research efforts have been undertaken in understanding the cause and the nature of the disease, its cause is still unclear [CBRE88, page 12] [TD87]. Without this knowledge, prevention of its occurrence is not possible [TD87]. Currently, the treatment for breast cancer is effective only when it is detected at an early stage. Once it has spread out, treatment can hardly be complete and has many negative side-effects [CBRE88, page 85]. Therefore, it is very important to detect the existence

of breast cancers at an early stage.

By now, mammographic screening is the only well-established method to achieve early detection [DBG86] [CBRE88, page 57]. It should be done regularly for women in a middle age or above. Implementing an exhaustive mammographic screening program is very difficult as it requires thousands of radiologists to interpret a large volume of breast images taken. Shortage of qualified radiologists and great financial costs are two important problems of such a mass screening program.

1.1 Computer Automation

In order to solve the problem of having to interpret a large number of breast images, Winsberg et al. [WEM⁺67] proposed to use a computer to automate the reading and interpretation of breast images. Some primitive experiments were conducted and it demonstrated the feasibility of computer automation.

Ackerman and Gose [AG72] and Smith et al. [SWGS77] showed that computers could be as good as experienced radiologists in differentiating malignant from benign tumors when the suspicious tumor areas were marked. However, this was true only when the tumor area was the only part seen by both radiologists and the computers. By no means could such a system be considered an automated tumor detection system as human assistance was needed to mark the suspicious areas first.

Other groups of researchers [KOS75] [HSAA79] [SSA⁺80] employed textural and shape measures to locate suspicious tumor areas in breast images. These techniques were successful in detecting tumors but at the same time many false alarms were generated. A false alarm is an area marked by the computer as suspicious but not considered suspicious by radiologists.

Most recent research undertaken by Lai [Lai88] took another approach to the tumor location problem. Instead of attempting to locate suspicious areas associated with any kind of breast tumor, a system was developed to detect a particular type of breast tumor called circumscribed tumors. Circumscribed tumors appear in breast images as bright circular objects. When compared with the general suspicious area detection, specific tumor detection allows more precise detection criteria and more effective detection algorithms.

1.2 Objective of Thesis

This thesis aims at developing a fully automated breast tumor detection system to detect another type of breast tumor, stellate tumors. The stellate tumor is a very important type of breast tumor as it bears the appearance of most breast cancers. The shape of a stellate tumor looks like a star. It contains two basic components: the *surrounding radiating structure* and the *central tumor mass*. Compared to the detection of a circumscribed tumor, the detection of a stellate tumor is considerably more difficult as stellate

tumors have a more complicated structure and a more diffuse boundary. The identification of a stellate tumor at its early stage often causes problem even to radiologists.

1.3 Overview of Thesis

The next chapter discusses breast cancer and how mammography can be used to effectively detect early breast cancer. A description about the visual appearance of all types of breast tumors is also included.

In chapter 3, a literature review on existing breast tumor detection techniques will be given. It provides a base for further discussion on how to design an effective method to detect stellate tumors.

Chapter 4 is considered to be the core of this thesis. It will begin with an overview of the entire detection strategy, where different alternatives for detection are explored. A new automated detection system is then proposed.

In the proposed system, there are four major processes. They are the *pre-processing process*, the *radial structure recognition (RSR) process*, the *central mass detection (CMD) process* and the *classification process*. Each of them will then be discussed in detail. Finally, experimental results from applying the proposed system on twenty-seven real mammograms (X-ray films of the imaged breasts) will be presented at the end of chapter 4.

As a conclusion, chapter 5 will summarize the contribution of this research and give suggestions for future research.

Chapter 2

Breast Cancer and Mammography

2.1 Introduction to Breast Cancer

Breast cancers are malignant and invasive tumors, made up of abnormal cells, which tend to infiltrate the surrounding tissues and eventually lead to death. Breast cancers are not only the most common cancers in women [Bau88] [CBRE88, page 12] [WB87] [TD85a], but also the leading lethal cancer in women [GBKS87]. Further, their incidence is still increasing [TD85a] [CBRE88, page 12]. According to recent statistics, breast cancers develop in one out of every ten to eleven women born in the United States [GBKS87] [Les84]. Since the size of the affected population is so large, intensive research has been undertaken in an effort to manage the disease.

2.2 Management of Breast Cancer

The best way to manage any disease is to find out the cause of the disease and prevent its occurrence. This applies equally well to breast cancer. Many studies investigated the nature of breast cancer and managed to identify a number of risk factors such as gender, age and heredity [RS88]. However, the cause of the breast cancer is largely obscure [CBRE88, page 12] [TD87]. Without this knowledge, prevention of breast cancer is practically impossible [TD87].

The second medical arm to combat the disease involves diagnosis and therapy. Many new improvements in therapeutic methods have made the treatments for early localized breast cancers effective [Les84] [TD87] [TD85a]. However, treatments for advanced breast cancers, which have metastatically spread to other parts of the body, are far from adequate [CR88] [MB88] and have many unpleasant side-effects [CBRE88, page 85]. Therefore, it is very important to diagnose breast cancer at a very early stage. In general, the earlier the breast cancer is detected, the more effectively the disease is cured.

Breast cancers must be detected first before any further diagnosis can be attempted. Currently breast cancers are detected by breast self-examination [CBRE88, pages 61–62], clinical examination [Chi88] and various imaging techniques, such as mammography, transillumination and magnetic resonance imaging [CBRE88, page 57]. The limitation of the first two methods

is that they can only detect symptomatic or palpable breast cancer. Asymptomatic and impalpable early breast cancer can only be detected by some imaging techniques.

2.2.1 Mass Screening with Mammography

At present, mammography is the only proven imaging technique which can detect breast cancer at its earliest preclinical stage [TD87] [Les84]. It is also the only well-established method for population screening [DBG86] [CBRE88, page 57].

According to the guidelines proposed by the American College of Radiology [Chi88] [GBKS87], mammography should be performed every 1–2 years on all women aged 40–49. The screening frequency depends upon some risk factors such as history of first-degree relatives having breast cancer. Annual mammography is recommended for all women over age 50.

Such an exhaustive and frequent screening program may require thousands of radiologists to interpret the large volume of breast images produced. An immediate problem will be the shortage of well-trained radiologists. It may take many years to train and recruit enough radiologists. Besides, it is difficult to maintain the interest of radiologists when interpreting large numbers of images with only a small number showing abnormalities [HSAA79]. Furthermore, such a mass screening program puts a great financial burden

on society.

One suggestion for solving the problems effectively is to employ advanced computer image processing technology to screen all breast images and filter out those containing suspicious cancerous areas. Radiologists can then focus on diagnosing small volumes of suspicious breast images in detail.

In order to build such an automated computer system to filter out those suspicious breast images, some well-defined programmable detection criteria for breast cancers should be obtained first. These criteria should correspond to those used by radiologists in detecting the breast cancers.

To provide a sufficient background for further discussion on automated detection of breast cancer in the coming chapters, a detailed description of mammographic techniques and mammographic features of breast cancers are given below.

2.3 Mammographic techniques

Mammography is a technique for imaging the internal structure of the breast by passing X-rays through the breast and capturing the image produced on film. The subsequently developed film is called a *mammogram*. By carefully analyzing all the fine details of the breast images in the mammograms, radiologists determine if there is any suspicious area of breast tumor¹.

¹If the breast tumor is of malignant nature, it is called breast cancer.

The visualization of deep structures of the breast on a mammogram is due to the small variations in absorption of radiation by the breast tissues. The difference in absorption power between normal and abnormal tissues will determine the contrast of the tumor on the mammogram [Ega70]. Since the breast tissues are soft, the overall contrast of the breast image is very limited.

On the other hand, radiologists require mammogram images of very high quality (in terms of contrast and resolution²) so that they can evaluate both tumors and their surrounding structures [Ega70]. In order to satisfy such a high quality requirement, most of the mammographic imaging systems are specially designed and built [CBRE88, page 31] [CBRE88, page 26].

2.3.1 Mammographic Imaging system

A mammographic system is very similar to an ordinary radiographic system as they both are used to image the internal structures of the body. However, a much higher quality is expected in a mammographic system so that it can image the fine structures of the soft breast tissues. The two basic components in a mammographic system are the *X-ray unit* and the *recording unit*.

In an X-ray unit, a fine focal spot of the X-ray tube is very important as it is one of the major factors affecting the image sharpness. Besides, the X-ray unit should be able to control the radiation dose accurately and generate an

²It is important because calcification, one kind of breast tumor, is very small in size.

X-ray spectrum of different energy levels. Since larger breasts absorb most low energy photons, a higher energy spectrum is required [CBRE88, page 27].

In the recording unit, a screen-film receptor and a xeroradiographic receptor are the two most commonly used image receptors. The advantage of a screen-film receptor is that it gives higher contrast images facilitating the detection of poorly defined masses [GB83], whereas xeroradiographic receptor inherits the edge enhancement phenomenon which assists in the visualization of microcalcifications³ [CBRE88, page 30] [GB83]. However, the overall performance using the two different imaging techniques is very close. Clinical investigation has failed to show any significant difference in their ability to detect breast cancers [GBKS87] [GB83].

2.3.2 Effectiveness of Mammography

At present, mammography is the only proven method that can detect breast tumors at an early stage [CBRE88, page 57] [TD87] [Kop87] [GBKS87] [DBG86]. Some early breast cancers that are not yet symptomatic nor palpable cannot be detected by self-examination and clinical check-up [GBC87]. Mammography, however, can effectively expose their existence. This is very important because early therapeutic intervention can control the cancer from

³Microcalcification is a kind of breast tumor whose size is extremely small.

developing into an advanced incurable stage. Besides detecting early breast tumors, mammography is also a very useful diagnostic tool as it provides more information about the characteristics of the tumors.

On the other hand, there is one limitation for a mammographic technique. It cannot definitively differentiate a malignant breast tumor (i.e. breast cancer) from a benign one [Hom87] [Kop87] [TD85b, page 20] [GBC87], even though it can effectively detect the existence of either tumor. Fortunately, this limitation does not severely degrade the power of mammography. Once suspicious tumors are detected using mammography, other imaging techniques, such as ultrasonography⁴, can be used to effectively differentiate the two [Kop87] [CBRE88, pages 53–57] [GBKS87].

2.4 Mammographic Classification of Breast Tumors

The classification of breast tumors has been a controversial problem for many years. The problem arises because nature and cause of the disease are still unclear [CBRE88, page 12] [TD87] [GBB88]. Paget's disease of the nipple, for example, is an extensively studied and well-documented breast disease. Yet there is no agreement on whether the abnormal cells develop locally within

⁴In ultrasonography, the suspicious tumor is imaged by recording the echoes of the ultrasonic waves directed into it.

the epidermis of the nipple or they originally develop in mammary ducts and subsequently move into the epidermis [Gal83]. Accurate classification, however, is very important as it is crucial in determining the therapeutic strategy [Gal83].

The *pathological approach*, which classifies breast tumors according to the site of origin and the existence of invasion, is currently the most common classification method [Bau88] [CBRE88, page 13]. *Clinical classification* which depends on observable symptoms is also used, especially when causes of the tumors are unclear and they represent a benign process. As a result, pathologists and clinicians often disagree on classification, and multiple labels are sometimes assigned to describe the same breast tumors [TD85b, page 89] [GBB88].

Radiologists emphasize the outward appearance of the breast tumors projected on mammograms. Instead of beginning with analyzing and classifying breast tumors, radiologists first place them into broader classification groups based on their apparent patterns in the mammograms.

Radiologists subsequently examine the mammograms to determine the details of the structures and other abnormalities due to the tumors [TD85b, page 16]. Based on these observations, pathological classifications are assigned. If ambiguity remains, radiologists can recommend a biopsy of the tumors in order to obtain some definitive diagnoses.

Breast tumors are generally divided into four major groups: stellate

tumors, circumscribed tumors, calcifications, and thickened skin syndrome [TD85b, page 16]. The labels reflect the basic mammographic appearance of the tumors. Stellate tumors are star-like structures with ill-defined border. Locating them in the early stage is especially difficult because they are hardly perceptible. Circumscribed tumors have circular or oval central masses. Calcifications are composed of small radiopaque spots. In the “thickened skin syndrome” the skin is thickened over much or all of the breast [TD85b, page 212].

In the following sections, descriptions of the mammographic features of each of the above classes and their corresponding implications (e.g. whether the tumor is benign or malignant) are given. Because the radiating structures of stellate tumors are considered to be the classic primary sign of malignancy [Chi88], and most breast cancers are classified as stellate tumors [TD85b, page 88], early detection of stellate tumors is extremely important. The present thesis is concerned with the development of an automated computer system for the detection of stellate tumors. To provide sufficient background for such a thesis, greater details of stellate tumors will be given.

2.4.1 Stellate Tumors

Most breast cancers have the appearance of a stellate tumor, but not all stellate tumors are malignant. Some stellate tumors, such as sclerosing duct

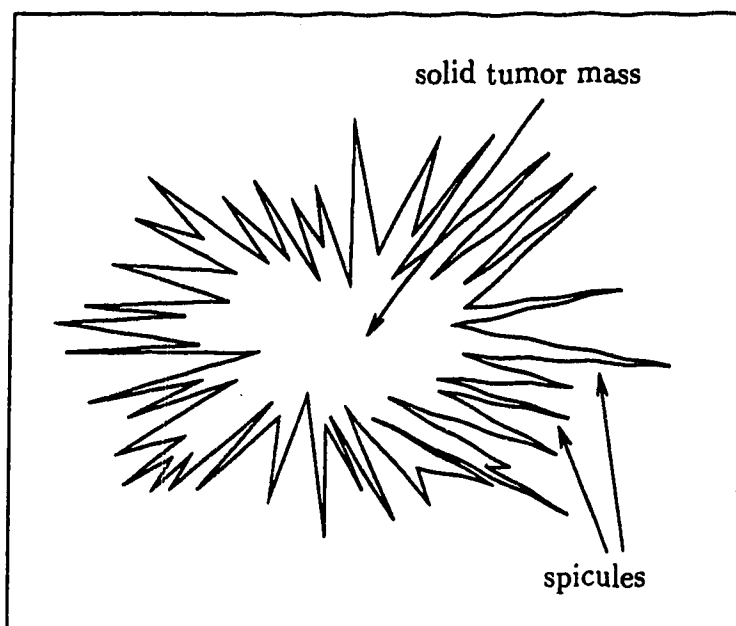
hyperplasia, are benign in nature and can be differentiated mammographically with very high accuracy.

In general, the shape of stellate tumors looks like a star. The radiating lines in the star are referred to as spicules. This is why stellate tumors are also called star-shaped tumors or spiculated tumors [Mar82b] [TD85a]. Basically, stellate tumors can be decomposed into two components: the *central tumor mass* and the *radiating structure*. The exact forms of these two components are related to the nature (malignant or benign) of the tumor. The malignant one tends to have a higher radiographic contrast in the mammogram. Figure 2.1a (2.1b) depicts the basic pattern of the malignant (benign) stellate tumors.

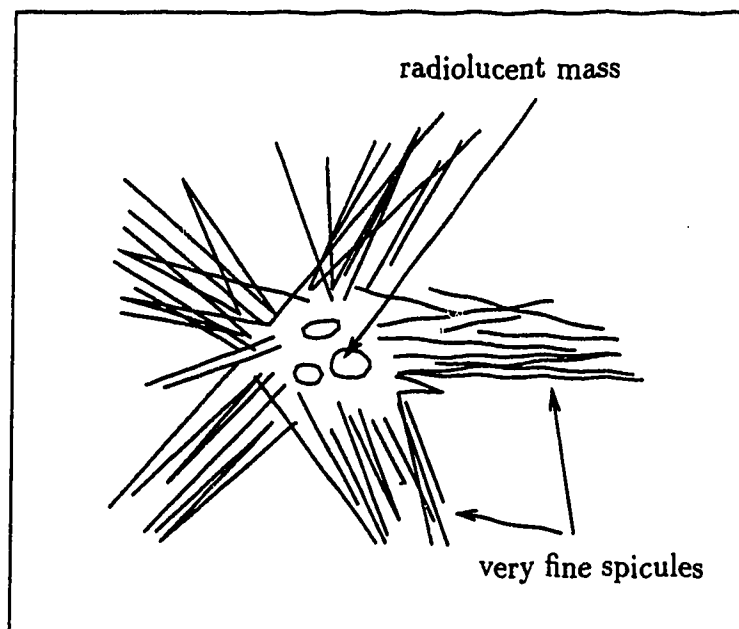
The detailed mammographic features of stellate cancers are listed below in point form so that they are readily translated to some programmable criteria for computer detection in later discussion.

The characteristics of the central tumor mass of a stellate cancer [TD85b, pages 88–89] [Chi88] are as follows:

- It is a solid distinct radiopaque mass.
- It is located at the center of the radiating structure.
- It may be very difficult to perceive when small (i.e. less than 10mm) but it always exists.



(a) malignant



(b) benign

Figure 2.1: Basic patterns of stellate tumors

The characteristics of the radiating structure of a stellate cancer [TD85b, pages 88–89] [Chi88] are as follows:

- It consists of some spicules which are sharp, dense, fine lines of variable lengths.
- The average length of the spicules increases with tumor size.
- Spicules are not usually bunched together.
- Spicules come out from the tumor surface.
- Spicules radiate outward in all directions.

Benign and malignant stellate tumors are similar in appearance. Instead of a solid distinct central mass, the benign stellate contains some circular or oval radiolucent areas at the center of the radiating structure. The spicules of such tumors are many but very fine and weak. They can be described as a sheaf of wheat [TD85b, page 88].

Plate 2.1 shows a real mammogram image which contains a typical malignant stellate tumor. Plate 2.2 depicts the same mammogram image except the central mass of the tumor is circled.

2.4.2 Circumscribed Tumors

Compared with a stellate tumor, the structure of a circumscribed tumor is relatively simple. It presents a circular or oval tumor mass, and the per-

ception of this kind of tumor is usually not a problem. In detailed analysis, radiologists study four different aspects of the tumors — contour, density, orientation and size [TD85b, pages 18–19]. The first two aspects are of primary importance in diagnosis. A sharply outlined contour indicates that it is benign, whereas a fuzzy periphery usually represents malignancy. A highly radiopaque tumor mass is a significant sign of circumscribed cancer. For a benign one, the tumor mass is either radiolucent or only partially radiopaque. The last two aspects are used only to confirm the diagnosis that has been made.

Plates 2.3 and 2.4 show the same mammogram image except the circumscribed tumor in Plate 2.4 is highlighted by a circle.

2.4.3 Calcifications

Calcifications can be described basically as collections of some small highly radiopaque spots. The form, size, density, number and distribution of different subclasses of calcifications vary a lot. The first three aspects are considered to be the primary signs for diagnosis [TD85b, page 138].

The malignant-type calcifications tend to be numerous, tiny, clustered, highly irregular in shape and of varying size [GBC87]. They also show great variations in density within an individual spot and among adjacent spots [TD85b, page 138].

Benign-type calcifications tend to be fewer, larger, more scattered, rounder and of less variation in size [GBC87]. Their densities are more uniform.

Since the size of the individual spot in calcifications is very small (0.25–0.75mm in diameter), microfocus magnification mammography is used for detailed analysis.

2.4.4 Thickened Skin Syndrome

The thickness of the skin of normal breasts seldom exceeds 1.5 mm [GBC87]. When the overall density of the breast is increased due to the high fluid content, the skin of the affected breasts will become many times thicker. Such an increase will give a very striking clinical and mammographic appearance [TD85b, page 218].

The cause of an increase in high fluid content may be due to lymphatic obstruction, lymphatic spread of breast cancer cells, inflammation or even heart failure [TD85b, page 213] [CBRE88, pages 43–44] [GBC87].

Chapter 3

Developments on Automated Breast Tumor Detection

3.1 Introduction

The importance of implementing a mass screening program using mammography to detect early breast cancers has been discussed in previous chapter. It was also mentioned that the problem of implementing such a screening program is the need for interpreting a large number of breast mammograms. Computer automation seems to be a potential solution to the problem.

This chapter contains a survey of the developments on automated detection of breast tumors and related research work. There are two reasons for including such a survey in this thesis. First, it provides an overview of

all techniques that have been used in this area. Secondly, by analyzing the strengths and the weaknesses of different approaches, it provides some ideas about new research trends.

Before looking into some of the work, it is important to understand the characteristics of breast images and the general tumor detection process. This is because most of the techniques employed are application dependent and oriented to overcome some of the limitations of mammograms. The characteristics of mammograms that make breast tumor detection difficult include:

- the small contrast between the tumor area and the normal tissue area.
This is due to the fact that the difference in density between normal and tumorous tissues is very small.
- the contrast reduction in the digitization process of the mammogram images.
- the presence of noise and anatomical structures that obscure the tumors.
- non-optimal exposure of the mammograms in some cases.
- the diffuse or partially obscured boundaries of the tumor areas.
- small tumor size (especially for tumors at an early stage).

In general, there are three basic steps in detecting tumors in the digitized mammogram images by computer:

- (i) locating the suspicious tumor areas in breast images.
- (ii) extracting the local features from the suspicious areas and their surrounding areas.
- (iii) classifying the suspicious areas (e.g. tumorous or non-tumorous areas).

It should also be noted that steps 1 and 2 are sometimes combined in a single step.

3.2 A Survey on Automated Breast Tumor Detection

This section contains a summary of previous research on using computers to analyze the breast images and to detect tumors in the images. The discussion is organized according to the approaches used, but it also follows the chronological order of development.

3.2.1 Feasibility Demonstration

Winsberg et al. [WEM⁺67] proposed computer automation of reading and interpreting mammograms as a solution to the problem of interpreting a large

volume of breast images produced in the exhaustive screening program using mammography. Some experiments were conducted to show the feasibility of this idea. The experiments involved computing two sets of statistics, one from the digitized left breast image and the other from the right breast of the same woman. Suspicious tumor areas were determined by comparing the two sets of statistics.

Outline of the Method

A mammogram was first positioned in a pre-defined manner relative to the chest wall, the skin and the nipple, and then digitized with 16 different grey levels. The image was divided into 8×8 windows. Four statistics characterizing the distribution of the grey levels within each window were obtained. If the statistics were greater than some preset thresholds, the corresponding window would be considered as a suspicious tumor area.

Results and Comments

Since the method described above did indicate the tumor areas as suspicious windows in some given examples, to some extent it successfully demonstrated the feasibility of computer automation of breast tumor detection. Besides, it introduced the concept of data reduction using small windows (also called primary resolution cells) which was considered useful in further research.

However, the work can only be considered as an initial step towards computer automation. The inadequacies include:

- (1) Human assistance is still required in positioning the mammogram in the digitization process.
- (2) The exact location, size and type of the tumor are not known. Only a rough location of the tumor is given.
- (3) There are too many false alarms in which the normal parts of the breast are mistaken as suspicious areas.
- (4) There is no solid mathematical support for the statistics used.
- (5) Other important features of breast tumors such as spiculation and calcification are not properly taken into consideration.

3.2.2 Tumor Classification

As mentioned in the introduction of this chapter, a fully automatic breast tumor detection process normally involves three basic steps: locating the suspicious tumors, extracting the local features and classifying the tumors. Since mammographic images are extremely noisy and of very low contrast, it would be too complicated to develop a system combining all three steps into a single step. Therefore, in some early research, the scope of the problem is first limited to extracting local features from given tumors and determining the

probable nature (benign or malignant) of the tumors. Since human assistance is needed to identify the tumors in the first place, such systems cannot be considered fully automated.

Among early work, two approaches in determining the nature of a tumor will be discussed. The first is more general and applicable to almost all kind of tumors (except thickening skin syndrome). The second is more restricted in application but it can be implemented very efficiently.

3.2.2.1 General Malignancy Differentiation

Ackerman and Gose [AG72] developed four malignancy measurements for isolated breast tumors. Based on these measurements, the given tumors are classified as either malignant or benign.

Outline of the method

A tumor in the mammogram is first circled manually with a black pen before digitization. The black circle in the digitized image is then identified by combination of thresholding and tracing. Only the area enclosed by the circle is considered as the tumor area.

Within the tumor area, four malignancy measurements are computed. They include calcification, spiculation, roughness and the perimeter-to-area ratio.

Calcification

Calcification is one of the primary signs of breast cancers, and shows up as a number of small spots (0.25–0.75mm in diameter). The existence of these small spots can be detected by the presence of four sharp edge points in a small area. The derivatives of these edge points should point inwards from 4 different directions and have a magnitude higher than a certain threshold. The threshold value is computed from the histogram of the derivatives of all points in the tumor. The number of small spots detected is taken as the calcification measure.

Spiculation

The shape of the tumor boundary is another important measure of the nature of the tumor. A tumor with a round border is probably benign in nature. Whereas, a malignant tumor tends to have a spiculated border formed by bunches of spicules radiating from the center of the tumor.

As a measure of radiality, the spicularity is defined as follows:

$$\text{Spicularity} = \frac{\sum_{\text{tumor}} |\vec{v} \cdot \hat{r}|}{\sum_{\text{tumor}} |\vec{v}_x| + \sum_{\text{tumor}} |\vec{v}_y|} \quad (3.1)$$

where \vec{v} is the brightness gradient vector at each tumor point with

components \vec{v}_x and \vec{v}_y in the x and y directions, respectively.

\hat{r} is the unit radial vector from the center of the tumor to the tumor point.

The function of the denominator in the expression is to normalize the spicularity by removing the gain in digitization.

Roughness

A dotted pattern which comes from the projection of some spicules on the top view should appear at the center of a spiculated malignant tumor. When the tumor is divided into two parts (the inner circle and the outer annulus) with equal area, the roughness of the central inner circle normalized by that of the outer annulus will give a malignancy measure of the tumor.

$$\text{Roughness} = \frac{\sum_{\text{inner circle}} |I(x, y) - I(x, y + w)|}{\sum_{\text{outer annulus}} |I(x, y) - I(x, y + w)|} \quad (3.2)$$

where $I(x, y)$ is the grey level intensity at point (x, y) , and

w has preset values (2, 4, 8 or 16).

Perimeter-to-Area

The boundary of a tumor cannot be located precisely since it is very diffuse. Different boundaries will be located when different thresholds are used. To achieve a better response, the threshold is defined on the intensity histogram of the tumor area. As a measure of the shape of the tumor, the perimeter-to-area ratio is defined as follows:

$$\text{Ratio} = \frac{\text{number of edge points on the perimeter}}{\text{number of points enclosed by the perimeter}} \quad (3.3)$$

Classification

The tumor to be analyzed is classified as benign or malignant on the basis of the values of the malignancy measurements described above. The classification method used is the modified nearest neighbor technique, which is shown to be superior to the multivariate gaussian analysis for this application.

Results and Comments

The results on the experiment performed by Ackerman and Gose showed that the computer can classify the tumors as well as the experienced radiologists when the tumor area is the only part of the mammogram which is available to the computer and to the radiologist. However, radiologists can do a better job if the entire mammogram and results of clinical examination are available to them.

3.2.2.2 Specific Malignancy Differentiation

Smith et al. [SWGS77] introduced a measure of malignancy called a linear mass ratio (LMR). By measuring the density pattern of a line across the tumor, the ratio differentiates the adenocarcinoma (of malignant nature) from the fibrocystic disease (of benign nature).

Outline of the Method

The end points of a line across the tumor are first marked manually. These end points should just go beyond the tumor and stop on the normal tissues. The intensities of the points $I(x)$ on the line are measured by a scanning microdensitometer. A function $L(x)$ is defined as a linear interpolation between the two end points. To remove the effect due to normal tissues, the abnormality function $A(x)$ is defined as the difference between $I(x)$ and $L(x)$.

The linear mass ratio (LMR) essentially measures the deviation of $A(x)$ from a uniform luminance profile (for further details, see [SWGS77]). If it exceeds a certain threshold, the corresponding tumor will be classified as a malignant tumor (i.e. adenocarcinoma).

Results and Comments

The linear mass ratio, as a measure of malignancy, is effective in discriminating the adenocarcinoma from the fibrocystic disease. However, its usefulness in real application is very limited as it confines the analysis to one of the two specific tumors. Besides, human assistance is needed to mark the tumor to be analyzed.

3.2.3 Detection of Abnormalities

In comparison with tumor classification, detection of abnormalities in mammograms using computers is more important. In the mass screening program, a large volume of mammograms of healthy breasts will be produced. If an automated computer system can filter out most of the normal mammograms which do not contain abnormal tumor areas, radiologists can then concentrate on those few mammograms with a high possibility of containing cancers.

3.2.3.1 Detection using Textural Statistics

Kimme et al. [KOS75] developed seven textural measures to locate the abnormal breast areas in the mammograms.

Outline of the Method

The breast region in the mammogram is first isolated and then divided into small rectangular sections. The locations of the sections are defined relative to the nipple, the chest wall and the skin.

For each section, four different histograms about the distribution of the grey level intensities are formed, and seven textural statistics based on the histograms are computed. The seven textural statistics include the variance and the third moment about the mean of the grey level intensities, two measures of directionality, the variance and the mean of the gradients and the bimodality of the grey level histogram.

In order to remove unwanted abnormalities, the textural statistics are first normalized with the other sections in the breast region and then normalized with the corresponding neighboring sections from the image of another breast of the same woman. Finally, the sections are classified according to the normalized textural statistics.

Results and Comments

The method successfully classified the cancerous areas of several tested breast images as abnormal, but in addition about 4 percent of healthy breast areas were labeled as abnormal.

3.2.3.2 Detection using Textural and Shape Measures

Hand et al. [HSAA79] developed a method to identify the suspicious areas from a digitized breast image. This method involves extraction of the textural features and shape features from the image. The textural measures used are similar to those used by Kimme et al. [KOS75], but shape measures gave additional strength to Hand's method.

Outline of the Method

Three basic textural features are extracted from 10×10 windows called primary resolution cells (PRCs). The textural features include intensity, roughness and directionality.

Two directionality parameters, for example, are defined as follows:

$$\begin{aligned} D_x &= \frac{1}{90} \sum_{x=1}^9 \sum_{y=1}^{10} [f(x, y) - f(x+1, y)] \\ D_y &= \frac{1}{90} \sum_{x=1}^{10} \sum_{y=1}^9 [f(x, y) - f(x, y+1)] \end{aligned} \quad (3.4)$$

where $f(x, y)$ is the intensity of a pixel in the PRC.

Among groups of PRCs, two shape measures, the “circle-likeness” and the “star-likeness”, are computed. These two measures determine if the group forms a circular or a star-like object.

Take star-likeness as an example. It counts the number of surrounding PRCs whose gradients are perpendicular to the radials of the target PRC. That is the number of times that the coordinate (x_p, y_p) represents the location of the target PRC in the following equations:

$$\begin{aligned} x_p &= x_i \pm \frac{-r \times D_y}{\sqrt{D_x^2 + D_y^2}} \\ y_p &= y_i \pm \frac{r \times D_x}{\sqrt{D_x^2 + D_y^2}} \end{aligned} \quad (3.5)$$

where (x_i, y_i) is the coordinate of the PRC whose distance from the target PRC is r PRC units. The value of r ranges from 2 to 16.

The results from applying the feature measurements on both left and right breast images of the same woman are then compared. After filtering, thresholding and clustering, the suspicious cancerous areas are determined.

Results and Comments

The strength of the method described above is to take the shape features into consideration. It is very important as the most common malignant tumors are characterized by their star-like shape. However, the gradient (D_x, D_y) of the surrounding PRCs computed in equation 3.4 is hardly perpendicular to the arm (spicule) of the star-like tumor even when the spicule falls in the PRC. The reasons for this will be explained with more detail later (in section 4.4). Besides, it should not take all the surrounding PRCs into account without discrimination, as the pixels at the center of the tumor do not to relate the star-likeness of the tumor.

Another problem is that this method has high false-negative rate, i.e. it has a high percentage of missed suspicious tumors. Later work by Semmlow and et al. [SSA⁺80] who combined the techniques employed by Hand and other researchers successfully relaxed this problem.

3.2.4 Detection of a Specific Kind of Tumor

Detection of abnormalities in mammograms locates all suspicious areas exhibiting features of breast tumors of any kind. Since there are four different kinds of breast tumors, it is hard to make the features to be extracted very specific, or it will be overwhelmed with a large number of detailed features. Therefore, some researchers aim at detecting one kind of breast tumor at a time so that the criteria for detection can be more specific and precise.

Lai [Lai88] (see also [LLB88]) described a method to locate a specific kind of tumor, *circumscribed tumors* (The appearance of the tumors is described in section 2.4.2).

Outline of the Method

Mammograms are first filtered using a variation of the median filter. The suspicious tumor areas are then located by template matching with the normalized cross-correlation as the similarity measure. Two heuristics are also used to reduce this false alarm. (Further details of this method will be discussed in section 4.5.)

Results and Comments

This method successfully identified the circumscribed tumors as suspicious areas in sample mammograms. However, the problem of false alarms still existed in most of the mammograms.

3.3 Related Research

Apart from major developments and research about automatic breast tumor detection, there are some other related research areas worth mentioning. They include mammographic image enhancement and radiographic tumor detection. Although this research is not exactly in the area of breast tumor detection, it does share some common techniques.

The discussion on object detection in general is too broad to include in this thesis. For interested readers, general references are available in [LZ85], [CD86], [BB82], [RK82, chapters 9–10] and [Hor86, pages 335–345].

3.3.1 Mammographic Image Enhancement

Image enhancement is a process to make selected features of an image more visible and to suppress useless information. Some enhancement techniques are specially designed for enhancing mammographic images or other radiographic images. These techniques may be used to improve the image quality in a way that subsequent detection can be done more easily.

Kruger and et al. [KHDL71] suggested some position variant and invariant, linear and nonlinear filtering techniques for enhancing the radiographic images. McSweeney and et al. [MSE83] applied a photographic technique (called blurred mass subtraction) on mammography to improve the visibility of small, low-contrast objects. This photographic technique can be easily implemented in a computer program. Dhawan et al. [DBG86] took an adaptive neighborhood processing approach to enhance the contrast of breast images.

Although the methods mentioned above do make some selected features more apparent to the human visual system, it does not necessarily imply that it can help subsequent detection in the computer. Besides, enhancement may introduce noise to the original images. Therefore, employing enhancement techniques in the detection process should be done with great care.

3.3.2 Radiographic Tumor Detection

Many techniques have been developed for detecting tumors in radiographs of any body part. Since these techniques may also be applicable to detecting breast tumors, they are worth mentioning.

Ernest and et al. [HKD⁺71] presented a survey about feature extraction techniques used in radiographic images. The techniques include directional signatures, contour tracing, frequency signatures and template matching. This survey can be viewed as an introduction to the feature extraction techniques in tumor detection.

Chow and Kaneko [CK72] developed the dynamic thresholding technique to detect the boundary of the left ventricle from the cardiac cineangiogram ¹. The threshold values are set dynamically according to the local characteristics computed from the intensity histogram of the local region in the image.

Ballard and Sklansky [BS73] employed the closed best curve approach with dynamic programming to find out the complete boundaries of tumors of roughly circular shape in chest radiographs and in liver isotope scans. When the entire boundary of a tumor is traced, the size and the location of the tumor are easily defined.

¹Cardiac cineangiogram is an X-ray motion picture of the heart image.

Sklansky [Skl76] discussed boundary detection techniques for some other objects in the radiographs. The objects include lung tumors, the entire lung, the ribs and the entire breast.

3.4 Conclusion

In this chapter, the developments of automated breast tumor detection and work in related areas are summarized. It provides an overview of the techniques that have been used before and gives a hint to the trends in new developments. The approaches used in the breast tumor detection range from the detection of abnormalities in general (in section 3.2.3) to the detection of a more specific kind of tumor (in section 3.2.4). More recent approaches allow the detection criteria to be more precise and so the detection programs can search for more specific features of that particular kind of tumor. This makes the detection more accurate and efficient.

Following this trend, we will analyze the problem of detecting another kind of breast tumor called *stellate tumors*, and present a way of detection in the next chapter. Compared with the detection of circumscribed tumors [Lai88], the detection of stellate tumors is considered to be more difficult as the structures of the stellate tumors are more complicated and their borders are more diffuse. Perception of stellate tumors in mammograms often causes a problem even to radiologists.

Chapter 4

Stellate Tumor Detection

4.1 Introduction

This chapter will focus on the detection of *stellate tumors*. The outward appearance of a stellate tumor (see Plate 2.1, page 15) looks like a star. It can be decomposed into two basic components: the central tumor mass and the radiating structure. Since most breast cancers have the appearance of a stellate tumor, the detection of stellate tumors is very important. However, it should not be mistaken that all stellate tumors are malignant. Some stellate tumors are in fact benign. When compared with the malignant stellate tumors, the benign ones have the same basic structure but different fine details (see Section 2.4.1). These fine details can be very useful in analyzing the nature of the stellate tumors, yet they give no definitive diagnosis. A definitive

diagnosis about the nature of a stellate tumor is usually obtained from other techniques such as biopsy and ultrasonography, after it has been detected in the mammogram (breast image). Therefore, an automated detection system should aim at detecting stellate tumors of either nature according to their basic structure projected in the breast image, and let subsequent other tests differentiate the two.

In the following sections, the discussion will begin with an analysis of different approaches to detect a stellate tumor in a breast image. An overall detection system is then proposed. In the proposed detection system, there are four processes which contribute to the successful detection of stellate tumors. They are the *preprocessing process*, the *radial structure recognition (RSR) process*, the *central mass detection (CMD) process* and the *classification process*. The preprocessing process isolates the breast region of the breast image and then normalizes it. The RSR process is responsible for identifying the existence of any radiating structure (one component of a stellate tumor) in a normalized breast image. The CMD process is responsible for detecting the existence of the central tumor mass of a stellate tumor (the second component). The classification process removes the non-tumorous areas detected by the previous detection process and classifies the tumor types of suspicious areas.

At the end of this chapter, experimental results are given to demonstrate the effectiveness of the proposed detection system.

4.2 Overall Strategy for Detection

There are two major approaches to the detection of stellate tumors in a breast image:

- (I) The entire stellate tumor is taken as one piece. A detection process is designed to search for all objects which have the star-like shape of a stellate tumor.
- (II) A stellate tumor is considered to be an object containing two components: a bright tumor mass and a radiating structure, where the tumor mass is located at the center of the radiating structure. Two processes are designed to recognize these two components individually. There may or may not be any communication or data flow between the two processes.

To design a process to detect the stellate tumor in approach I is very difficult as the star-like structure of a stellate tumor is complicated. Methods like template matching are not even applicable. In template matching, there should be a representative of the object to be searched for. This representative or template should match the search object after a simple transformation such as translation, rotation, scaling and reflection is applied. However, good representatives of stellate tumors do not exist. As shown in Figure 2.1 (page 15), the spicules of a stellate tumor radiate from the central tumor mass in all directions. Their number, length and width are not confined

to some fixed number. Besides, the orientations of some spicules deviate from the radial directions. Therefore, template matching does not work in detecting stellate tumors.

Other methods which involve detection of the boundary of a stellate tumor do not work, either. This is because the boundary of a stellate tumor is too diffuse and the image is too noisy. Even good edge detectors cannot locate the boundary of stellate tumors. Plate 4.2b (page 76) shows the edge image formed by applying the Canny edge operator (discussed in Section 4.4.2.1) to a segment of a breast image (Plate 4.2a) containing a stellate tumor. The boundaries of the stellate tumor are difficult to trace, in particular boundaries defined by fine spicules. Similarly, a contour detection algorithm based on the minimum radial inertia criterion [GPP88] does not work in this case. The situation is even worse when the stellate tumor is not as easily perceivable as in Plate 4.2a (page 76) or when the stellate tumor is partially obscured by anatomical structures.

Thresholding seems to be a way to identify the entire stellate tumor as a bright object, and then some other method can be used to identify the shape of the segmented object. Among different thresholding techniques, dynamic thresholding suggested in [CK72] seems promising as it employs local information to determine local threshold values to segment the image and has been shown to be successful in detecting the boundary of the left ventricle of the heart in a cardiac cineangiogram image. However, there

are still several potential problems in this method. First, it only focuses on the brightness of the segmented object but has not integrated its shape information. This not only degrades the quality of segmentation but also introduces a large number of false alarms. Second, segmentation through thresholding is not a direct way of detection since it does not directly produce a list of suspicious tumors. A subsequent process is needed to locate the segmented objects and identify their shape. Third, the spicules of a stellate tumor are easily lost in segmentation, and anatomical structures of the breast may be attached in the segmented stellate tumor. These make the subsequent shape identification process prone to errors.

Some of these problems can be overcome in the second major approach to detection of stellate tumors. This approach tries to simplify the detection problem by considering stellate tumor as an object composed of two components: a bright central mass and a radiating structure. Two separate processes are designed, each responsible for identifying one component of the stellate tumor. Since the two processes may or may not share information, there are four variations in approach II:

- (a) The two processes work in a totally independent manner. One process forms a list of bright (approximately circular) objects detected in the breast image, and the other process forms another list which contains objects with radiating structures. If objects with compatible size and location exist in both lists, they are taken as suspicious stellate tumors.

- (b) All objects with radiating structures are detected first and then a process is used to check if the central parts of the radiating structures are of high intensity.
- (c) All roughly circular, bright objects are detected first and then a process is used to check if they are surrounded by radiating structures.
- (d) The two processes work in a cooperative manner. Temporary data and images are passed between the two processes.

Variations a and b require a process to detect the radiating structure of a stellate tumor independently. The problems that arise will be similar to those when the stellate tumor should be detected as one piece. It is difficult to detect the radiating structure formed by various number of spicules whose lengths and orientations are not fixed. Comparatively, variations c and d are more flexible as the roughly circular, bright tumor mass of a stellate tumor is more outstanding and of a simpler structure. Between these two variations, variation d is preferable. Since it allows the sharing of information and interleaving between the two detection processes, a more effective detection algorithm is developed.

In conclusion, approach II variation d which employs two individual processes to detect the two components of a stellate tumor in a cooperative manner is the preferable approach, and it will be employed in the proposed detection system.

4.2.1 Overview of the Detection System

The proposed detection system is a system which takes mammograms as input, converts them into computer-accessible images, locates suspicious stellate tumors in the images, and finally produces a corresponding list of suspicious stellate tumors. Figure 4.1 shows an overview of the proposed detection system.

Digitization

The first step in the automated detection system is to convert the mammograms to numerical data that the computer can access and process. This is done by using a TV camera to scan the mammograms and convert them into raster images.

In this implementation, mammograms are digitized as raster images of size 512×512 and with a grey level range from 0 (dark) to 255 (bright). Symbolically, the grey level intensity of a pixel P at (x, y) in the image is denoted by $I(x, y)$ or $I(P)$.

Preprocessing

The digitized breast images then undergo a preprocessing process. The processing process has two major functions. The first one is to isolate the breast region from the nonbreast region in the entire breast image. This

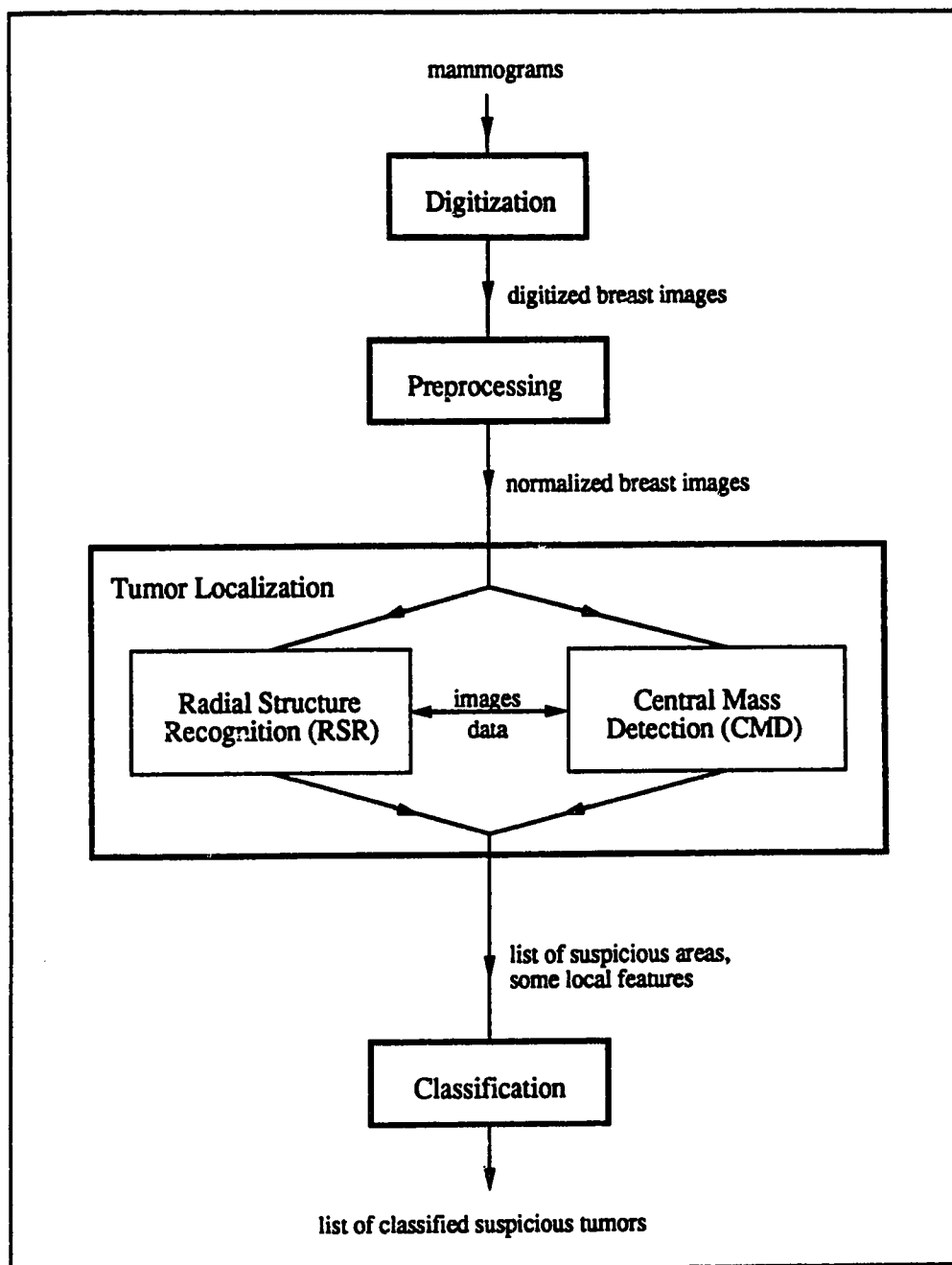


Figure 4.1: Data Flow of the Detection System

is very important as it will speed up the subsequent processes and avoid potential problems caused by high-contrast labels in the nonbreast region of the breast image. The other function of the preprocessing process is to standardize the breast regions across all breast images. Images produced in this preprocessing process are called *normalized breast images*.

Tumor Localization

Tumor localization is the most important part in the entire automated detection system. It has to locate all suspicious tumor areas in the normalized breast images according to the basic structure of a stellate tumor. As discussed in the beginning of this chapter, it is good to use two cooperative processes for tumor localization. The process which is responsible for identifying the radiating structure is called the *radial structure recognition (RSR) process*, and the other which is responsible for detecting the bright central tumor mass is called the *central mass detection (CMD) process*. Figure 4.2 shows the interaction between these two process.

The RSR process will first detect all the spinal pixels (pixels on the central line of a spicule). By using the information obtained from the spinal pixels, it removes the spicules in the images. The reason for doing this is to remove the corresponding radiating structure which surrounds the central tumor mass of the stellate tumor. The resultant images will then pass to the CMD process. Since the central tumor mass of a stellate tumor will look like a

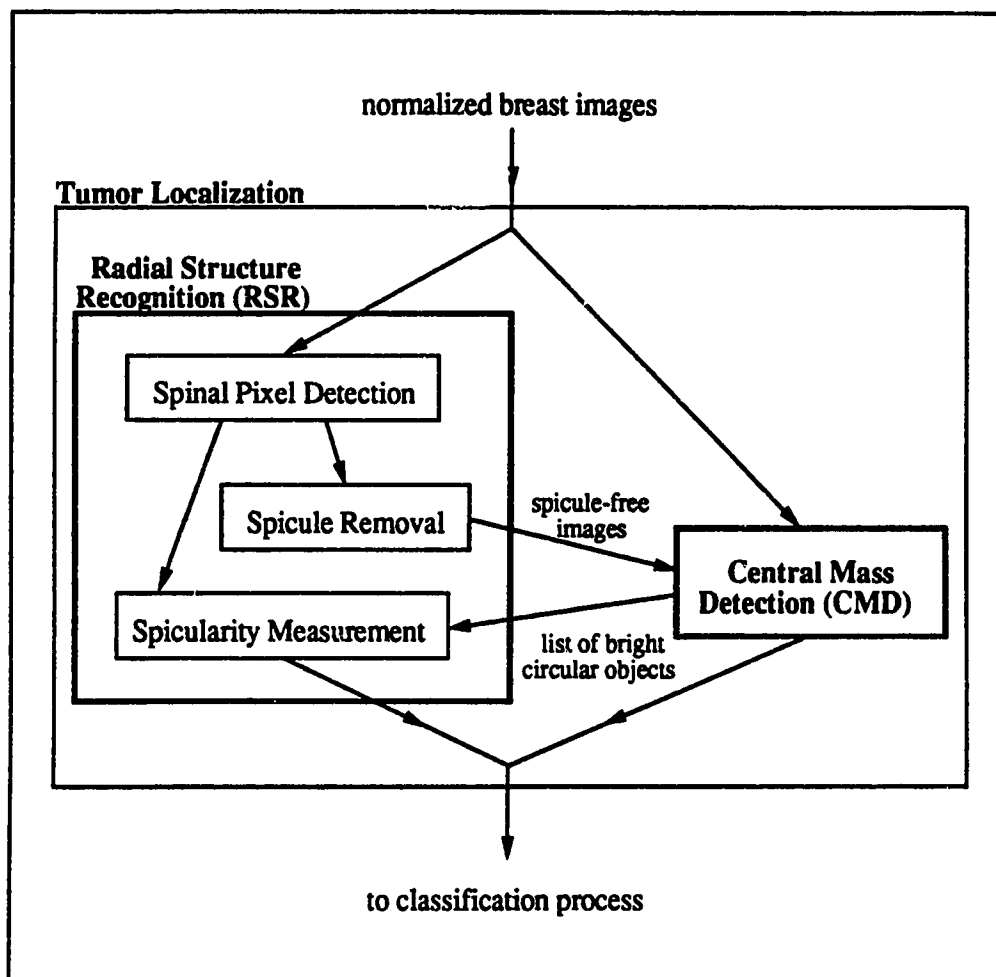


Figure 4.2: Data Flow between RSR and CMD processes

roughly circular bright object after the removal of the surrounding radiating structure, the CMD process can effectively detect all circular objects in the image as suspicious tumor areas. Then the list of suspicious tumor areas will be passed back to the RSR process and the likelihood of the existence of a radiating structure around each suspicious area will be measured. These likelihood measures will also be referred as the spicularities of suspicious areas. Finally, both the list and the spicularity measures will be forwarded to the classification process for further examination.

Classification

The classification process has two major functions. The first one is to determine whether the bright objects produced by the CMD process are really suspicious tumor areas. It will extract local information from the areas containing the bright objects and their surrounding regions. Combined with the spicularity measures obtained from the RSR process, the classification process will classify those detected objects and produce a list of suspicious stellate tumors for each mammogram input to the system.

In the following sections, there will be a detailed account for the rationale and implementation of each process in the system.

4.3 Preprocessing

The objective for the preprocessing stage is to prepare a given image for subsequent processing. Preparation may involve operations such as removing irrelevant information and identifying the target region for a given image. Preprocessing can reduce potential errors at later stages and make subsequent operations more effective and efficient.

In the breast tumor detection system, operations applied to the digitized breast images in the preprocessing stage (see Figure 4.3) include *breast region isolation* and *breast region normalization*.

There are several reasons why these two preprocessing operations are included. The original digitized breast image basically contains two regions: a *breast region* and a *nonbreast region*. If all subsequent operations can be restricted to the breast region, it will significantly shorten the subsequent processing time. Further, it can also eliminate the chance of picking the high contrast labels in the nonbreast region as suspicious tumor areas. These labels are usually found in mammograms and give patient numbers and dates.

The goal of breast region normalization is to standardize contrast across all breast images. The grey level intensities of the pixels in the breast region of each image are stretched to the same range. This makes the results generated in the subsequent processes comparable. As it will be seen in Section 4.3.2, breast region normalization also eliminates the potential abnormalities near

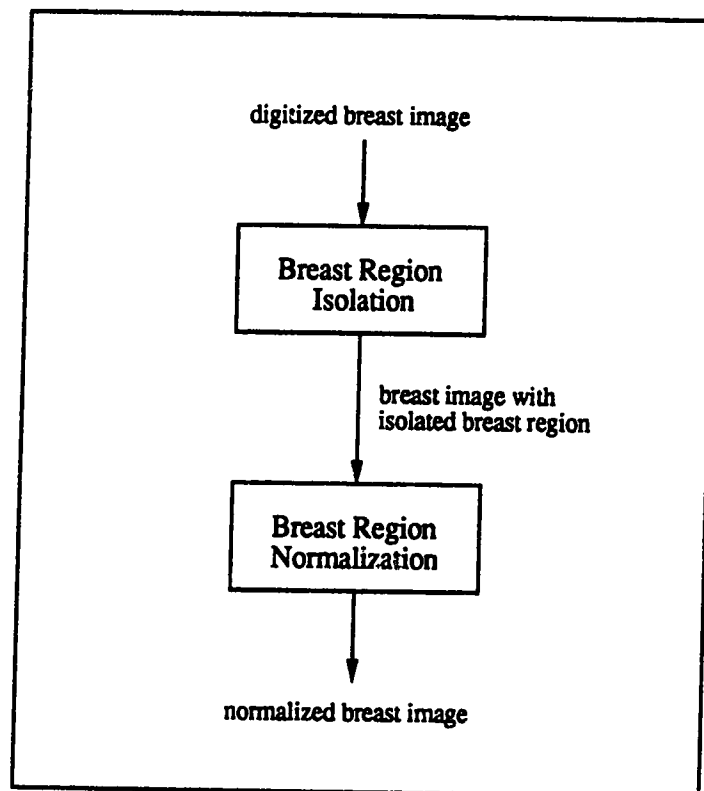


Figure 4.3: Preprocessing Process

the boundary of the breast region.

Other operations such as enhancement and noise filtering are not included in the preprocessing stage as they are detection-dependent. Their selection depends on what features in the images have to be enhanced or suppressed in the subsequent detection stage. Besides, they may enhance noise in the image.

In the following sections, we will look into the details of the two preprocessing processes: breast region isolation and breast region normalization.

4.3.1 Breast Region Isolation

Breast region isolation is a process to segment the breast region from the nonbreast region. The resultant output image of this process should satisfy the following two requirements:

- (1) It must contain exactly two 4-connected regions. They are the breast region and the nonbreast region. The intensities of the pixels in the breast region should not be modified but those in the nonbreast region should be set to zero. This allows subsequent processes to differentiate the two regions easily.
- (2) The boundary of the breast region in the output image should not be too rough. This avoids unexpected abnormalities which may occur near the boundary. The smoothed boundary should roughly correspond to the skin of the imaged breast and the chest wall.

Previous Work

There are many different techniques for segmentation available in the literature. They may range from simple grey-level thresholding [RK82, pages 61–83] to the more complicated ones such as edge-based segmentation [BB82, pages 119–144] or region growing [BB82, pages 149–164]. Some others may even employ topological information [BB82, pages 166–193] for segmentation.

However, only edge-based segmentation has been used in breast region isolation in the previous work [KOS75] [Sk176] [HSAA79] [SSA⁺80]. There are three steps in tracing the boundary of the breast:

- (1) The edges in the digitized breast image are enhanced by a gradient operator such as the Sobel operator [RK82, page 96].
- (2) The nipple is located and taken as the starting point for boundary tracing.
- (3) The subsequent pixels in the boundary are traced according to some heuristic rules.

Even though this breast boundary tracing method used in segmenting the breast region was considered successful in previous work, it has several weaknesses. Selection of the initial point for boundary tracing is complicated. In some cases, nipple points may not even exist on the breast boundaries. Besides, heuristics have to be used in guessing the next edge pixel on the

boundary when there are broken edge points. Further, a subsequent process is needed to identify pixels inside the detected breast boundary.

In this thesis, a new method for breast region segmentation is proposed. This method combines the simple grey level thresholding [RK82, pages 61–83] and blob coloring techniques [BB82, pages 151–152]. When compared with the edge-based segmentation mentioned above, it is simpler and direct.

Simple Grey Level Thresholding

The digitized breast image first undergoes an initial segmentation. It attempts to segment the breast region from the nonbreast region by a simple thresholding process. Pixels are simply classified according to their grey level intensities. If the grey level intensity of a pixel in the breast image is higher than or equals a particular threshold value, say T , the pixel is taken as part of the breast region and is set to a value of one. Otherwise, the pixel is assigned a value of zero which corresponds to a pixel in the nonbreast region. This thresholding process can be formulated as follows:

$$B_1(x, y) = \begin{cases} 1 & \text{if } I_0(x, y) \geq T \\ 0 & \text{otherwise} \end{cases} \quad (4.1)$$

where I_0 is the original image, and

B_1 is the resultant binary image.

The selection of the threshold value is quite critical as it directly determines the classification of a pixel. As listed in a survey paper [SSWC88], there are various methods to select the threshold value automatically. The choice of a particular method depends mainly on the nature of the image. For example, if the image contains homogeneous objects on a homogeneous background of sharp contrast, the lowest point between the two peaks in the bimodal intensity histogram of the image can be taken as a good threshold value. However, it does not work well in breast images as the breast region is not a homogeneous region. The intensity histogram of a breast image is simply not bimodal.

Instead of using some sophisticated methods, simple thresholding with preset threshold value is proposed here. There are several arguments to support this proposal:

- (1) The mammographic process and the digitization can be standardized. Therefore, the nonbreast background in the digitized breast image can be segmented with a predetermined threshold value.
- (2) Small deviations from the true breast boundary (the skin of the breast and the chest wall) are acceptable in this application.
- (3) A simple method can reduce the possibility of unpredictable abnormalities.

Blob Coloring

After applying simple thresholding to the original digitized breast image, a binary image is formed. If this binary image contains exactly one cluster of 4-connected 1's with no internal clusters of 0's, the breast region is simply represented by the area occupied by the cluster of 1's. However, this is often not the case. The binary image may contain several clusters of 1's. The extra "1" clusters may come from the high-contrast patient number or other labels in the original mammogram. Besides, clusters of 1's and 0's may enclose each other as it is shown in the binary image B_1 in Figure 4.4.

The following method (see Figure 4.4) is used to extract the breast region and to remove unwanted clusters. All subregions in the binary image B_1 are identified using blob coloring [BB82, page 151], a simple method for labeling 4-connected regions in a binary image. Then all regions with a pixel value of "0", with the exception of the largest region, are changed to "1". This results in image B_2 . Finally, all regions with a pixel value of "1", with the exception of the largest region, are changed to "0". This results in image B_3 .

Boundary Smoothing

When the breast region is initially segmented, its boundary may be rough. A simple method is designed to smooth out the boundary. Any pixel in the binary image, that has less than four (8-connected) neighboring pixels of the same value, is inverted (i.e. from 1 to 0 or 0 to 1). This is achieved by

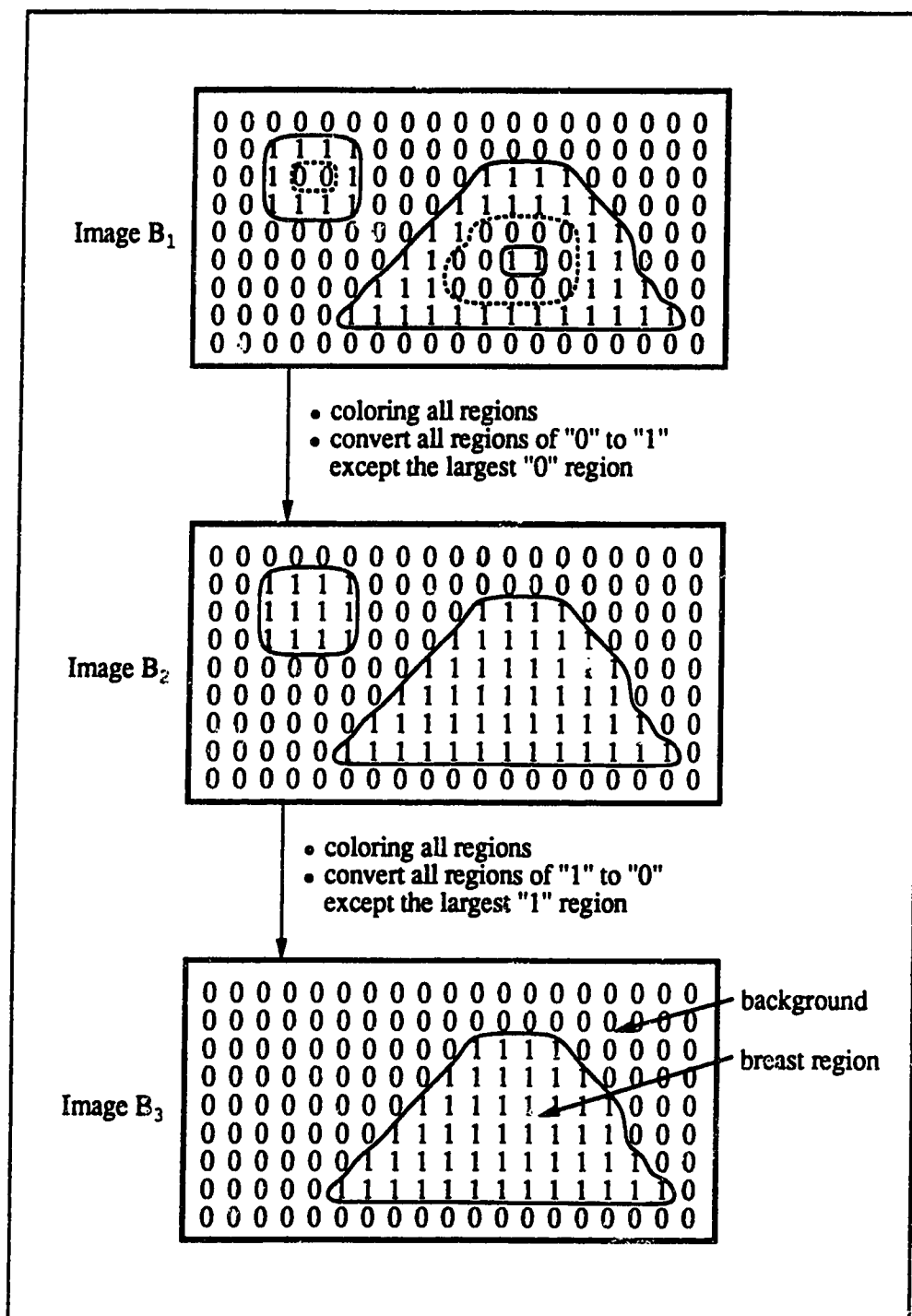


Figure 4.4: Breast Region Isolation

convolving the binary image B with the following kernel G:

-1	-1	-1
-1	9	-1
-1	-1	-1

and then the new pixel values are determined by:

$$\text{new } B(x, y) = \begin{cases} 1 & \text{if } G*B(x, y) \in [-8, -5] \cup [1, 5] \\ 0 & \text{if } G*B(x, y) \in [-4, 0] \cup [6, 9] \end{cases} \quad (4.2)$$

This process is repeated until no more changes are made. Since the operation in this smoothing process can be applied to all pixels simultaneously, it can be executed very fast by an array processor.

Normally, the resultant binary image after boundary smoothing should contain only one region of 1's which corresponds to the breast region. However, it is possible that new small regions of 1's are generated. Therefore, blob coloring should be applied once more in order to select the largest region of 1's as the breast region.

Finally, the image with the isolated breast region is formed by replacing the 1's in the binary image with the corresponding pixel values in the original breast image.

4.3.2 Breast Region Normalization

The breast region normalization process has two purposes. The first one is to standardize the range of grey level intensities of the breast pixels across all

breast images. The second one is to remove any abrupt intensity change at the boundary of the breast region. To achieve these two purposes, the grey level intensities of the pixels in the breast region are stretched to the range from 1 to 255. The normalized breast image is then given by:

$$\text{new } I(x, y) = \begin{cases} 1 + 254 \cdot \frac{I(x, y) - \min I}{\max I - \min I} & \text{if } I(x, y) \neq 0 \\ 0 & \text{if } I(x, y) = 0 \end{cases} \quad (4.3)$$

where $I(x, y)$ is the breast image before normalization, and $\min I$, $\max I$ are respectively the minimum and maximum grey level intensities of the pixels in the breast region.

This normalized breast image obtained after breast region isolation and breast region normalization is now ready for subsequent tumor detection. At this stage of processing the following has already been achieved:

- (1) Artifacts such as high-contrast labels have been removed.
- (2) Pixels in the nonbreast region (intensity = 0) can be easily identified and excluded from further processing.
- (3) The boundary of the breast region has been smoothed.
- (4) The abrupt intensity drop near the boundary has been removed.
- (5) The range of the grey level intensities of the pixels in the breast region has been standardized across all breast images.

Plates 4.5 and 4.6 (page 123) show two normalized breast images. They are formed by applying breast region isolation and breast region normalization to the breast images s1 and c1 (Plates 2.1 and 2.3, pages 18 and 19).

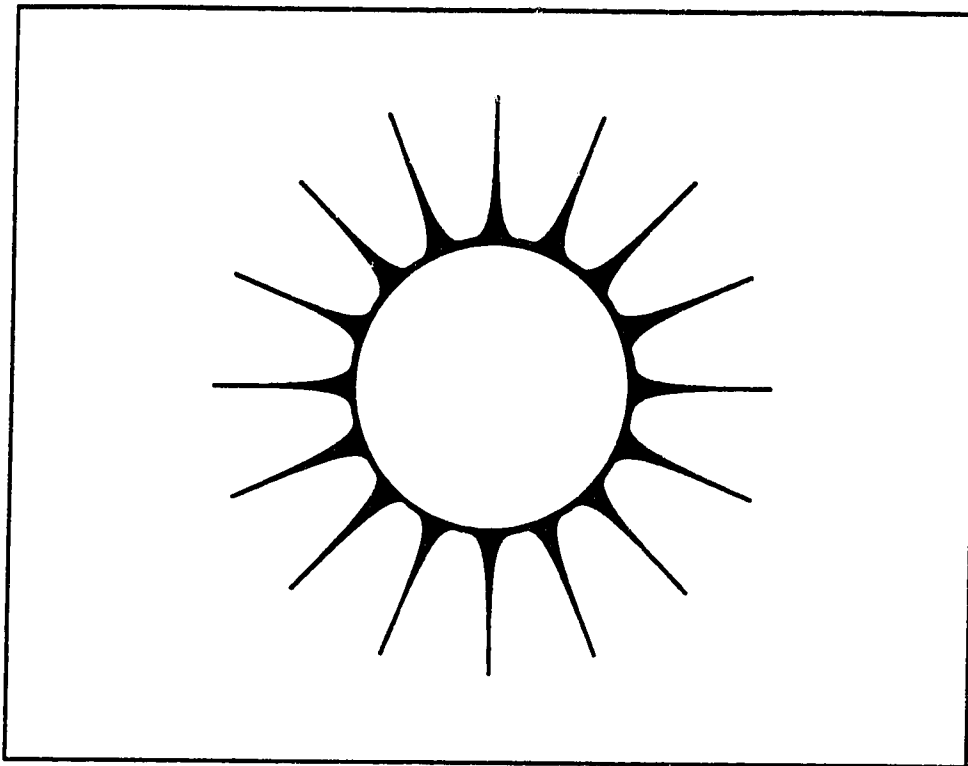


Figure 4.5: An Ideal Radiating Structure

4.4 Radial Structure Recognition (RSR)

The radiating structure of a stellate tumor is formed by a number of spicules which surround and radiate from the central tumor mass of the stellate tumor. Figure 4.5 gives an ideal example of such a radiating structure. It should be noted that the circular area at the center of the structure is not considered as part of the radiating structure. Only the spicules contribute to the radiating structure.

The radial structure recognition (RSR) process is a process designed to capture the radiating structure of a stellate tumor. The RSR process cooperates with the central mass detection (CMD) process to locate suspicious stellate tumors in a breast image. Figure 4.2 (page 48) shows the relationship and the data flow between these two processes.

Three different approaches to radial structure recognition (RSR) are proposed and tested. The best one is selected and used in the RSR process to locate and remove the spicules in the radiating structure. After the completion of the central mass detection (CMD), it will also be used in measuring the spicularity (the likelihood of the existence of a surrounding radiating structure) of objects detected by the CMD process.

4.4.1 Gaussian Operator

Before going into details of the three proposed RSR approaches, the Gaussian operator and its derivatives are introduced first as they will be referenced quite often in the later discussion.

The Gaussian operator is a smoothing filter which can be used to blur (smooth) an image and effectively to remove fine details [Mar82a]. The Gaussian operator (also referred to as the “normal distribution” in statistics) is defined in one dimension by:

$$G(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{x^2}{2\sigma^2}} \quad (4.4)$$

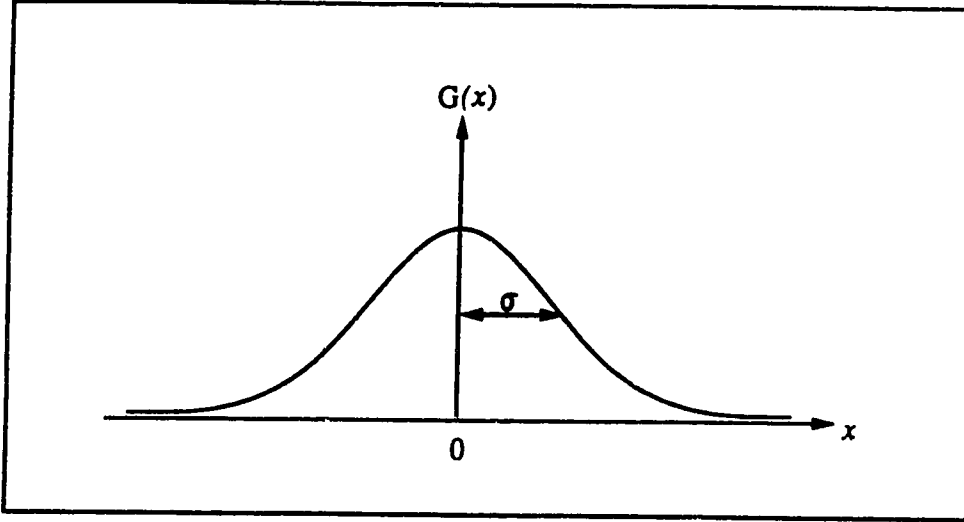


Figure 4.6: Gaussian Distribution

where σ is the standard deviation. Figure 4.6 shows the distribution curve.

and in two dimensions by:

$$G(x, y) = \frac{1}{2\pi\sigma^2} \cdot e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (4.5)$$

The advantage of the Gaussian operator lies in the fact that it is smooth and localized in both the spatial and the frequency domains. It does not introduce any changes that did not exist in the original image [YP86].

Further, the Gaussian operator is quite flexible. By selecting the value of the space constant σ (which is the standard deviation of the corresponding Gaussian distribution) in equation 4.5, the smoothing power of the Gaussian operator can be adjusted. The larger the space constant is set, the stronger is the smoothing power.

Mathematically, applying the Gaussian operator G to an image I is to convolve the function G with I . In the discrete domain, simple sampling is used. The resultant image F is given by:

$$F(x, y) = \frac{1}{(2N + 1)^2} \cdot \sum_{v=-N}^N \sum_{u=-N}^N G(u, v) I(x - u, y - v) \quad (4.6)$$

This is written as $F = G * I$, where $*$ denotes convolution throughout this thesis.

Plate 4.1 shows the results of applying the Gaussian operator with different σ values to a small image segment which contains a stellate tumor. The resultant images are pseudo-colored in order to make the effect of smoothing more apparent. Every 5 consecutive grey levels are taken as an interval and a color assigned to them. In the smoothed image c (of Plate 4.1), the structure of the stellate tumor is more apparent. However, if it is over-smoothed (as shown in image d), all fine details of the structure are smeared.

The first derivative of the Gaussian operator is also very useful in image processing. Its x and y components are given by:

$$\frac{\partial G}{\partial x} = \frac{-x}{2\pi\sigma^4} \cdot e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (4.7)$$

$$\frac{\partial G}{\partial y} = \frac{-y}{2\pi\sigma^4} \cdot e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (4.8)$$

When convolved with an image, it gives the gradient information about the corresponding smoothed image. Edge points in the image can then be detected as the locations having peak gradient values (see Figure 4.7c) at some

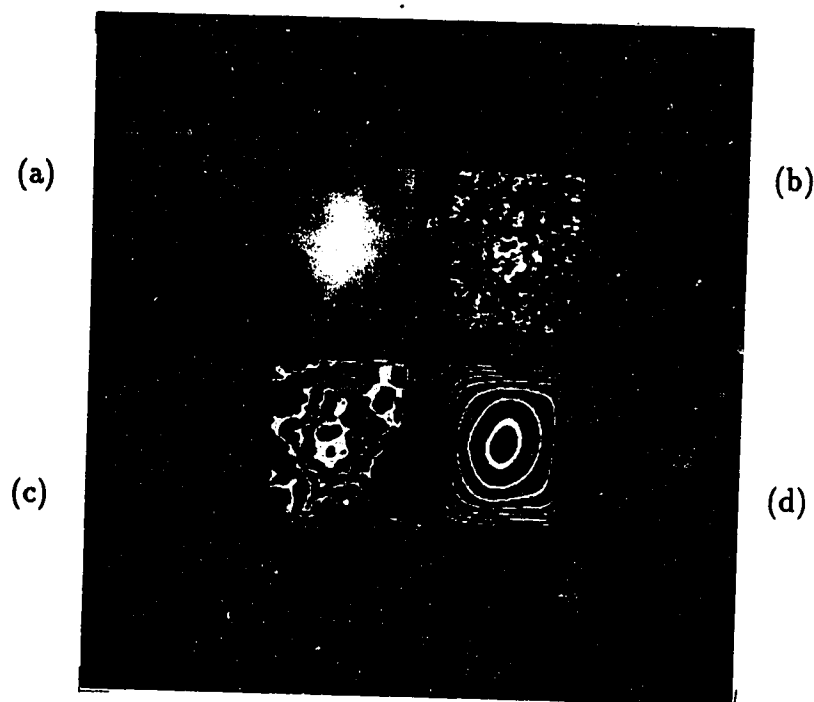


Plate 4.1: Effect of Gaussian Operator

- (a) original image
- (b) pseudo-colored original image
- (c) pseudo-colored smoothed image, $\sigma = 1.5$
- (d) pseudo-colored smoothed image, $\sigma = 10$

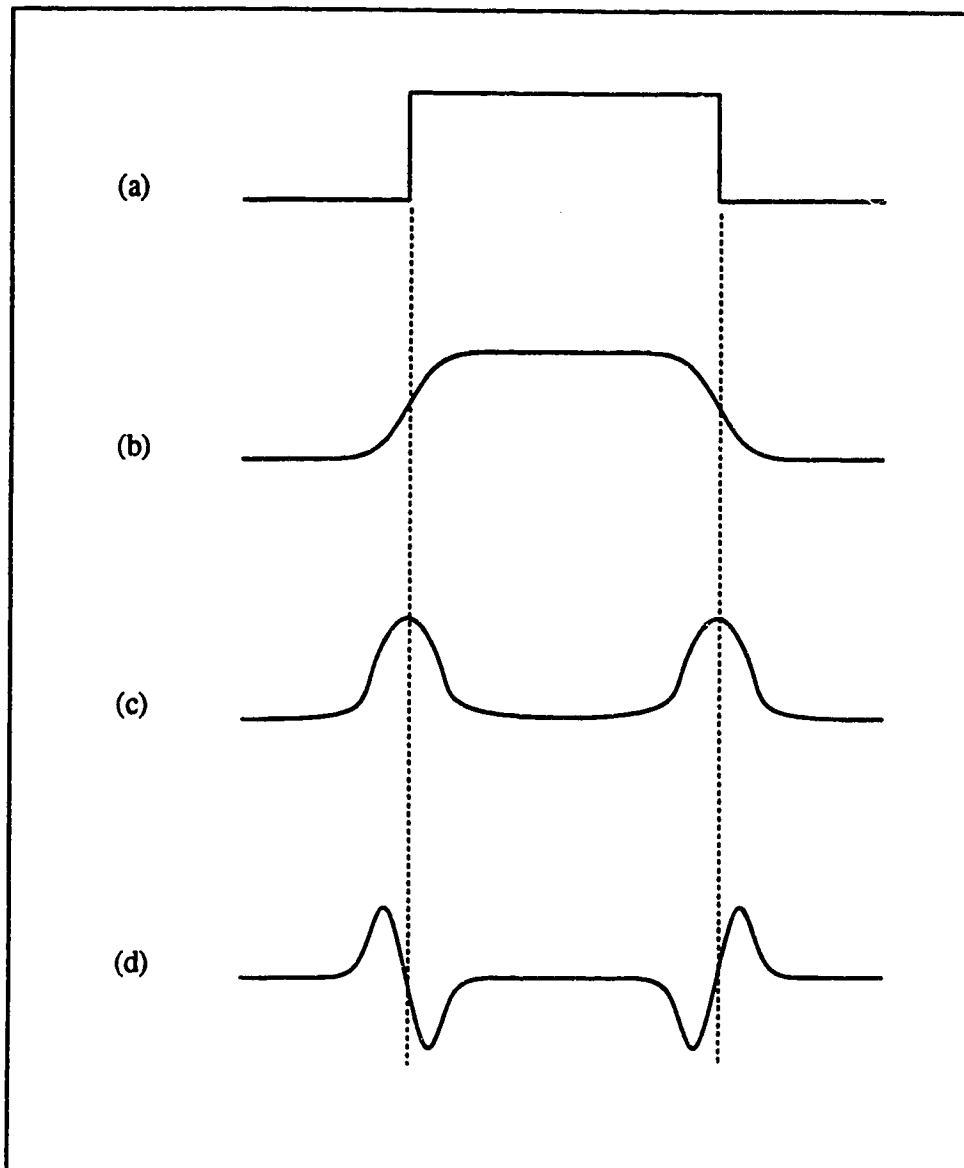


Figure 4.7: Response in Edge Detection

(a) original image(I)

(b) Gaussian smoothing ($G \ast I$)

(c) magnitude of the first derivative of $G \ast I$

(d) second derivative of $G \ast I$

orientations. The scale (coarse or fine) of edges detected is controlled by the value set for the space constant σ .

It should also be noted that the x and y components of the Gaussian operator and its first derivative are separable. If the x and y components of a filter H are separable (i.e. H can be rewritten as $H_x * H_y$), the filter can be decomposed into two linear filters. Convolving an image with the two linear filters in two steps is equivalent to convolving the image with the original filter before decomposition. For example, if I is an image and H is the separable filter, then

$$\begin{aligned} I * H &= I * (H_x * H_y) \\ &= (I * H_x) * H_y \end{aligned} \quad (4.9)$$

This allows more efficient implementation of the convolution.

Another way to detect edge points in an image I is to apply the Laplacian $\nabla^2 (= \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2})$ to the Gaussian smoothed image. Edge points in the original image are then given by the zero-crossings¹ of $\nabla^2(G*I)$. Figure 4.7d (page 66) shows an example. The Laplacian is important as it is the lowest-order isotropic differential operator. Further, the following mathematical identity allows an efficient implementation of the Laplacian:

$$\nabla^2(G*I) = (\nabla^2 G) * I \quad (4.10)$$

¹Clark [Cla89] showed that under certain conditions a zero-crossing may not correspond to a real edge in the image.

4.4.2 Three Different Approaches to RSR

Three different approaches to radial structure recognition (RSR) are proposed, discussed and tested in this section. The results are illustrated with a small image segment (which has been used in testing the effect of the Gaussian operator). It contains a stellate tumor and was segmented from the breast image shown in Plate 2.1 (page 18). The size of the test image is 200×200 . There are two reasons for using a smaller image. First, it saves processing time during testing. Secondly, it reduces the interference from other parts of the breast image. The best approach will then be used to identify radiating structures in full-sized breast images.

The three proposed RSR approaches are:

- (a) *Edge-oriented RSR*. This approach first detects all the edge pixels in the image and then determines how well these edge pixels form the (outer) boundary of a radiating structure.
- (b) *Field-oriented RSR*. In this approach, the radiating structure is taken as a radial field. By measuring the convergence of the field, it determines the likelihood of the existence of a radiating structure.
- (c) *Spine-oriented RSR*. This approach considers each individual spicule in the radiating structure as a unit. Pixels on the *spines* (central lines) of spicule-like objects are first detected, and then it is determined if they are pixels on the spines of the spicules in a radiating structure.

The design of each approach described above is based on assumptions about the shape of the spicules in the radiating structure. Therefore, their success depends on how close these assumptions match real cases. Experiments which employ a technique called *infinite line accumulation* (ILA) are used to test the appropriateness of the assumptions of the three RSR approaches and to compare their efficiencies in identifying a radiating structure.

Infinite Line Accumulation (ILA)

An oriented pixel in an image is defined as a pixel which has three parameters: image coordinate, orientation and strength. A set of oriented pixels is said to be convergent if their orientation vectors intersect at the same point.

Infinite line accumulation (ILA) is an instance of the Hough transform [BB82, pages 123–131] [RK82, pages 121–126]. It determines the degree of the convergence of a set of oriented pixels in an image (see Figure 4.8). An accumulation image which has the same size as the given image is formed first and is initialized by assigning all its elements zero. For each oriented pixel with image coordinate (x, y) , orientation (slope m) and strength (s) in the given image, a line passing through (x, y) with slope m is drawn on the accumulation image. The values of the pixels along the line on the accumulation image are incremented by s .

The convergence of the oriented pixels in the given image will then be determined by the existence of a high-intensity area (bright spot) in the

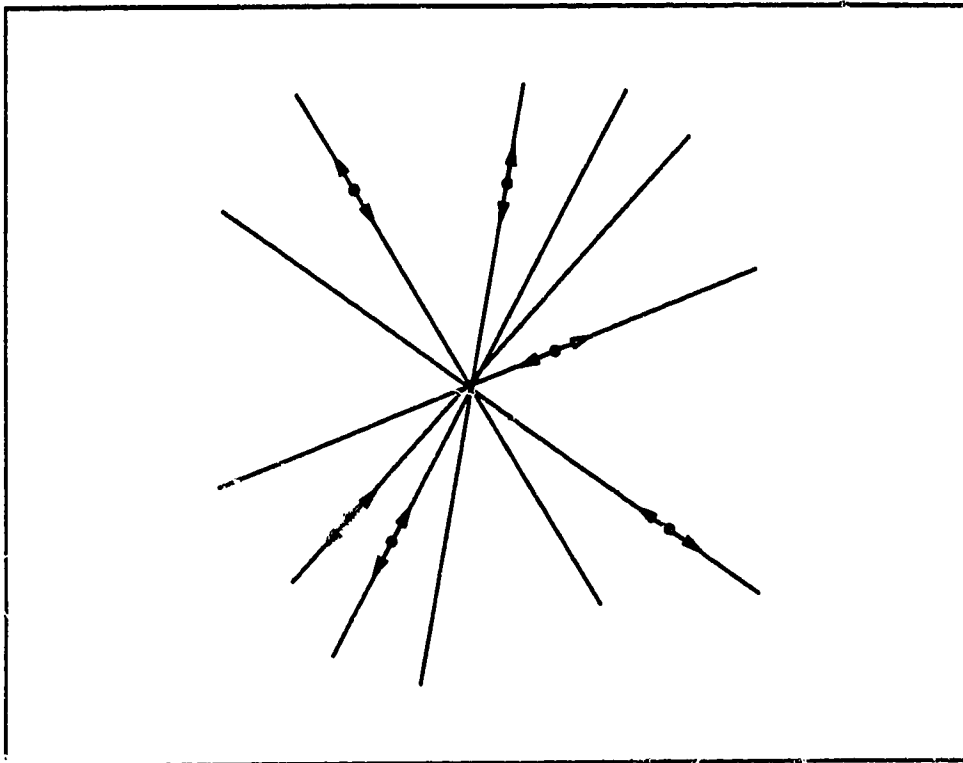


Figure 4.8: Infinite Line Accumulation

accumulation image.

4.4.2.1 Edge-oriented RSR Approach

The edge-oriented radial structure recognition approach is based on the assumption that the radiating structure is formed by a number of very sharp, radially arranged spicules. The edges of the spicules point close to the center of the radiating structure. If a line is drawn passing through the edge of a spicule and in a direction (edge direction) normal to the principal gradient change at that edge point, it should cut through the center of the corresponding radiating structure. Figure 4.9 illustrates the situation.

When all edge points in an image are detected and taken as oriented pixels with strength equal to the gradient, the infinite line accumulation (ILA) method can be applied to determine if radiating structures exist. Any bright spot in the accumulation image formed in ILA indicates a possible location of the center of a radiating structure.

Edge Detection

The most crucial step in the edge-oriented RSR is to detect the edge pixels in the given image. There are many edge-detection techniques available in the literature. Most of them are based on estimating the first or second derivative over some support. Examples include the simple difference operator, the Sobel operator [RK82, pages 85–96] and the Laplacian of the

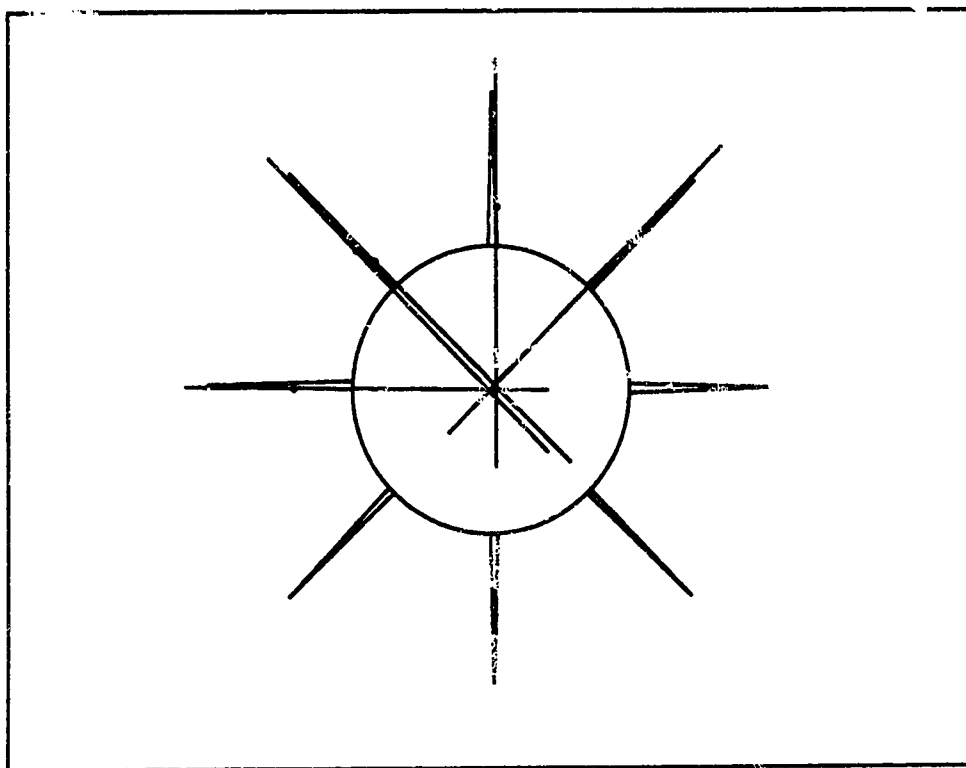


Figure 4.9: ILA in Edge-oriented RSR

Gaussian operator [Mar82a]. The edge operator used is also a differential operator, known as “Canny edge detector”. It employs the directional second derivative and detects edge pixels as zero-crossings.

Canny [Can83b] (see also [Can83a]) started with a set of edge detection criteria which capture desirable properties of a detector. They include:

- (1) high signal-to-noise ratio. This in turn minimizes the probability of failing to mark real edge points and the probability of falsely marking a non-edge point.
- (2) good localization. The detected edge points should be as close to the true edge as possible.
- (3) only a single response to an edge.

According to these three criteria, Canny deduced that the first derivative of a Gaussian is the optimal filter (with some approximation) to detect one-dimensional step changes. (For detailed derivation, see [Can83b].) A peak or trough in the first derivative filter indicates the location of a step change. In one dimension, this is equivalent to the zero-crossing of the second derivative of the Gaussian.

In two-dimensional images, the direction of an edge contour is perpendicular to the direction of the principal gradient \vec{n} (direction of maximum gradient change). By projecting the two-dimensional edge along \vec{n} , it gives

an one-dimensional step change. Therefore, detecting an edge in a two-dimensional image can be done by detecting the zero-crossing of the second derivative of the Gaussian in the direction of the principal gradient. i.e.

$$\frac{\partial^2}{\partial \vec{n}^2}(G*I) = 0 \quad (4.11)$$

where $\vec{n} (= \frac{\nabla G*I}{|\nabla G*I|})$ is the direction of principal gradient.

When compared with the Laplacian of the Gaussian ($\nabla^2 G$) operator, the directional second derivative of the Gaussian ($\frac{\partial^2}{\partial \vec{n}^2} G*I$) operator is superior for the following reason. Let \vec{z} be a unit vector along the edge contour (i.e. $\vec{n} \cdot \vec{z} = 0$). Then,

$$\nabla^2 G = \frac{\partial^2}{\partial \vec{n}^2} G + \frac{\partial^2}{\partial \vec{z}^2} G \quad (4.12)$$

Since the second component ($\frac{\partial^2}{\partial \vec{z}^2} G$) is along the edge, it can contribute nothing to the edge detection and the localization of the edge. However, it does increase the output noise of the edge detector by 60% [Can83a] [Cla89]. Therefore, $\nabla^2 G$ detector is not as good as the Canny detector.

When an edge detector is applied to breast images, it may produce some weak edges which do not correspond to the edges of any spicule. These weak edges may be due to random noise or due to normal characteristics of a breast image (its intensity increases gradually towards the center of the breast). In order to remove unwanted weak edge pixels, the edge detector is modified in the following way. Let P be a pixel in the image and M be the maximum gradient magnitude in the image. The magnitude of the gradient at P is

given by $|\nabla G * I(P)|$. In addition to the zero-crossing condition, a pixel P taken as an edge pixel should also satisfy one of the following conditions:

- (1) $|\nabla G * I(P)| > T \cdot M$, where T is a preset threshold value (0.05 in our implementation).
- (2) $|\nabla G * I(P)| > \frac{T}{2} \cdot M$ and P has a (8-connected) neighboring pixel satisfying condition (1) or (2).

Plate 4.2b shows the edge image formed by applying the modified directional edge detector (described above) to the test image (Plate 4.2a). When the edge pixels detected are then taken as the oriented pixels and the gradients as their strengths, the accumulation image (Plate 4.2c) is formed by applying the infinite line accumulation (ILA) technique. Plate 4.2d shows another accumulation image formed in the same way as Plate 4.2c except the strengths of all edge pixels are taken as value one.

Interpretation of the experimental results will be given in Section 4.4.3 where the edge-oriented RSR approach is compared with other RSR approaches.

4.4.2.2 Field-oriented RSR Approach

The field-oriented radial structure recognition (RSR) approach will conceive the radiating structure of a stellate tumor as a radial pattern formed by a large number of fine, radially arranged spicules or lines. The central part

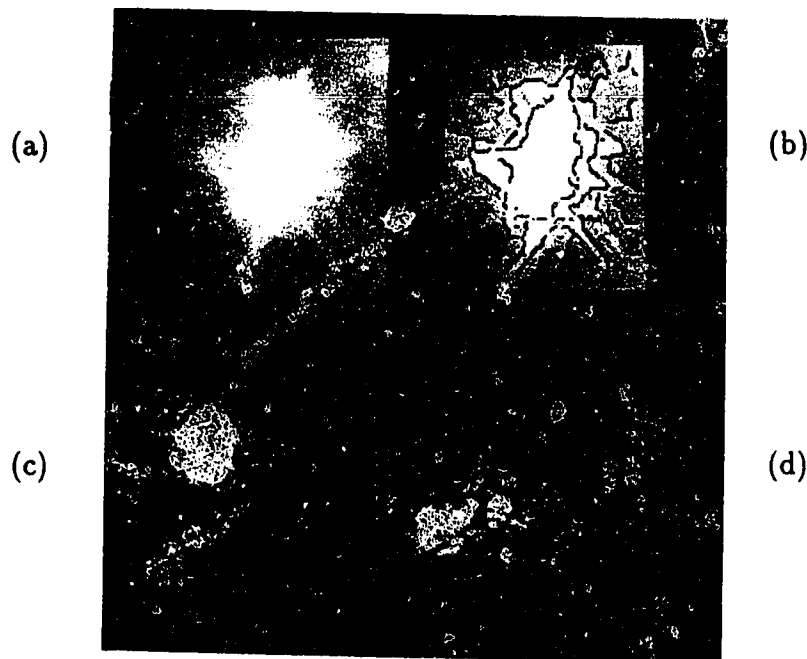


Plate 4.2: Result of Edge-oriented RSR

(a) original image

(b) edge image

(c) accumulation image (strength=gradient)

(d) accumulation image (strength=1)

of the radial pattern, which corresponds to the central tumor mass of the stellate tumor, is considered an area free of any recognizable pattern.

Since the lines forming the radial pattern are weak and fine, and many of them may even be broken, it is too difficult to detect each of them individually. However, as a whole they form a field which flows towards the center of the structure. Each pixel in the field can be taken as an oriented pixel with a strength value and a specific orientation. The strength is determined by the coherence of the local flow field. This coherence value measures the extent to which fine lines in a local area point into the same direction. Since an area not covered by the radiating structure normally does not have a preferred local orientation, the pixels in that area should have small coherence values. Therefore, by thresholding the coherence values of the pixels in the image, the oriented pixels in the radial pattern area can be detected. The flow direction (direction of the fine lines) in a small area centered at an oriented pixel can be taken as the orientation of that oriented pixel. Once all oriented pixels are detected and their strengths and orientations are computed, a radiating structure can be detected as an area of high intensity in the accumulation image formed by applying the infinite line accumulation (ILA) technique. Figure 4.10 illustrates the situation.

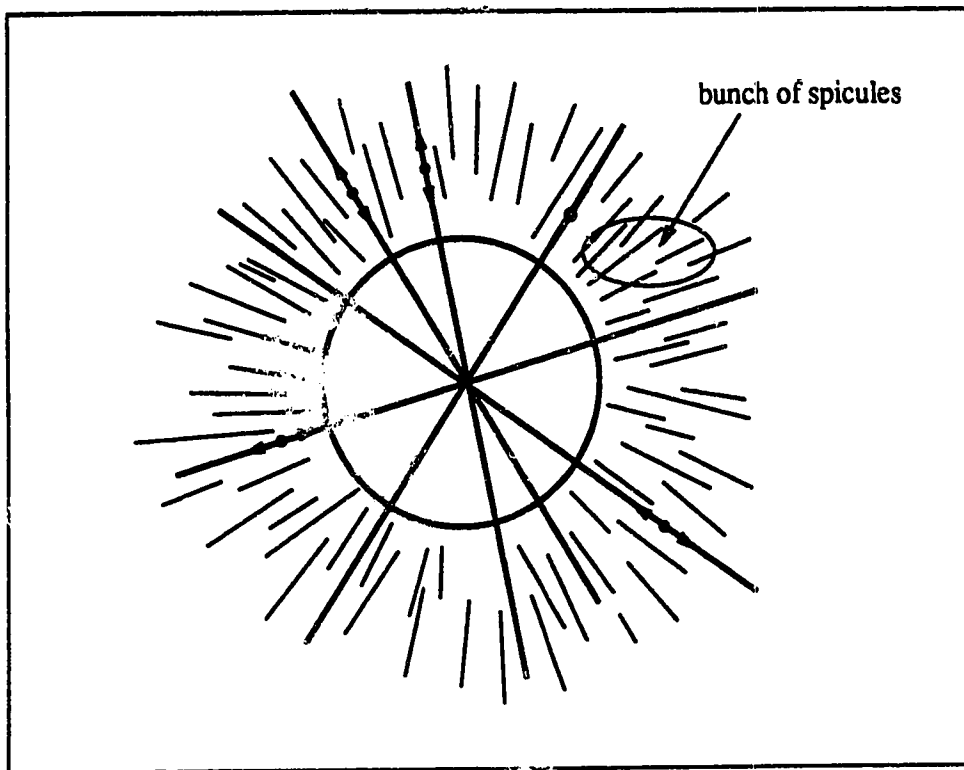


Figure 4.10: ILA in Field-oriented RSR

Flow Field Measurement

A simple way to measure the flow directions in an image (I) is to first smooth the image by convolving it with the Gaussian (G_1) operator. This will remove some high frequency noise. The flow direction at a point is then given by a direction normal to the gradient (g_x, g_y) of that point in the smoothed image. This two-step process can be further simplified by convolving the original image with the first derivative of the Gaussian to give the gradient images directly.

However, the gradient calculated by this simple method may not be good enough to estimate the principal flow direction in a local area. In order to improve the accuracy of the flow direction, it may be good to smooth the gradients within a small local area. Unfortunately, simple averaging or smoothing will cause a severe cancellation which will offset all potential benefits. The reason is that the gradient vectors of the nearby points in flow field are often π radians out of phase. When they are added together, they cancel each other. The gradient vectors \vec{a} and \vec{b} in Figure 4.11a are an example.

In [KW87], Kass and Witkin suggested a method to solve the cancellation problem. The key is to first double the angles of the gradient vectors before smoothing. This will make the gradient vectors, which are originally π out of phase and cancel each other, become in phase and reinforce each other. Figure 4.11b shows the result of doubling the angles of the gradient vectors

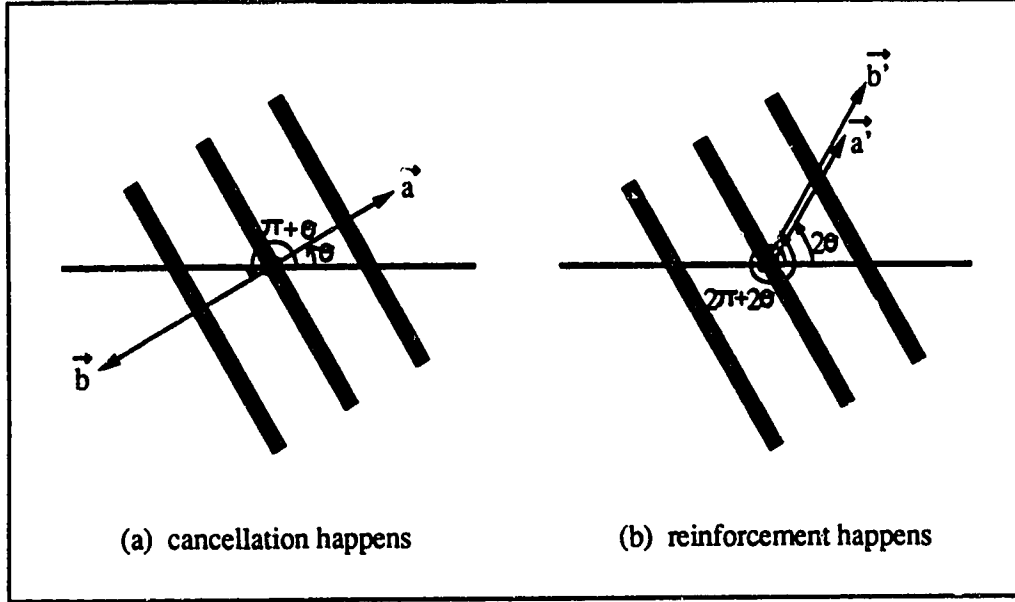


Figure 4.11: Effect of Angle Doubling

in Figure 4.11a. Mathematically, the double-angled gradient vector (g'_x, g'_y) at a point (x, y) can be obtained by:

$$g'_x(x, y) = \frac{g_x^2(x, y) - g_y^2(x, y)}{g(x, y)} \quad (4.13)$$

$$g'_y(x, y) = \frac{2g_x(x, y) \cdot g_y(x, y)}{g(x, y)} \quad (4.14)$$

where (g_x, g_y) is the gradient given by $(\nabla G_1) \cdot \mathbf{i}$, and $g = \sqrt{g_x^2 + g_y^2}$.

It should be noted that the strengths of the gradient vectors are the same² before and after angle doubling, i.e. $g = g'$.

The double-angled gradient vectors are then smoothed by a Gaussian

²The magnitude of the double-angled gradient vector used in [KW87] is squared in the angle-doubling process.

operator (G_2). Dividing the angle of the smoothed double-angled gradient vector by 2 will finally give a direction normal to the flow direction. Mathematically, the inclination of the flow direction is given by:

$$\phi = \frac{\pi}{2} + \frac{1}{2} \arctan \left(\frac{G_2 * g'_y}{G_2 * g'_x} \right) \quad (4.15)$$

This is a better estimation of the local flow direction.

The next measurement that should be computed is the local coherence of the flow field. Since the coherence (χ) of the flow field corresponds to the coherence of the gradient vectors in a small local area, it can be defined by:

$$\chi = \frac{\sqrt{(G_2 * g'_x)^2 + (G_2 * g'_y)^2}}{G_2 * g'} \quad (4.16)$$

(as suggested in [KW87]). This coherence value ranges from 0 (absence of any preferred orientation) to 1 (existence of a highly oriented pattern).

Plate 4.3b shows a coherence image. Its intensity at a point corresponds to the coherence of the flow field centered at the respective pixel in the original image (Plate 4.3a). Plate 4.3 c and d show the accumulation images formed by applying the infinite line accumulation (ILA) method in the field-oriented approach with the strength of the oriented pixel set to χ and 1, respectively.

Interpretation of the experimental results will be given in Section 4.4.3, where the field-oriented RSR approach is compared with other RSR approaches.

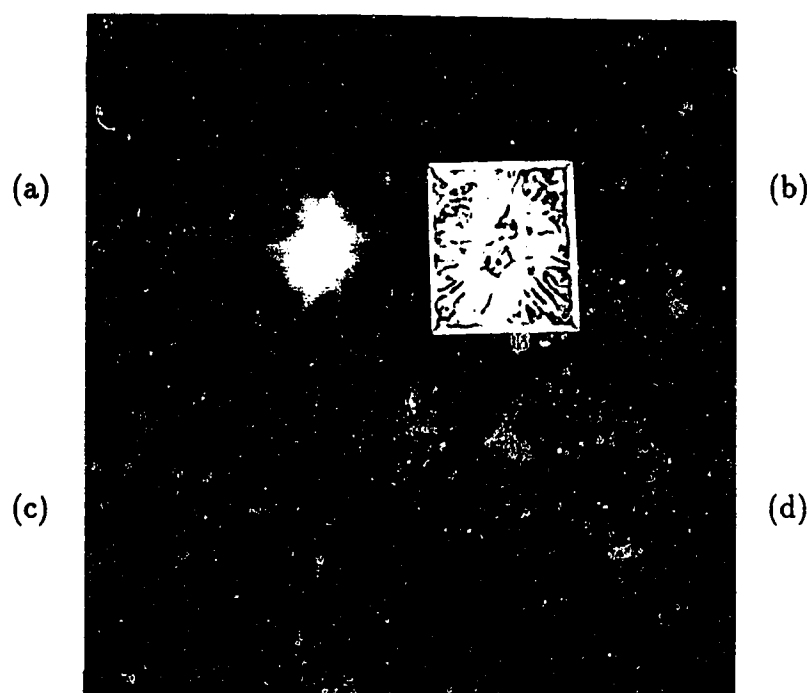


Plate 4.3: Result of Field-oriented RSR

- (a) original image
- (b) coherence image
- (c) accumulation image (strength=coherence)
- (d) accumulation image (strength=1)

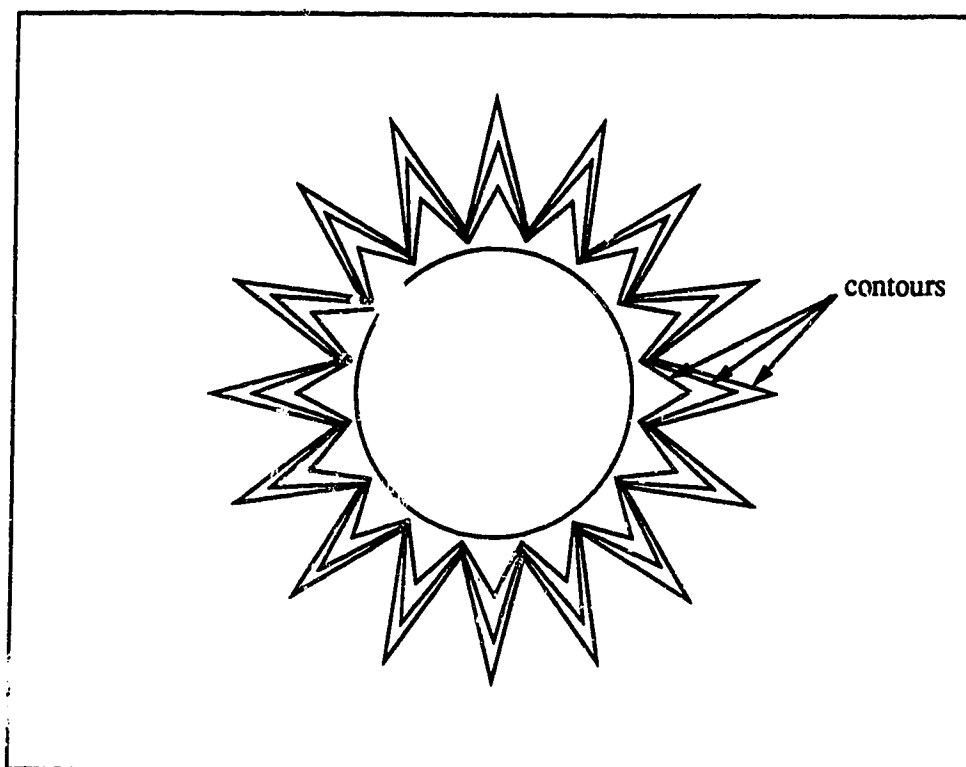


Figure 4.12: Ideal Radiating Structure for Spine-oriented Approach

4.4.2.3 Spine-oriented RSR Approach

The spine-oriented radial structure recognition (RSR) approach will conceive the radiating structure of a stellate tumor as a structure formed by a number of radially-arranged distinct spicules. Figure 4.12 shows an ideal case of such a radiating structure.

Each spicule, in turn, is viewed as a structure of a specific form. Figure 4.13 shows the fine structure of such a spicule. The central line of the spicule is called the *spine*. It determines the orientation of the spicule. The

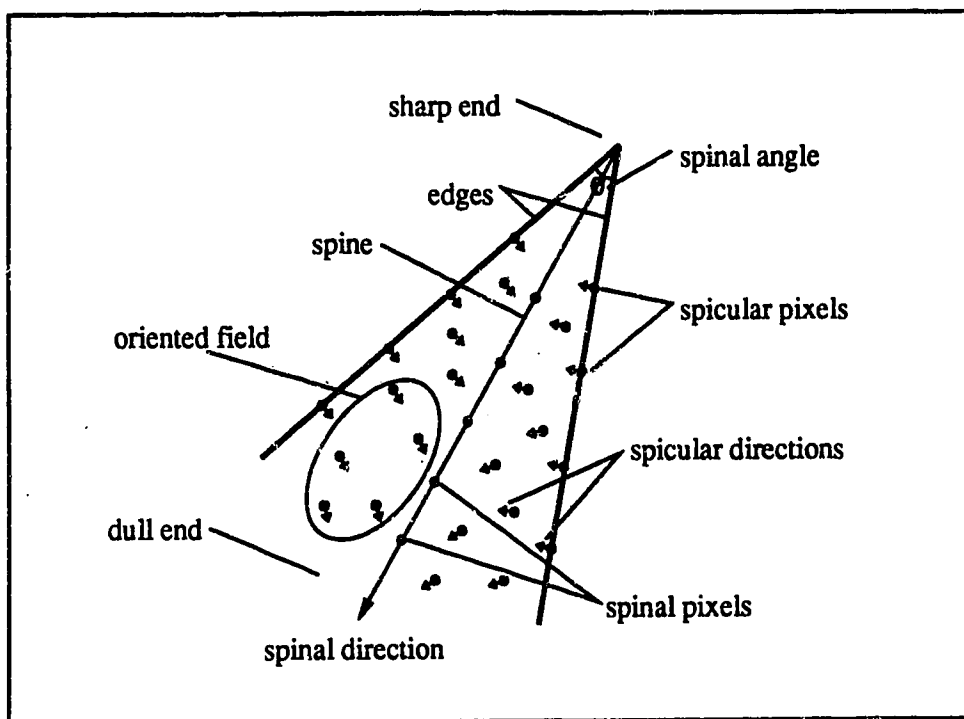


Figure 4.13: Fine Structure of a Spicule

direction along the spine which points to the dull end of the spicule is taken as the *spinal direction*. The reason why this direction is selected is simply that it points to the center of the radiating structure. The pixels lying on the spine are called *spinal pixels*. Their directions are defined to be the same as the spinal direction.

The pixels which are part of the spicule but not on the top of the spine are called *spicular pixels*. The intensities of the spicular pixels in a spicule are not all the same. They increase gradually towards the spine and the dull end of the spicule. This phenomenon is made even more apparent and dominant after smoothing (see Plate 4.1c, page 65). The direction of a spicular pixel, or simply called the *spicular direction*, is defined as the gradient direction at that pixel.

Since the spicular pixels in a small local area point almost to the same direction, they form a consistent oriented field. The field direction³ is given by the smoothed local spicular direction. If the spicular pixels are close to an edge of the spicule, the corresponding field direction will be roughly normal to the edge. The angle between the two edges of a spicule is called the *spinal angle*. It determines the acuteness of the spicule.

As it has been shown, the spicules in the radiating structure contain a lot of fine details. It will be very hard to encode all the fine details of the spicules

³The field direction here is different from the flow field direction defined in the previous section. They differ by $\pi/2$ radians.

in the RSR process. The problem is further aggravated by the fuzziness of the spicules and the existence of noise. In order to solve this problem, an abstract representation of a spicule is proposed in the RSR process.

The spine of a spicule may be the best candidate for abstraction. It represents the backbone of the spicule. The following four parameters capture most of the information that exists in the corresponding spicule.

- (a) **Location** — It defines the location of the spicule.
- (b) **Spinal Direction** — It gives the orientation of the spicule.
- (c) **Spinal Angle** — It specifies the acuteness of a spicule. The smaller this angle is, the more acute the spicule is. In an extreme case, a spicule may actually come from a straight line (see Figure 4.14). On the contrary, a very obtuse spicule may actually be part of a round object (see Figure 4.15).
- (d) **Spinal Strength** — It specifies the likelihood of the existence of a corresponding spicule. Its value depends on the coherence of the spicular directions of the spicular pixels on the two sides of the spine.

The spinal pixels are the basic elements of the spine and they inherit all the four characteristics (parameters) of a spine. However, the spinal angle of a spinal pixel will increase as it moves towards the dull end of the spicule.

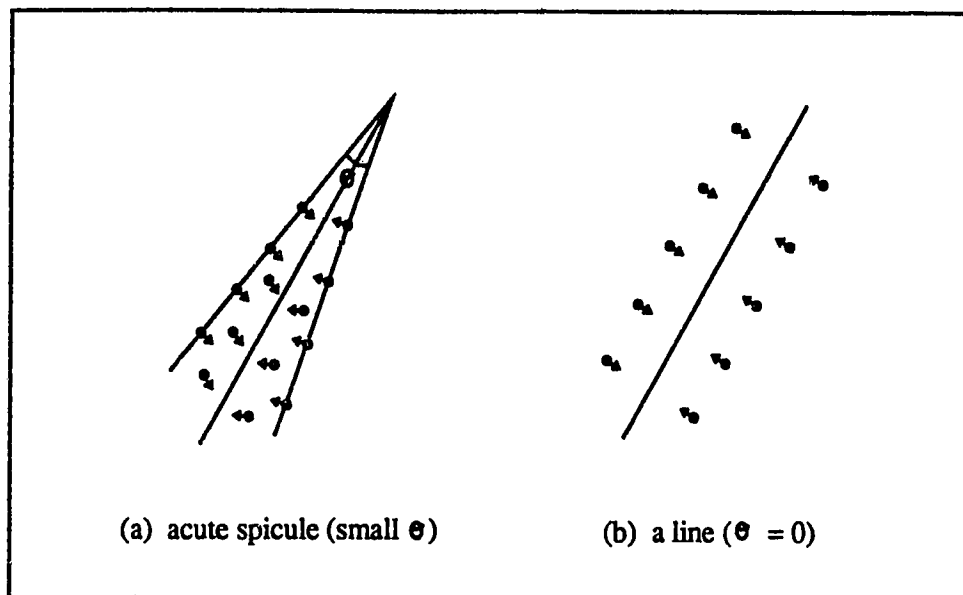


Figure 4.14: Examples of Acute Spicules

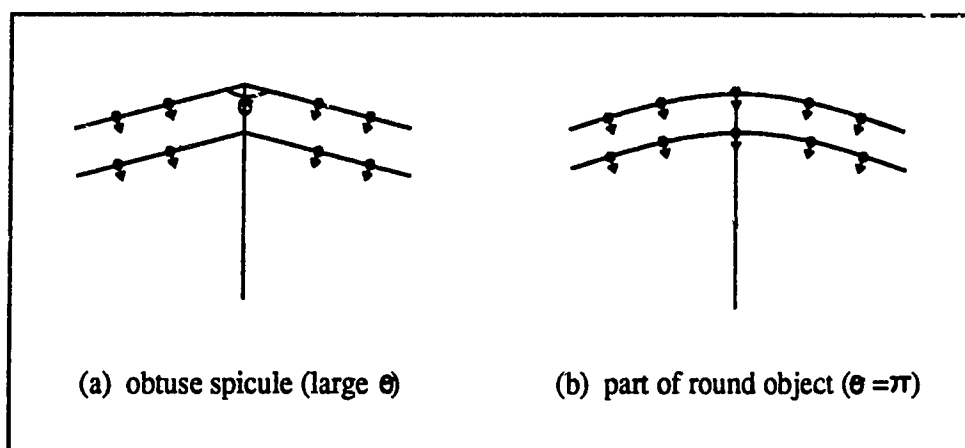


Figure 4.15: Examples of Obtuse Spicules

This can be observed from the contours of ideal radiating structure shown in Figure 4.12 (page 83).

In the spine-oriented RSR approach, conceptually all potential spines in an image should be located first, and then a high-level process is used to determine if a subset of the detected spines forms a radial pattern that in turn corresponds a radiating structure. However, the process to extract individual spines is complicated as it involves the formation of lines from scattered pixel points that satisfy the criteria of being part of the spine to be formed. Due to noise and anatomical structures of the human breast, there is hardly any complete spines. Most of them may be broken. The high-level process which is responsible for the identification of radial patterns will add another complication into the entire RSR process. Therefore, in practice individual spinal pixels will be used directly in the spine-oriented RSR.

When the spinal pixels detected in a breast image are taken as the oriented pixels in the infinite line accumulation (ILA) process (see Figure 4.16), an accumulation image will be formed. Any high-intensity spot which exists in the accumulation image would indicate the existence of a radiating structure and the location of the bright spot would represent the center of the radiating structure.

However, there can be some modifications (or improvements) to the ILA technique when it is used in the spine-oriented RSR. The modifications include:

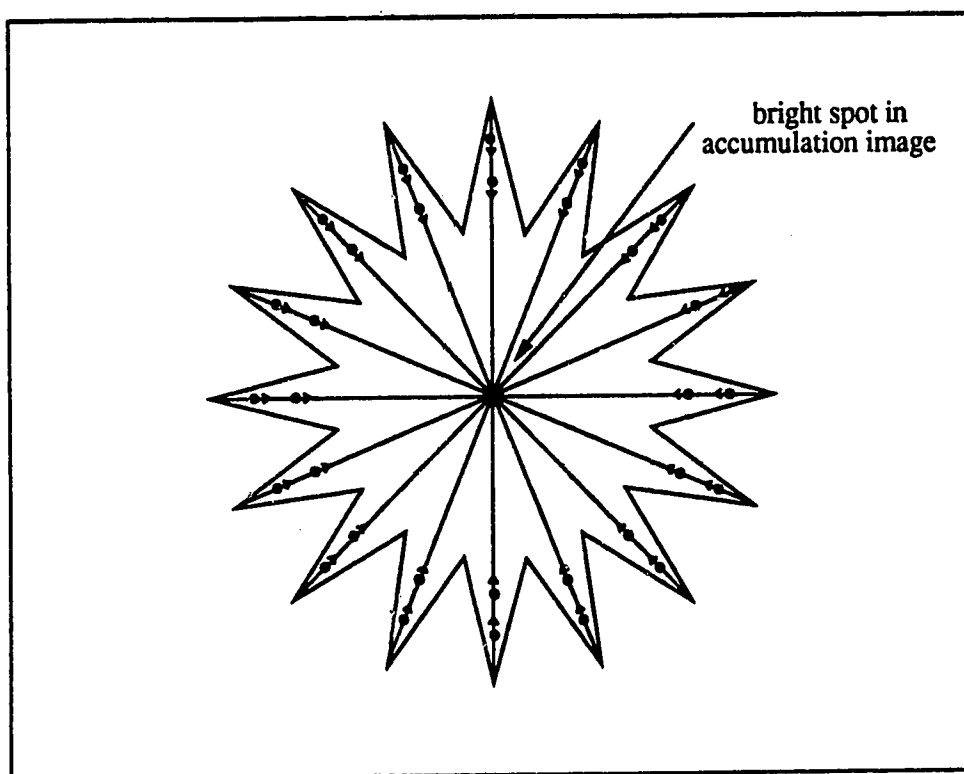


Figure 4.16: ILA in Spine-oriented RSR

- Not all the detected spinal pixels are used in the ILA process. Their spinal angles should be smaller than a preset threshold θ_0 . This inhibits some very obtuse “spicules”, which may come from some round objects, to affect the result.
- The infinite lines drawn on the accumulation image in the ILA process are no longer bidirectional. They start from the spinal pixels and extend only in spinal directions.
- The lines no longer extend without limit (or control). They stop when the intensity level along the extended line in the Gaussian smoothed image drops more than a value of 2 from the current maximum intensity of the line. This is because the intensity of a point on the extended line should increase when it moves towards the center of a radiating structure.

The discussion of the spine-oriented RSR is mainly focused so far on the overview of the entire process. We now give the details on how to detect the spinal pixels in a breast image and how to measure the four parameters of a spinal pixel.

Measurement of Spinal Parameters

The original breast image I is first convolved with the Gaussian operator (G_1) to remove the high frequency noise and to enhance the image. The

resultant image is given by $G_1 * I$. Plate 4.1c (page 65) shows the pseudo-colored Gaussian image of Plate 4.4a (page 100). (Plates 4.1a and 4.4a are the same.) It can be seen that the structure of the spicules in the smoothed image is very clear.

The gradient at any point P in the image is given by $\vec{g}(P) = (g_x(P), g_y(P))$ and its strength is given by $|\vec{g}(P)|$, where

$$g_x = \frac{\partial}{\partial x}(G * I) = \left(\frac{\partial G}{\partial x} \right) * I \quad (4.17)$$

$$g_y = \frac{\partial}{\partial y}(G * I) = \left(\frac{\partial G}{\partial y} \right) * I \quad (4.18)$$

$$|\vec{g}| = \sqrt{g_x^2 + g_y^2} \quad (4.19)$$

Since a spinal pixel is a pixel lying on the spine of a spicule and there are consistent fields on both sides of a spine, a target pixel qualified as a spinal pixel should have corresponding consistent fields nearby. Further, the fields should have appropriate directions (i.e. they cannot point outward). Figure 4.17a shows a spinal pixel and its consistent fields nearby.

Figure 4.17b shows a target pixel T which is tested if it is a spinal pixel. N_1, N_2, N_3 and N_4 are four nearby pixels of T . Their respective locations are as in the figure. By applying the same coherence formula 4.16 given in page 81, the coherence $\chi(P)$ of a field centered at pixel P can be computed. However, no angle doubling is applied here as the spine-oriented approach makes different assumptions about the pattern of the radiating structure.

The small area centered at the target pixel T is divided into four quad-

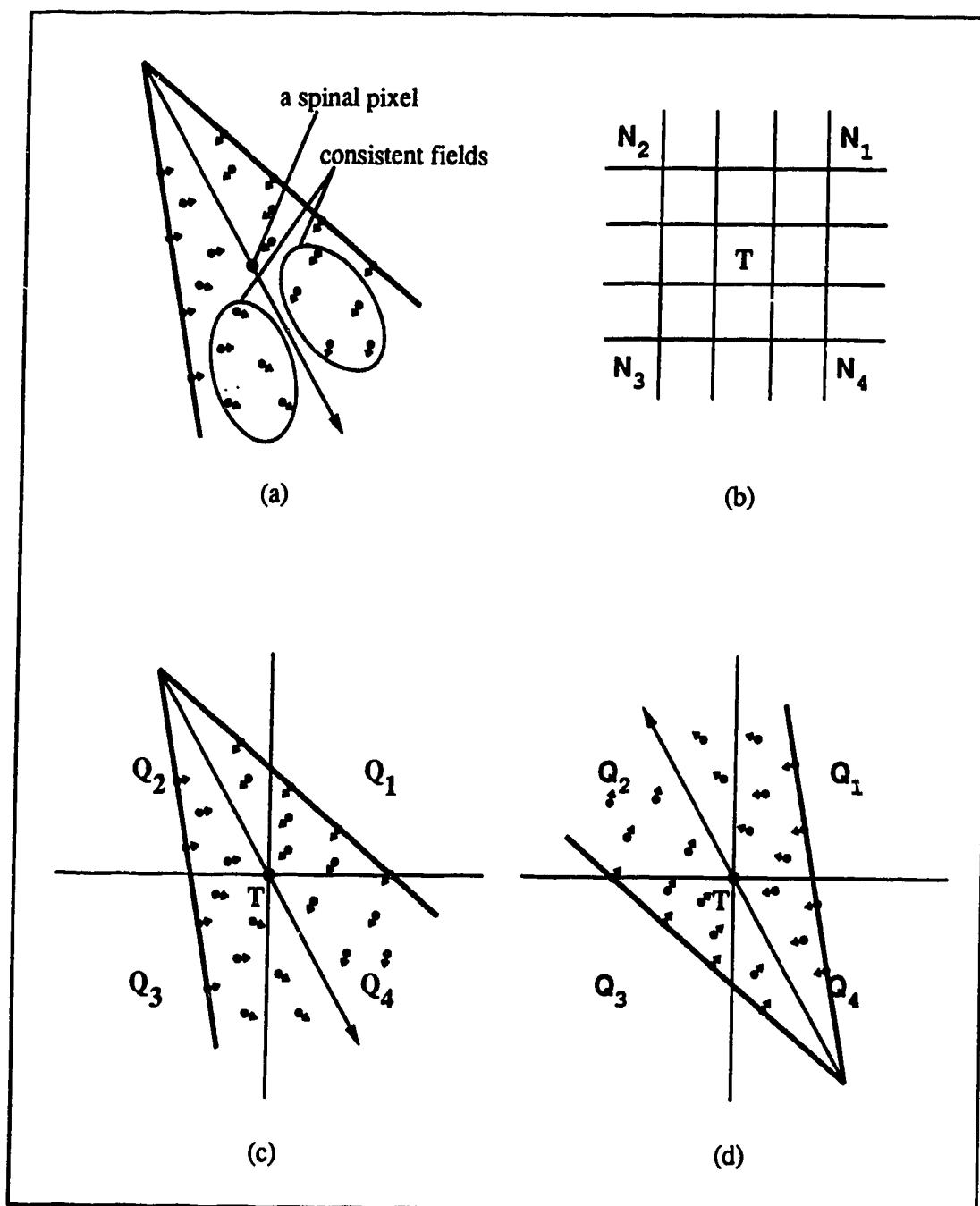


Figure 4.17: Detection of Spinal Pixel

rants Q_1 – Q_4 . If a spine exists, crosses T and falls in quadrants Q_2 and Q_4 , then the coherence of the fields in Q_1 and Q_3 should be higher than those in Q_2 and Q_4 (see Figure 4.17 c and d). This is because the coherence of the area very close to the spine would have a smaller value. By comparing the product of $\chi(N_1)$ and $\chi(N_3)$ with that of $\chi(N_2)$ and $\chi(N_4)$, one can tell which pair of quadrants the spine falls in.

Without loss of generality, $\chi(N_1) \cdot \chi(N_3)$ is assumed to be higher than $\chi(N_2) \cdot \chi(N_4)$. (i.e. the spine is assumed to lie in quadrants Q_2 and Q_4 .) When compared with the smoothed gradients measured at N_2 and N_4 , those measured at N_1 and N_3 will be better estimations of the directions of the fields close to T. For simplicity, the pixels N_1 and N_3 will be referred to as A and B in the following discussion, and their corresponding smoothed gradients (i.e. $\vec{J}(N_1)$ and $\vec{J}(N_3)$) as $\vec{a} = (a_x, a_y)$ and $\vec{b} = (b_x, b_y)$, respectively. The angles that they make with the x -axis are given by θ_a and θ_b .

When the gradient vectors \vec{a} and \vec{b} are taken as the spicular directions of the fields on the two sides of the spine, the spinal direction, say $\vec{c} = (c_x, c_y)$, can be computed as: (see Figure 4.18)

$$\begin{aligned}\vec{c} &= \frac{\vec{a}}{|\vec{a}|} + \frac{\vec{b}}{|\vec{b}|} \\ &= \left(\frac{a_x}{|\vec{a}|} + \frac{b_x}{|\vec{b}|}, \frac{a_y}{|\vec{a}|} + \frac{b_y}{|\vec{b}|} \right)\end{aligned}\quad (4.20)$$

The next step is to test if the gradient vectors \vec{a} and \vec{b} will lead to some contradicting situations when the spine is supposed to lie in Q_2 and Q_4 . A

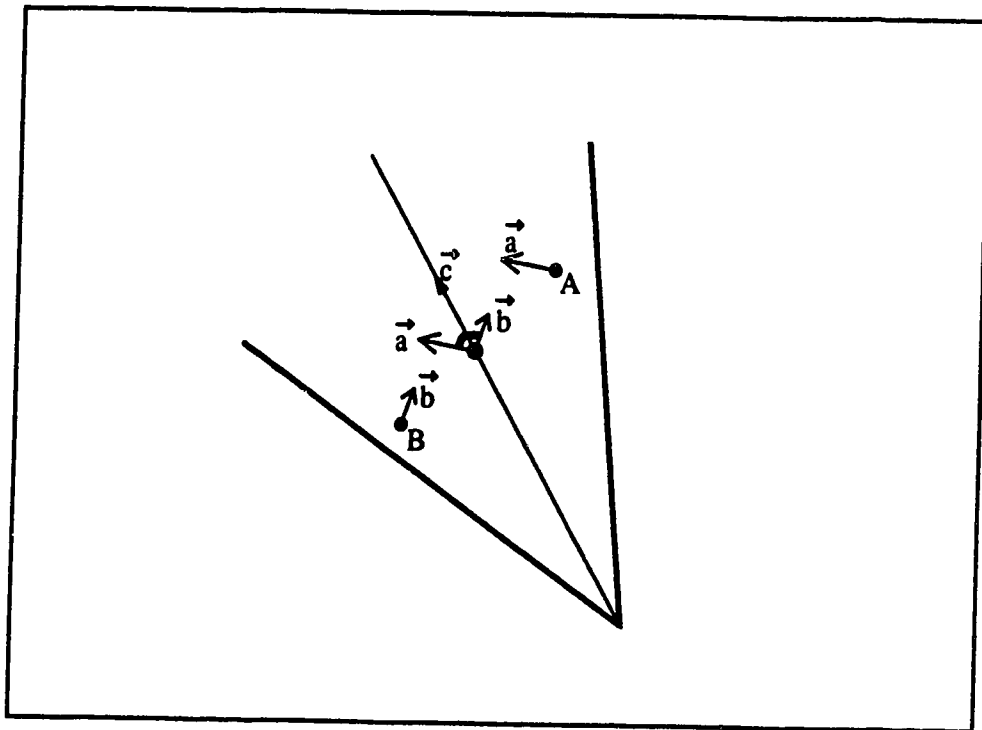


Figure 4.18: Spinal Direction Measurement

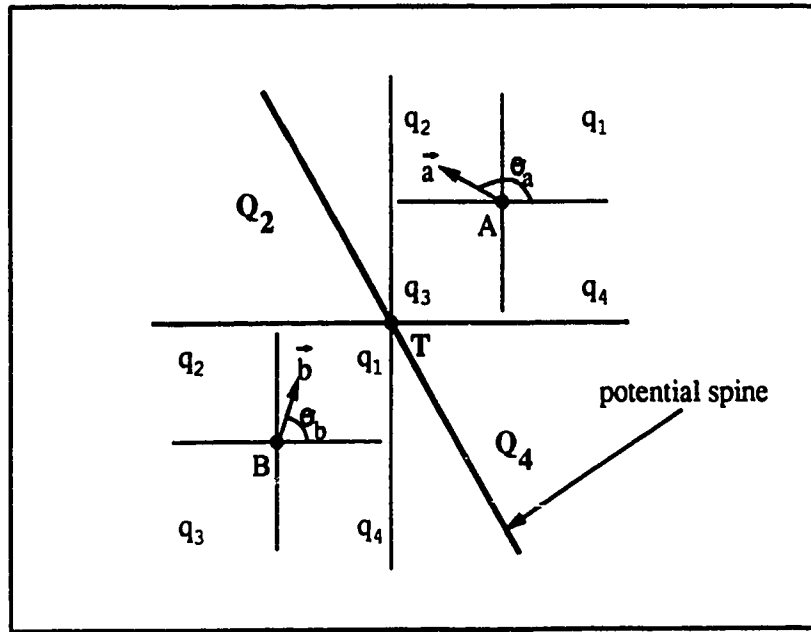


Figure 4.19: Consistency Test for a Potential Spinal Pixel

small area centered at A (and B) is divided into four sub-quadrants q_1 - q_4 (see Figure 4.19). In order to be consistent with the orientation of the spines, \vec{a} and \vec{b} should satisfy the following two conditions:

(I) $\vec{a} \notin q_1$ and $\vec{b} \notin q_3$

(“ $\vec{a} \notin q_1$ ” means “vector \vec{a} does not point to quadrant q_1 ”.)

(II) If $\vec{a} \in q_2$, $\vec{b} \in q_1$ then ($\vec{c} \notin q_1$ or tolerance test passes)
 $\vec{b} \in q_2$ then $\theta_a > \theta_b$ (i.e. $a_x b_y - a_y b_x < 0$)
 $\vec{b} \in q_4$ then fail

If $\vec{a} \in q_3$, $\vec{b} \in q_1$ then pass
 $\vec{b} \in q_2$ then ($\vec{c} \notin q_3$ or tolerance test passes)
 $\vec{b} \in q_4$ then ($\vec{c} \notin q_3$ or tolerance test passes)

If $\vec{a} \in q_4$, $\vec{b} \in q_1$ then ($\vec{c} \notin q_1$ or tolerance test passes)
 $\vec{b} \in q_2$ then fail
 $\vec{b} \in q_4$ then $\theta_a < \theta_b$ (i.e. $a_x b_y - a_y b_x > 0$)

The tolerance test is considered as passed if and only if $(\left|\frac{c_y}{c_x}\right| < \tan \alpha)$ or $(\left|\frac{c_x}{c_y}\right| < \tan \alpha)$, where α is a preset threshold (e.g. 10°). This test is needed when the orientation of the spine is close to either horizontal or vertical.

When the spine is horizontal (or vertical), the accuracy of the spinal direction \vec{c} will decrease. This is because the spine will be closer to pixel B (or A) and affect the field direction \vec{b} (or \vec{a}) measured at B (or A). To improve the accuracy of the measurement of the spinal direction in general, new locations of A and B are computed relative to the current spinal direction \vec{c} and the target pixel T. Let A' and B' be the new locations of A and B respectively. The line joining A' and B' should be normal to \vec{c} , and both A' and B' are

$2\sqrt{2}$ units away from T (see Figure 4.20). By taking the notation (x_p, y_p) representing the coordinate of any point P in the image, the new A' and B' can be obtained by:

$$(x_{A'}, y_{A'}) = \left(x_T + 2\sqrt{2} \cdot \frac{c_y}{|\vec{c}|}, y_T - 2\sqrt{2} \cdot \frac{c_x}{|\vec{c}|} \right) \quad (4.21)$$

$$(x_{B'}, y_{B'}) = \left(x_T - 2\sqrt{2} \cdot \frac{c_y}{|\vec{c}|}, y_T + 2\sqrt{2} \cdot \frac{c_x}{|\vec{c}|} \right) \quad (4.22)$$

The new spinal direction \vec{c}' is given by:

$$\vec{c}' = \frac{\vec{a}'}{|\vec{a}'|} + \frac{\vec{b}'}{|\vec{b}'|} \quad (4.23)$$

It should be noted that the lines XY and YZ in Figure 4.20 may not necessarily be the two real edges of the spicule. They are simply lines perpendicular to \vec{b}' and \vec{a}' , respectively. The angle (θ) between them is the spinal angle of the potential spinal pixel T. This angle may range from the spinal angle of the spine to π radians, depending on the location of T on the spine. As T moves towards the dull end of the spicule, the corresponding spinal angle will increase.

If the target pixel T is really a spinal pixel, the old and the new spinal directions (\vec{c} and \vec{c}') should be close. As another consistency test for being a spinal pixel, the angle between \vec{c} and \vec{c}' should be less than a preset threshold β (e.g. 15°). That is the following inequality should be satisfied:

$$\frac{|c_x c'_y - c_y c'_x|}{|\vec{c}| \cdot |\vec{c}'|} < \sin \beta \quad (4.24)$$

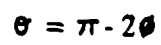


Figure 4.20: Measurement of New Spinal Direction

It should be noted that $\vec{c} = (c_x, c_y)$, $\vec{c}' = (c'_x, c'_y)$ and the cross product equality ($\vec{u} \times \vec{v} = |\vec{u}| \cdot |\vec{v}| \cdot \sin \theta \cdot \hat{n}$) are used in deriving the inequality above.

If a pixel in the given image passes all the consistency tests described above, it will be taken as a spinal pixel and used in the later stages of the spine-oriented RSR (as it is described in the beginning of this section). The four parameters of a spinal pixel are given as follows:

- (a) The location is given by the image coordinate of the detected spinal pixel.
- (b) The spinal direction is given by \vec{c}' .
- (c) The spinal angle θ is obtained by computing: (see Figure 4.20)

$$\theta = \pi - \arcsin \left[\frac{|\vec{a}' \times \vec{b}'|}{|\vec{a}'| \cdot |\vec{b}'|} \right] \quad (4.25)$$

- (d) The spinal strength can be defined as the product of the coherence of the fields centered at A' and B' , that is $\chi(A') \cdot \chi(B')$. However, a later experiment shows that there is no significant difference when the spinal strength is set to a fixed value, say 1.

Plate 4.4b shows the image formed by overlaying the (red-colored) spinal pixels, detected by the method described above, on the original breast image (Plate 4.4a). Plate 4.4 c and d shows the accumulation images generated by employing the spinal pixels as oriented pixels on the modified infinite line

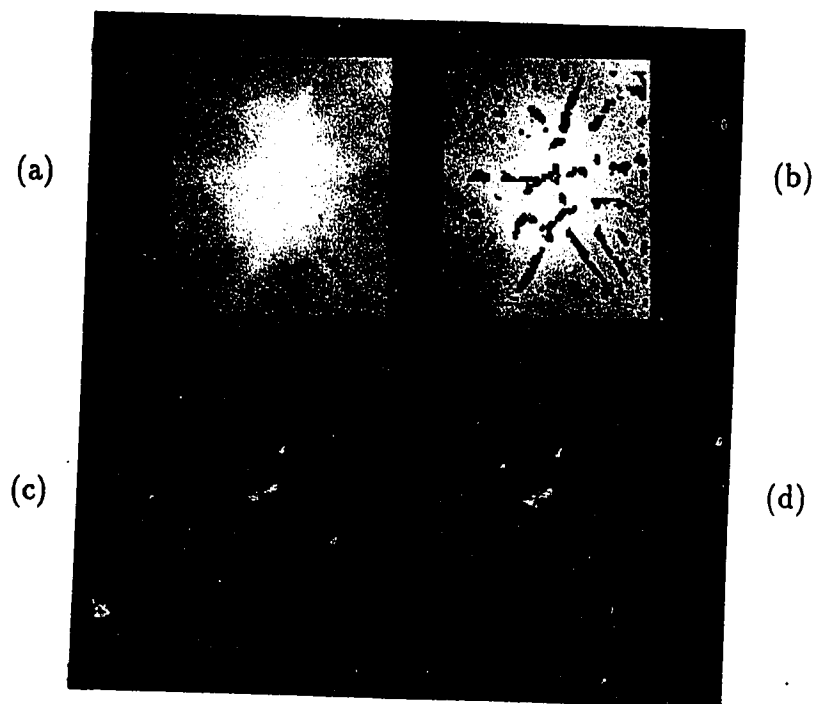


Plate 4.4: Result of Spine-oriented RSR

- (a) original image
- (b) overlayed with red-colored spinal pixels
- (c) accumulation image (strength=coherence)
- (d) accumulation image (strength=1)

accumulation (ILA) process, when the product of coherence and value of one are taken as the strength of a spicule, respectively.

Interpretation of the experimental results (Plate 4.4 c and d) will be given in the next section, where the spine-oriented RSR approach is compared with other RSR approaches.

4.4.3 Comparison among Different RSR Approaches

Three different radial structure recognition (RSR) approaches have been introduced in the previous sections. They are the edge-oriented approach, the field-oriented approach and the spine-oriented approach. They put their emphases on different aspects of the spicules which form the radiating structure. The edge-oriented approach focuses on the edges of a spicule. The field-oriented approach takes the spicules collectively as a field that flows into the center of the radiating structure. The spine-oriented approach looks at individual spicules and focuses on the spine (central line) of a spicule. Together with different assumptions about the exact pattern of an ideal radiating structure, they apply the infinite line accumulation (ILA) technique to detect the radiating structure as high-intensity area in the accumulation image.

In this section, the effectiveness and the robustness of the three RSR approaches will be compared. The best approach selected will be fine-tuned and

employed in the RSR process of the entire stellate tumor detection system.

Since the accumulation images formed in each of the three RSR approaches have the same basic structure regardless the selection of the strength definition, only the accumulation image formed with the strength of the oriented pixels set to the value one is taken for comparison.

Theoretically, a bright spot should appear at the center of each accumulation image since there is a radiating structure which exists and is located at the central part of the corresponding original images (Plates 4.2a, 4.3a and 4.4a in pages 76, 82 and 100 are all the same). However, instead of bright central spots, there are dim annular areas in the accumulation images (Plates 4.2d and 4.3d) formed by the edge-oriented process and the field-oriented process. The existence of the annulus indicates that the infinite lines used in the accumulation process do not point directly to the center of the radiating structure. The reason of the inappropriate line directions is probably caused by incorrect assumptions about the shape of the spicules.

In the edge-oriented approach, spicules are assumed to be very sharp and their edge contour points directly to the center of the radiating structure. In the field-oriented approach, spicules are assumed to be fine lines which collectively flow to the center of the radiating structure. If the assumed “very sharp spicules” or “fine lines” are in fact obtuse spicules (Figure 4.15a, page 87), the line directions computed in the edge-oriented and the field-oriented approaches will certainly deviate a lot from the true radial direction. Inspection

of the original test image by employing the pseudo-coloring technique (see Plate 4.1, page 65) shows that some of the spicules in the radiating structure are obtuse. This explains why the edge-oriented and the field-oriented approaches fail to produce a bright spot at the center of the accumulation image.

The spine-oriented approach, on the other hand, does produce the expected result. The following analysis suggests that it has advantages over the other two approaches:

- (1) It is more robust. Since the spine of a spicule is located at the central line of a spicule, the spinal direction is unaffected by the acuteness of the spicule. Therefore, obtuse spicules have no severe adverse effect on the spine-oriented approach.
- (2) When compared with the line directions measured in the other two approaches, the spinal direction computed in the spine-oriented approach is a better estimate of the true radial direction of the radiating structure since it represents the direction of the central line (rather than the edges) of a spicule. It has a higher chance to point right at the center of the structure.
- (3) The extension of the (infinite) line is under better control. The “infinite lines” drawn on the accumulation image can no longer be extended without limit in both directions. It can only be extended in the direc-

tion of an increase in intensity and stops when the intensity along the line in the smoothed image begins to drop.

In addition to the advantages stated above, the spine-oriented approach also provides an effective way to remove the spicules in the original breast image (which is not available in the other two approaches). The details of this spicule removal process will be given in the next section. Once the spicules in the breast image have been removed, the central mass detection (CMD) process can be applied to detect the bright central tumor mass of the stellate tumor and to form an initial list of suspicious tumors. The reason why the initial list is first formed by the CMD process instead of the RSR process is that when compared with a radiating structure (or the corresponding bright spot in the accumulation image), the bright central tumor mass of the stellate tumor is easier to detect. After the initial list is formed, the spine-oriented RSR process can be applied again to measure the *spicularity* of each suspicious tumor. Spicularity is a measure of the likelihood of the existence of a surrounding radiating structure. The details about the spicularity measurement will be given in Section 4.4.5. According to this spicularity measure, stellate tumors can be differentiated from non-stellate tumors in the initial tumor list.

4.4.4 Spicule Removal

Spicule removal is a filtering process specially designed for smearing the spicules of stellate tumors in breast images. Since some anatomical structures such as the ducts in the breasts have a similar appearance as a spicule, they will be smeared indiscriminately. Fortunately, this would not cause any problem. On the contrary, such anatomical structures reduce the efficiency of the tumor detection process and it is thus desirable to have them removed. In addition to spicule-smearing, the spicule removal process will have a smoothing effect on the image as it will also remove some other spicule-like structures and noise.

The spicule removal process is obtained through application of a spicule-smearing filter. This filter is responsible for smoothing out the spicules in the image. The application of this filter is location-specific and orientation-specific. It is only applied to image positions where there is a spicule and at an orientation matching the orientation of the spicule. This allows effective spicule-smearing and reduces the chance of incurring undesirable noise. Since both location and orientation are known for spinal pixels in a spicule, the spicule-smearing filter can be applied to the spinal pixels as shown in Figure 4.21.

The design of the spicule-smearing filter is shown in Figure 4.22. It can be considered as a modification to the directional selective median filter.

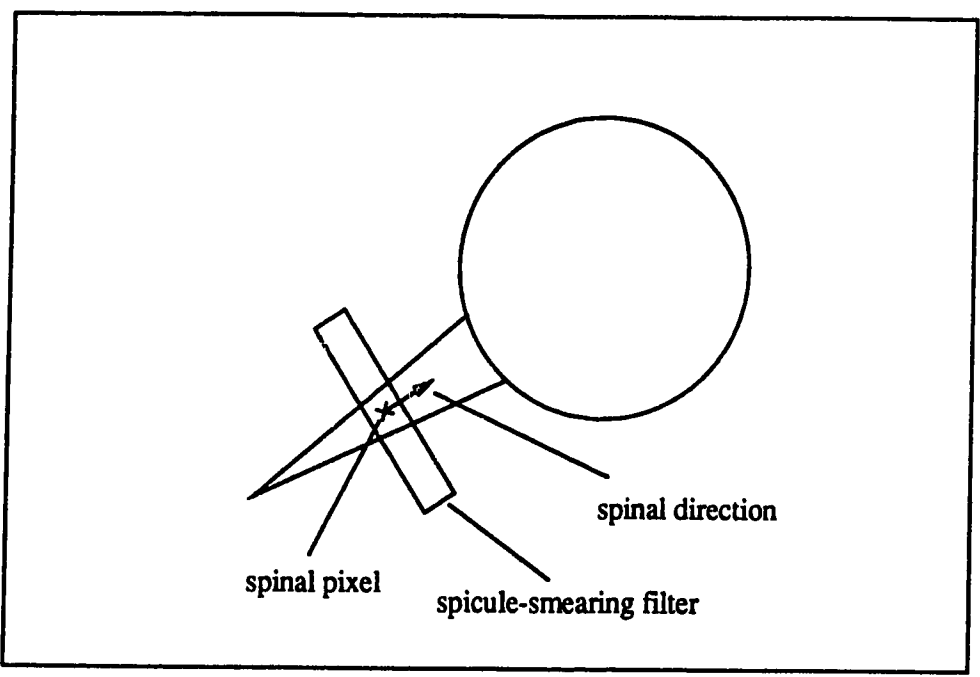


Figure 4.21: Spicule Removal

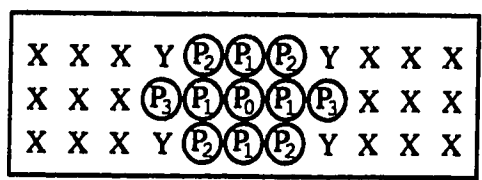


Figure 4.22: Spicule-smearing Filter

The principal axis of the filter is oriented normal to the spinal direction of the target spinal pixel at P_0 . When the filter is applied, the new grey level intensity at P_0 will be given by the median of the intensities of the 18 pixels marked with X in the filter window (Figure 4.22), provided that the new grey level intensity is lower. Otherwise, it is kept unchanged. This is because the intensity of a spinal pixel normally should be higher than that of the nearby pixels.

The pixels marked X are selected and used in computing the new grey level value of the spinal pixel because they are at a distance neither too close nor too far from the spine of the spicule and may be good to represent the background of the spicule.

The filter having its orientation fixed will then be applied to the neighbors (marked P_i ($i = 1, 2, 3$) in the filter window, see Figure 4.22) of P_0 , provided that these neighbors have not been marked as P_j , where $j < i$ (i.e. provided that these neighbors are not closer to another spinal pixel).

The resultant image formed will be spicule-free as all the spicules in the image have been smoothed out by assigning them the intensity of the nearby background.

4.4.5 Spicularity Measurement

After the spicules in the normalized breast images have been removed, the spicule-free image will be passed to the central mass detection (CMD) process (which will be discussed in Section 4.5) and an initial list of suspicious tumors will be formed. The suspicious tumors in this list may not necessarily be stellate tumors. They may actually be circumscribed tumors because circumscribed tumors appear as bright objects of circular or oval shape in both the normalized and the spicule-free breast images.

Spicularity is a measure designed to differentiate between stellate tumors and circumscribed tumors. It measures the likelihood of the existence of a radiating structure surrounding a suspicious tumor. Since a radiating structure is formed by a number radially arranged spicules, spicularity can be defined as the number of spicules (or the number of spines) connected to suspicious tumors in a normalized image. However, it is too hard to say definitely if a spicule (or spine) exists. Instead, its existence should be expressed as a likelihood. The number of spinal pixels can be a good candidate for this likelihood measure. Therefore, a new definition of the spicularity of a suspicious tumor can be defined as the number of spinal pixels “connected” to the suspicious tumor. A spinal pixel is considered to be connected to a tumor if the following two criteria are satisfied:

- (1) The spinal pixel is not located inside any suspicious tumor.

- (2) When a line starts at the spinal pixel and extends in the spinal direction, the line should meet the tumor. Besides, there should be no significant drop on intensity along the extended line in the Gaussian smoothed version ($G_1 * I$) of the normalized image before it meets the tumor.

Since the spicules tend to be longer in a large stellate tumor and there is a higher chance for a large tumor to catch some scattered spinal pixels, the spicularity measurement should be normalized by the radius of the tumor and defined as:

$$\text{Spicularity} = \frac{\text{number of spinal pixels connected to a tumor}}{\text{radius of the tumor}} \quad (4.26)$$

It should be noted that the uniformity of the spinal pixel distribution has not been taken into account for computing the spicularity even though the spicules in a stellate tumor are often evenly distributed. This is because the noise or “pseudo-spicules” are often evenly distributed around a tumor and sometimes the spicules of a genuine stellate tumor in certain orientations may be totally obscured by some anatomical structures of the breast. Therefore, the uniformity factor is not included in the definition of the spicularity.

The spicularity measurement process first forms four images:

- **Spinal direction images** — They are a pair of images whose elements are the x and y components of the spinal directions for spinal pixels.

For non-spinal pixels, the corresponding elements in the spinal direction images will be set to zero.

- **Gaussian smoothed image** — This image is formed by smoothing a normalized breast image with a Gaussian operator.
- **Tumor image** — This image is formed by firstly initializing all its elements to 0. For each entry in the initial tumor list, a circle of value i is then drawn on this tumor image, where i is the index in the tumor list. The index starts with a value one and is incremented by one for each successive tumor.

For each spinal pixel, a line is then drawn passing through it and extended in the spinal direction. If the intensity along the line in the Gaussian smoothed image does not drop significantly before the line meets a tumor in the tumor image, the spinal pixel count of the tumor is incremented by one. Figure 4.23 outlines the process diagrammatically. Finally, the spicularity of a tumor is given by the quotient of the spinal pixel count and the radius of the tumor.

The spicularity value computed in this way can then be used to differentiate a stellate tumor from a circumscribed tumor, or to measure the star-likeness of a tumor.

When compared with the spicularity measure suggested in [AG72] and the star-likeness measure suggested in [HSAA79] (see Sections 3.2.2.1 and 3.2.3.2

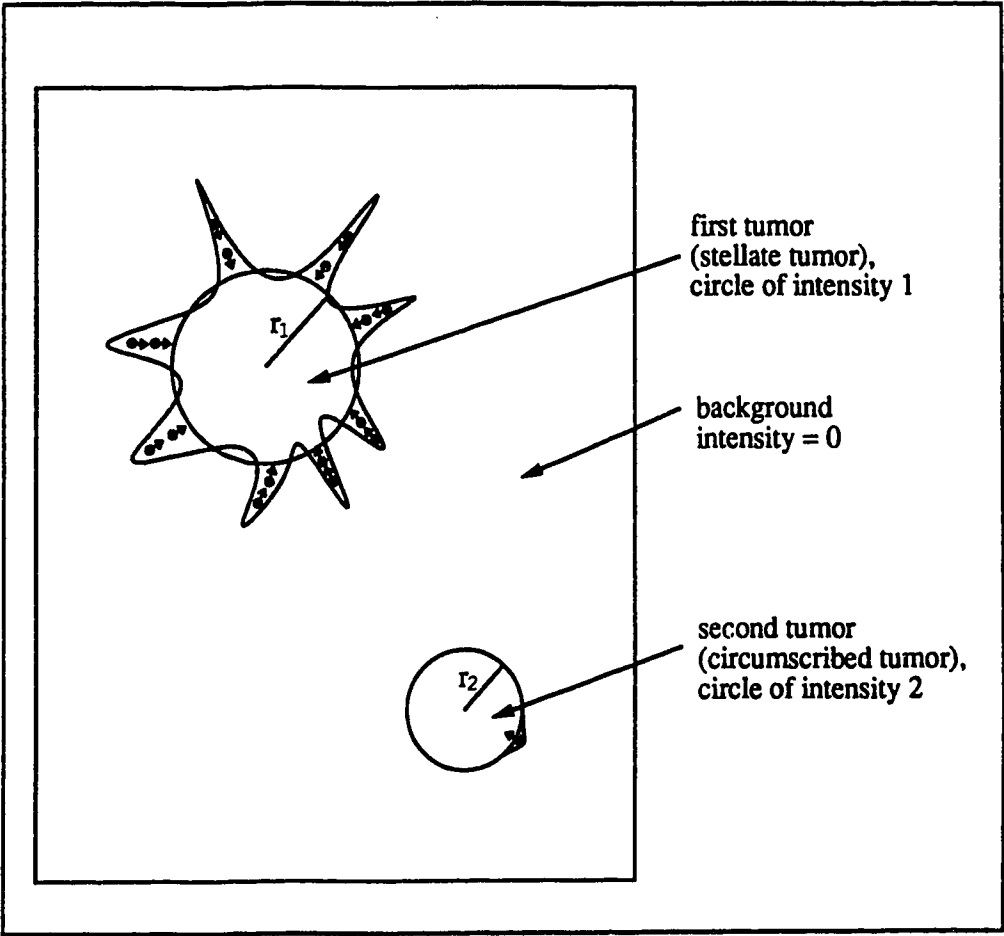


Figure 4.23: Spinal Pixel Accumulation

of this thesis), the spicularity measure discussed in this section is considered to be superior. This is because it has the following two advantages over the other two measures:

- (1) In computing the spicularity of a tumor, it takes the spinal pixels to represent the spicules of a tumor and ignores all irrelevant pixels in the image. However, the other two measures employ all pixels in the tumor area and its surroundings indiscriminately. The irrelevant pixels will then introduce additional noise to the measurement.
- (2) Since spinal pixels point to the center of a tumor, our spicularity measure can directly count the spinal pixels connected to the tumor. The other two measures either have to take the projection of the edge contour on a range of potential radial directions or assume the direction of the edge contour to be the true direction of a spicule.

4.5 Central Mass Detection

A stellate tumor appears in a breast image as an object composed of two components: the radiating structure and the central tumor mass. The radiating structure of the stellate tumor is removed in the spicule removal process. The stellate tumor left in the resultant spicule-free breast image will then contain the central tumor mass only. The central mass detection process described in this section is designed to detect these tumor masses and to form an initial list of suspicious tumors, specifying both location and size of each suspicious tumor.

The tumor mass of a stellate tumor left in the spicule-free image has the following features:

- (1) The tumor mass appears as a bright object in a comparatively darker background.
- (2) The shape of the tumor mass is approximately circular as the surrounding radiating structure has been removed.
- (3) The boundary of the tumor mass is very fuzzy because it is the place where the spicules of the tumor attach to the tumor. After spicule removal, it at best leaves a barely perceivable boundary.

The appearance of the central tumor mass of a stellate tumor in a spicule-free image is similar to that of a malignant circumscribed tumor. They both

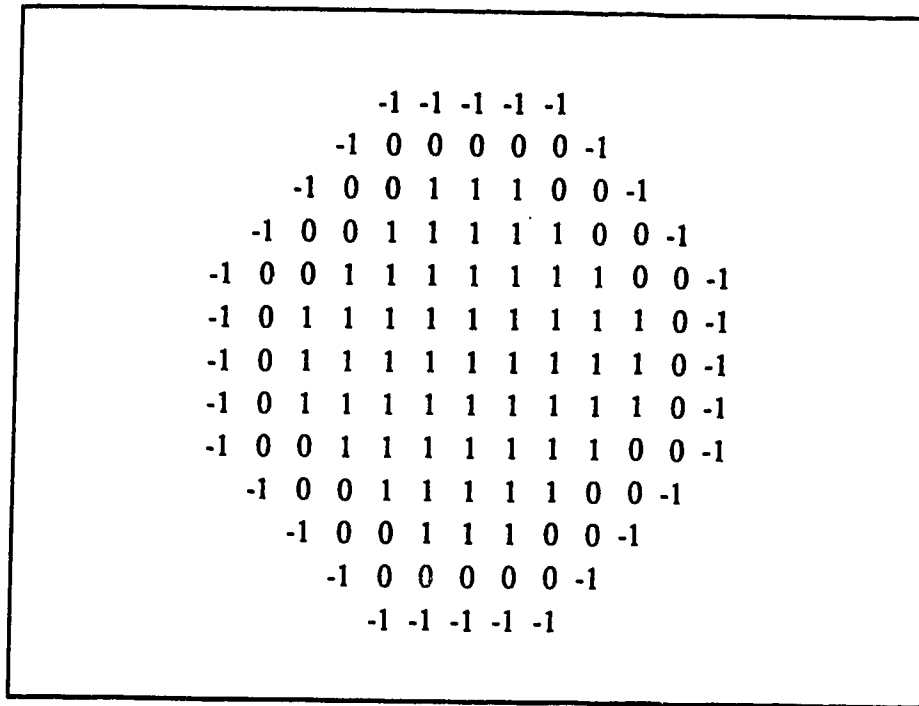


Figure 4.24: A Tumor-like Template

appear as an approximately circular, bright object with a fuzzy periphery. Owing to these similarities, the techniques used in detecting a circumscribed tumor are also applicable to the detection of the central tumor mass of a stellate tumor in the spicule-free image. Since the template matching technique has been shown to be effective in detecting the circumscribed tumor (see [Lai88]), it is adopted here to detect the central tumor mass of the stellate tumor.

Templates such as the one shown in Figure 4.24 are used to detect central tumor masses (see [Lai88]). The size of a tumor template shown in

Figure 4.24 is fixed. However, the size of a tumor mass in real life may vary from case to case and from stage to stage. To circumvent this problem, multiple templates can be used, with each of them representing a tumor of a particular size.

Likely tumor positions are given by the peaks in the (normalized) cross-correlation of a template and the image. Let F be the data image, and W be the circular tumor-like template. The center of the template is defined as the origin $(0,0)$ and all other pixels in the template are referenced relative to it. The normalized cross-correlation at a point (x, y) in the data image is given by:

$$R(x, y) = \frac{\sum_{(i,j) \in T} [(W(i, j) - \mu_w)(F(x + i, y + j) - \mu_F)]}{\sqrt{\sum_{(i,j) \in T} (W(i, j) - \mu_w)^2 \sum_{(i,j) \in T} (F(x + i, y + j) - \mu_F)^2}} \quad (4.27)$$

where $T = \{(i, j): W(i, j) \text{ is a valid reference in the template}\}$

$$\mu_w = \frac{\sum_{(i,j) \in T} W(i, j)}{N}$$

$$\mu_F = \frac{\sum_{(i,j) \in T} F(i, j)}{N}$$

$$N = \sum_{(i,j) \in T} 1$$

After a few simple algebraic operations, the normalized cross-correlation can be rewritten as:

$$R(x, y) = \frac{S_{wF} - C_1 S_F}{C_2 \sqrt{S_{FF} - S_F^2/N}} \quad (4.28)$$

$$\begin{aligned}
\text{where } S_{WF} &= \sum_{(i,j) \in T} W(i,j)F(x+i, y+j) \\
S_{FF} &= \sum_{(i,j) \in T} F^2(x+i, y+j) \\
S_F &= \sum_{(i,j) \in T} F(x+i, y+j) \\
C_1 &= \mu_w \\
C_2 &= \sqrt{\sum_{(i,j) \in T} W^2(i,j) - N\mu_w^2}
\end{aligned}$$

Since the computation of the normalized cross-correlation is expensive, it is important to realize that the computations of S_{WF} , S_{FF} and S_F can be done in integer arithmetic. Further, the values C_1 and C_2 are fixed for each template, and have to be computed only once. To improve efficiency further, a Fast Fourier Transform algorithm or a hierarchical scene matching algorithm [WH78] can be used.

The next step in the central mass detection (CMD) process is to form an initial list of suspicious tumors according to the correlation values computed at each image location. This can be done as follows:

- (1) Since multiple templates have been used in template matching to represent tumors of different sizes, there should be an equivalent number of correlation values measured at each image coordinate. Among these correlation values measured at a point, the largest is taken to represent the correlation value of that point and the corresponding tumor size is marked down.

- (2) The correlation measure ranges from 1 to -1. A value of 1 corresponds to a perfect match. As the degree of similarity between the template and the subimage decreases, the corresponding correlation value decreases. A value of 0 indicates that there is no likeliness between the template and the subimage. Image coordinates with negative correlation values can be ignored as they correspond to darker objects on a relatively bright background. The intended objects are of higher intensity than the background and so their corresponding correlation values should be positive.
- (3) Image coordinates with very low correlation values (between 0 and 0.2) can also be ignored as they do not constitute a significantly suspicious tumor. The image coordinates left (correlation value > 0.2) will form a list. Each entry in the list contains an image coordinate and its corresponding correlation value and tumor size.
- (4) The list is then reduced by clustering. Entries corresponding to overlapping tumors and of lower correlation values are removed from the list.
- (5) A fixed percentage of the entries in the list is taken as suspicious and forms an initial list of suspicious tumors. The use of this percentile method is based on the assumption that suspicious tumors at most occupy a fixed percentage of the breast area. Besides, the percentile

method is good as it can adapt to breast images of different richness in texture. A simple thresholding with a preset threshold value on the other hand cannot.

The initial list formed will then be passed to the radial structure recognition process to measure the spicularities of the suspicious tumors, and to the classification process (discussed in next section) to extract some local features of each suspicious tumor. Based on the local features measured, the classification process will determine the types of the suspicious tumors in the list. They may come out as false alarms (non-tumorous areas), suspicious stellate tumors or suspicious circumscribed tumors.

4.6 Classification

The classification process in the automated stellate tumor detection system first extracts local features (information) of each suspicious area in the initial tumor list. Combining this with information obtained from some other processes, it classifies the suspicious areas into non-tumorous areas, suspicious stellate tumors and suspicious circumscribed tumors. To perform these functions, the classification process is composed of two sub-processes: *false alarm removal process* and *tumor type differentiation process*.

4.6.1 False Alarm Removal

Since the percentile method is used to determine the correlation threshold value, some of the suspicious areas selected may not really correspond to genuine suspicious tumors. Lai [Lai88] has used two tests to remove the non-tumorous areas (false alarms).

(1) Bimodality Test

This test was first suggested by Sklansky [Skl76]. It is based on the fact that there should be a significantly high contrast between a genuine tumor and its surrounding background. When a grey-level histogram is constructed on a region containing both the suspicious area and its surrounding background, the histogram must at least be bimodal (or even multimodal). The

two peaks in the histogram correspond to the bright tumor area and the dark background. The bimodality should sustain even after subsequent smoothing of the histogram. On the other hand, a false alarm which comes from some anatomical structures or other noise does not have high contrast with its surrounding background, and so the corresponding histogram has one peak only.

It should be noted that the original image (before Gaussian smoothing and spicule removal) should be used. Smoothing and spicule removal will reduce the contrast of the tumor to the background and may introduce artifacts to the image. This test is good as it also removes some false alarms due to the artifacts introduced in the previous processes.

(2) Neighborhood Test

If a pixel P in the spicule-free image is the center of a suspicious area in the initial tumor list, the normalized cross-correlation values of the four immediate neighbors of P should be close to that of P when the same template is applied. This is because the boundary of a real tumor mass is very fuzzy and a good matching between a suspicious area and a template should not be confined to a single pixel located at the center of the suspicious area. Therefore, if the average cross-correlations of the immediate neighbors of P falls below the threshold value set (dynamically by percentile method) for that image, the suspicious area centered at P will be rejected as a suspicious

tumor.

Since the neighborhood test is quite a strong test, it may remove all suspicious areas in the initial tumor list. Therefore, the suspicious area with the highest correlation value is not discarded if it passes the bimodality test.

4.6.2 Tumor Type Differentiation

Since both the stellate tumors and the circumscribed tumors in the spicule-free image have the appearance of bright, roughly circular objects, they will be detected indiscriminately by the central mass detection (CMD) process. A suspicious area in the initial tumor list may correspond to a tumor of either type.

The spicularity measure suggested in Section 4.4.5 can be used to make the differentiation. It gives the likelihood of the existence of a surrounding radiating structure in the original image. Therefore, if spicularity of a suspicious area is higher than a threshold, the suspicious area should be taken as a suspicious stellate tumor or else a suspicious circumscribed tumor.

4.7 Experimental Results and Discussion

To test the effectiveness of the proposed detection system, twenty-seven mammograms have been used in our experiment. Among the twenty-seven mammograms, twenty-one of them contain one suspicious stellate each, four of them contain one suspicious circumscribed tumor each, and the last two contain two suspicious stellate tumors each. The determination of an area in the mammogram as a genuine suspicious tumor area is based on the radiologist's diagnosis.

The inclusion of the four mammograms containing circumscribed tumors is aimed at testing the differentiation power of the classification process in the proposed system. Once a suspicious tumor is located, the classification process should be used to identify its type.

Plates 2.1 and 2.3 (pages 18 and 19) show two of the twenty-seven mammogram images. The first one (Plate 2.1) contains a stellate tumor and the second one (Plate 2.3) contains a circumscribed tumor. In the preprocessing process, breast region isolation and breast region normalization are applied to the images, and the normalized breast images (Plates 4.5 and 4.6) are formed. These normalized images are then used in all subsequent processes.

In the radial structure recognition (RSR) process, spinal pixels in the normalized breast images are detected. Plates 4.7 and 4.8 show the images overlayed with the spinal pixels. Based on the spinal pixels detected, the

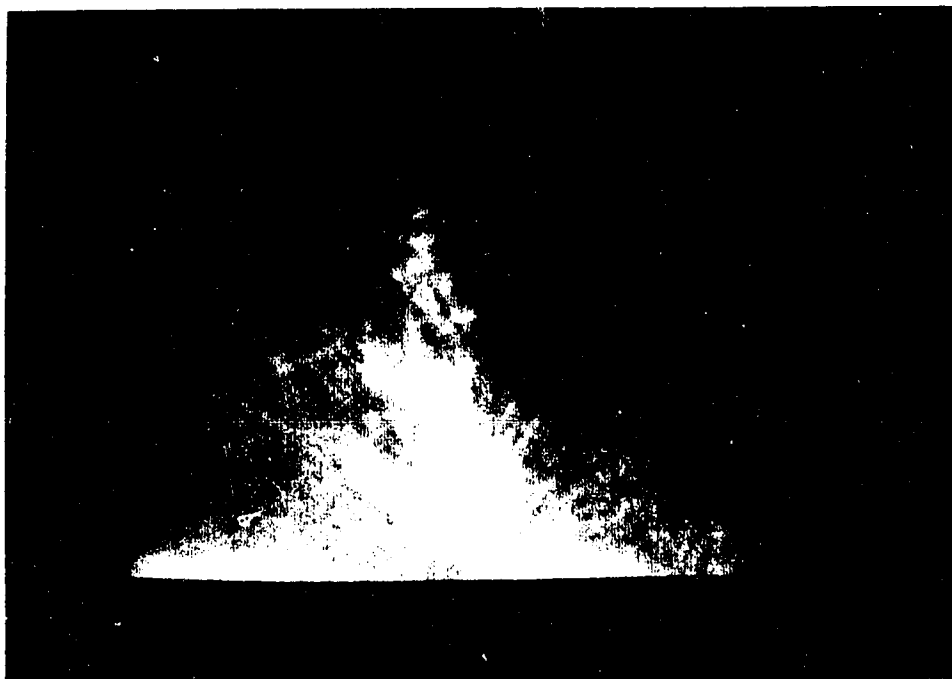


Plate 4.5: Image s1 after the preprocessing stage
(result of Breast Region Isolation and Breast Region Normalization)



Plate 4.6: Image c1 after the preprocessing stage
(result of Breast Region Isolation and Breast Region Normalization)

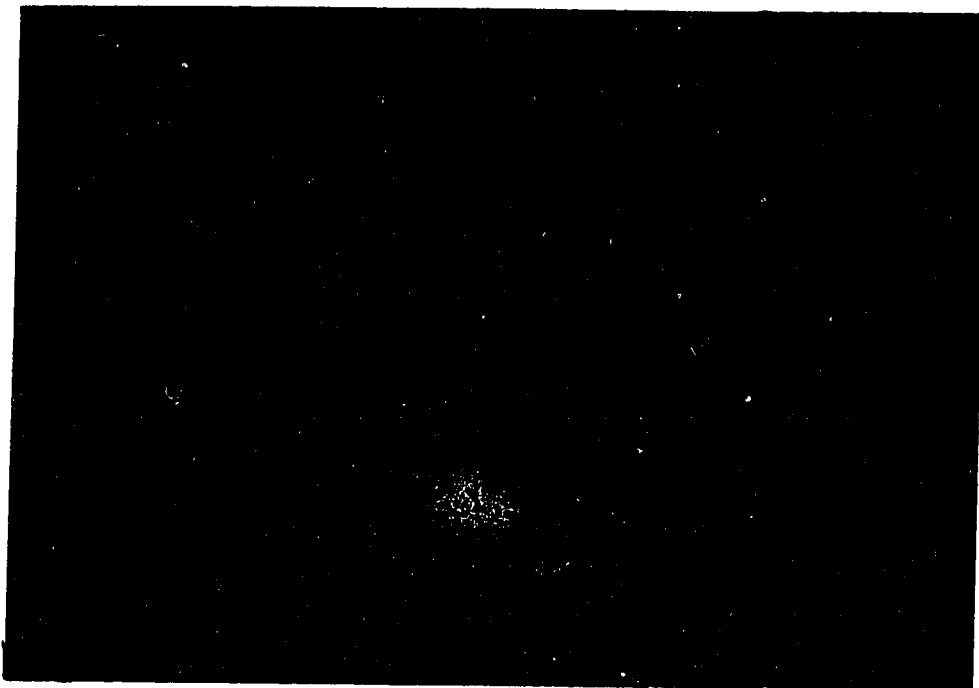


Plate 4.7: Image s1 overlayed with red-colored spinal pixels
(detected by the Radial Structure Recognition process)

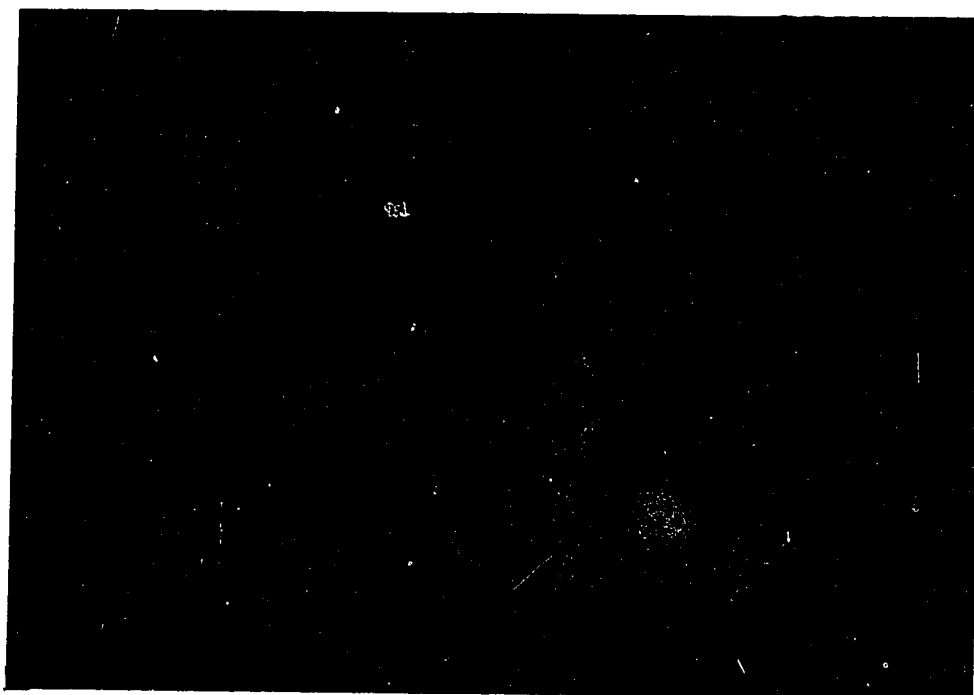


Plate 4.8: Image c1 overlayed with red-colored spinal pixels
(detected by the Radial Structure Recognition process)

spicules in the images are removed. The spicule-free images formed are processed by the central mass detection (CMD) process which produces a list of suspicious tumors. Plates 4.9 and 4.10 show the images with the suspicious tumors circled.

The list is then passed back to the RSR process in which the associated spinal pixels of each suspicious tumor in the list are identified and the spicularity is computed. Plates 4.11 and 4.12 show the images overlaid with red-colored spinal pixels. These red-colored spinal pixels are the ones with associated suspicious tumors.

In the final classification stage, the suspicious tumors in the initial tumor list undergo further analysis. Some of them may then be identified as non-tumorous areas (false alarms) and removed from the list. For those suspicious areas left in the list, their types are determined according to the spicularity measures. Plates 4.13 and 4.14 show the final results of detection, where the suspicious tumors are classified respectively as suspicious stellate tumor and suspicious circumscribed tumor. These results agree with the diagnosis given by an expert radiologist, Dr. Bill Castor of the Radiology Department at the Cross Cancer Institute of the University of Alberta.

When the diagnostic results produced by radiologist and computer are compared, one cannot expect a 100% match in locations and sizes of the suspicious tumors. A more reasonable and generally used criterion [HSAA79] [Lai88] is to define two specific areas as matching if they overlay each other

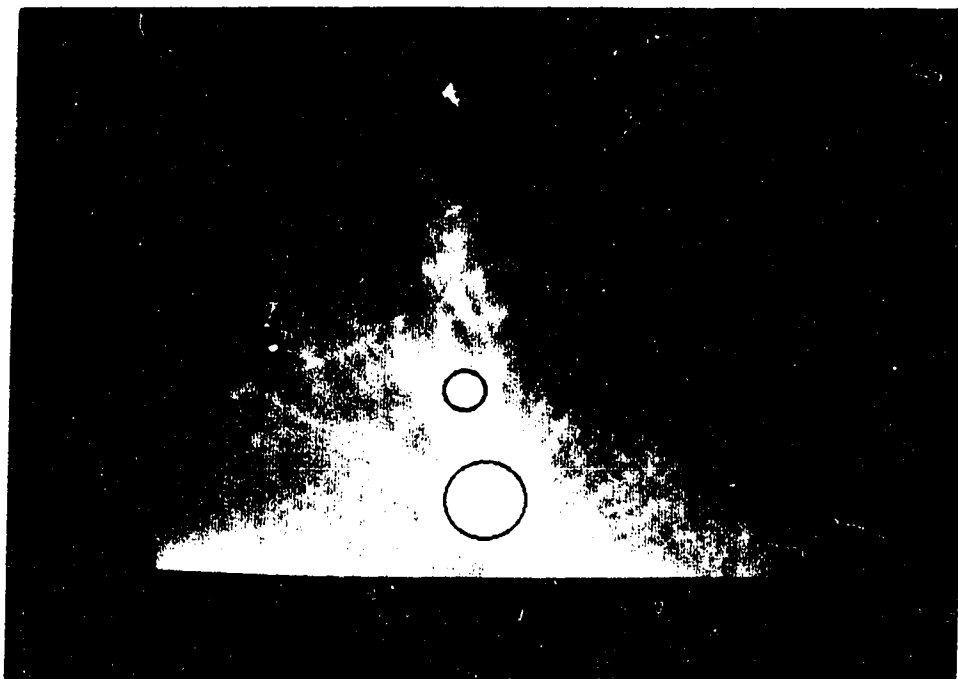


Plate 4.9: Image s1 with the suspicious areas circled
(detected by the Central Mass Detection process)



Plate 4.10: Image c1 with the suspicious areas circled
(detected by the Central Mass Detection process)

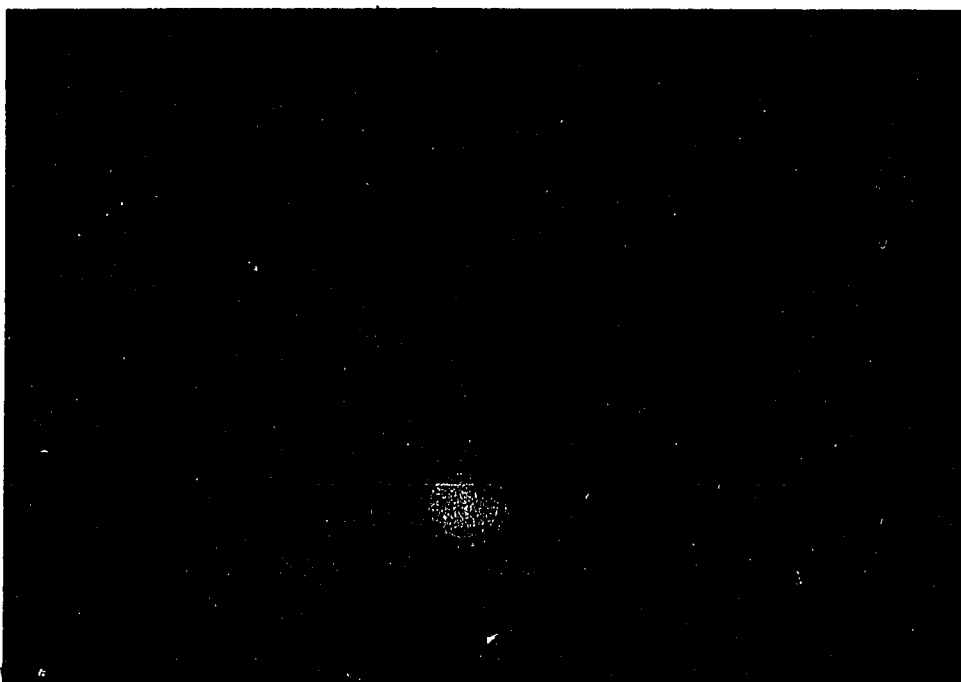


Plate 4.11: Image s1 overlayed with the red-colored spinal pixels of the suspicious areas



Plate 4.12: Image c1 overlayed with the red-colored spinal pixels of the suspicious areas

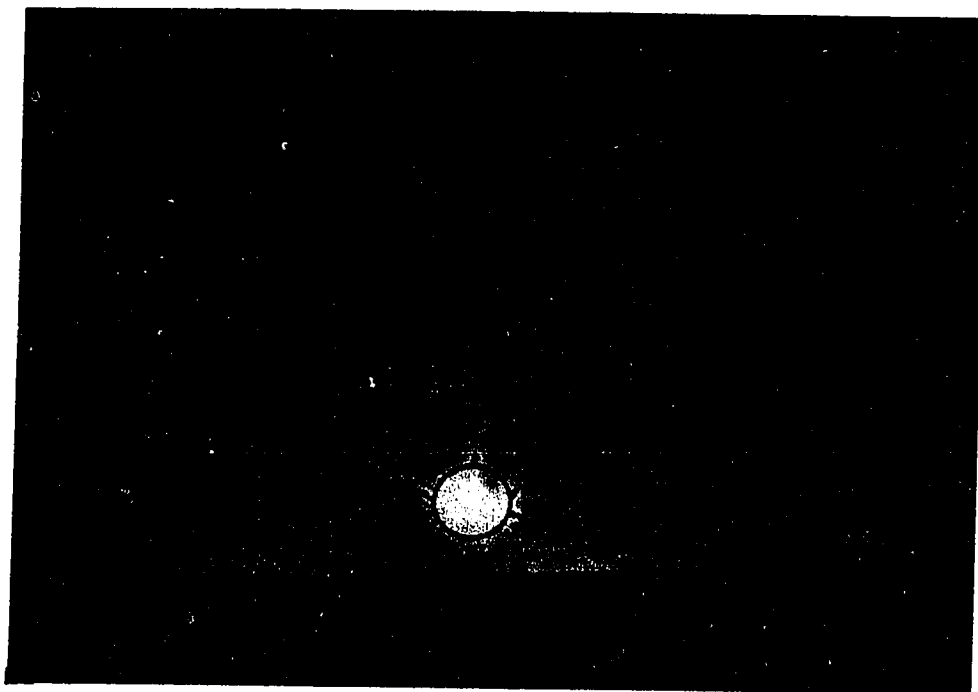


Plate 4.13: Image s1 with the detected suspicious stellate tumor marked



Plate 4.14: Image c1 with the detected suspicious circumscribed tumor marked

50% or more. In fact, there can be three different outcomes in detection:

- (1) **hit** – A suspicious tumor marked by radiologist overlaps 50% or more with a suspicious tumor detected by the computer.
- (2) **miss** – No suspicious tumor detected by the computer can match with the suspicious tumor marked by radiologist.
- (3) **false alarm** – A suspicious tumor detected by the computer cannot match with any suspicious tumor marked by radiologist.

Using the criterion above, we found a hit rate of 69% and a false alarm rate of 0.3 occurrence per test image with the twenty-seven test images. For the tumor type identification, all except two of the detected tumors are correctly classified.

In reviewing the test images, there are three problems which may cause the detection system to miss a suspicious tumor. Firstly, a suspicious tumor may be partially obscured by a bright parenchyma (essential part of the breast). Plate 4.15 shows an example. Another problem is that the central tumor mass of the suspicious stellate tumor barely exists and/or strongly deviates from a circular shape. It is anticipated that when the “roughly circular shape” criterion of the central mass detection (CMD) process is relaxed to “roughly circular or elliptical shape”, the detection (hit) rate will be improved without greatly affecting the false alarm rate.

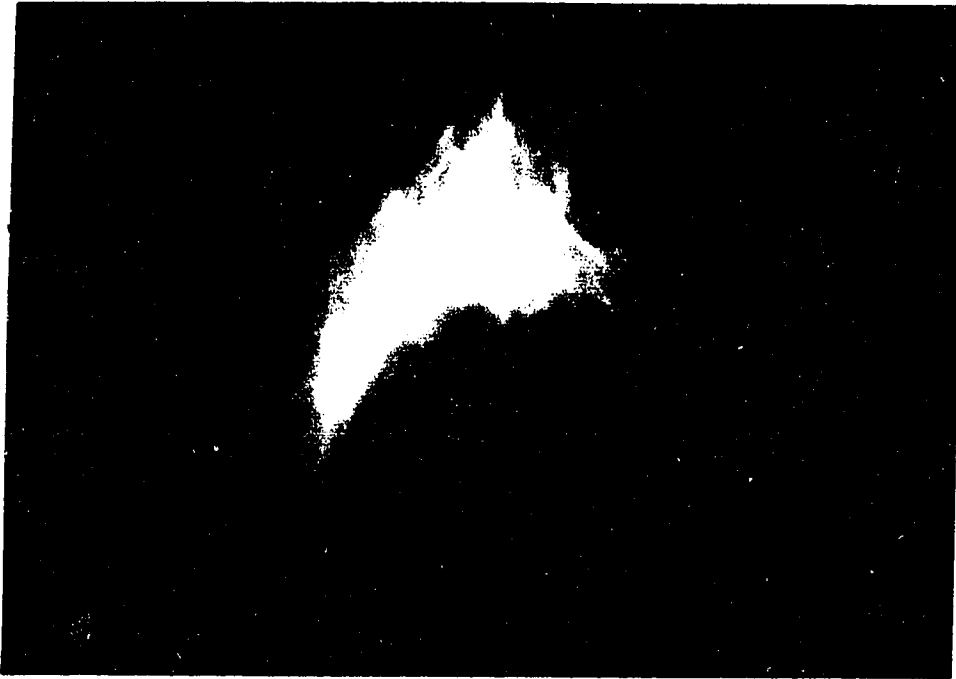


Plate 4.15: Image s4 having a partially obscured suspicious stellate tumor

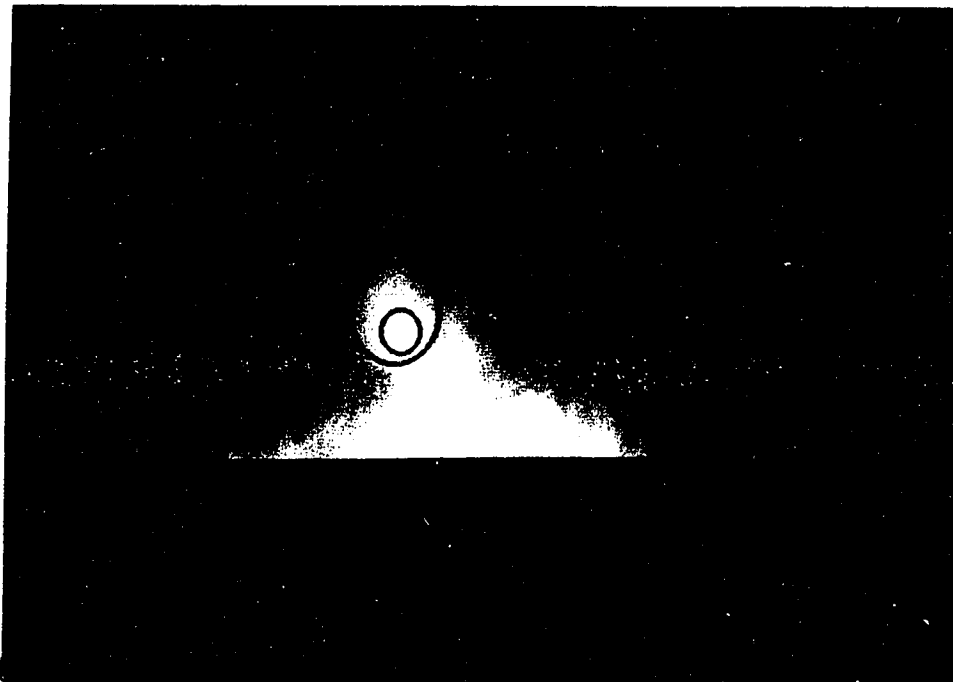


Plate 4.16: Image s9 having a suspicious stellate tumor of unusual structure

The last problem is related to the unusual structure of some tumors. In two cases, there is another brighter circular area inside the suspicious tumor. Plate 4.16 (page 130) shows one of the two cases. Since the inner area is nearly circular and of high contrast, it produced a high correlation value in template matching. In the clustering stage, the outer larger suspicious tumor is removed from the initial tumor list. Even though the locations of the suspicious tumors in these two test cases are correctly marked by the detection system, the tumor sizes are not compatible with the radiologist's diagnosis and so they are considered as misses. In order to circumvent this problem, the clustering process may have to give higher weighting to suspicious areas of larger size.

Chapter 5

Concluding Remarks

The objective of this research is to construct an automated breast tumor detection system that can detect all suspicious stellate tumor areas in mammograms. It begins with a study on the nature of the mammographic images and the tumors to be detected. Several factors that make detection difficult have been identified. They include the existence of noise and anatomical structures, and low contrast between areas which come from the projection of normal cells and cancerous cells in the breast. The problem is further aggravated by the complexity of the tumor structure and the fuzziness of the tumor boundary.

A method which combines simple thresholding and blob coloring techniques has been developed to first isolate the imaged breast from the background. This removes some potential false alarm and greatly speeds up

subsequent processes. Three different radial structure recognition (RSR) approaches, namely edge-oriented RSR, field-oriented RSR and spine-oriented RSR, are proposed and tested. The best one, spine-oriented RSR, is then chosen to identify the existence of a radiating structure.

In addition, a modified directional selective median filter is suggested for removing the spicules in the breast images. The resultant spicule-free images are then passed to the central mass detection (CMD) process, where template matching is used to locate all suspicious tumor areas. A classification process further examines each suspicious tumor area and determines its type.

When the proposed system is applied to twenty-seven mammograms, it gives a good detection rate (69%) with an extremely low false alarm rate (0.3 occurrence per mammogram). In more than half of the cases, the true tumor area is the only suspicious area reported by the computer.

Future Research

In future, some further improvements to the proposed detection system can be made. The shape requirement for the detection of stellate tumors in the spicule-free images could be relaxed in the way that the filter response to a stellate tumor would generally increase. This might improve the detection rate.

Besides, mammograms taken from different viewing positions can be read to the system and the results of detection are correlated. This not only allows

a more accurate diagnosis but also avoids the problem which arises when part of a tumor is obscured at a certain viewing angle. In addition, the diagnosis may be further improved by comparing the results obtained from the left and right breast images of the same woman.

The last suggestion is to incorporate the retraction of the parenchyma (essential part of the breast) and the localized skin thickening into the detection mechanism, as they are the phenomena when the spicules of some stellate tumors reach the skin [TD85b, page 89]. However, the implementation of such system will require intensive research effort.

It is my sincere hope that this work will contribute to the development of a competent automated system which can detect breast cancer at an early stage, and that many lives can be saved as a result of this research effort.

Bibliography

- [AG72] Laurens V. Ackerman and Earl E. Gose. Breast lesion classification by computer and xeroradiograph. *Cancer*, 30:1025–1035, 1972.
- [Bau88] Thomas W. Bauer. The relationship between breast pathology and prognosis. In Sharon Grundfest-Broniatowski and Caldwell B. Esselstyn, Jr., editors, *Controversies in Breast Disease: Diagnosis and Management*, pages 123–150. Marcel Dekker, Inc., 1988.
- [BB82] Dana H. Ballard and Christopher M. Brown. *Computer Vision*. Prentice Hall, 1982.
- [BS73] D. Ballard and J. Sklansky. Tumor detection in radiographs. *Computers and Biomedical Research*, 6:299–321, 1973.
- [Can83a] John Canny. A variational approach to edge detection. In *Proceedings of The National Conference on Artificial Intelligence*, pages 22–26, August 1983.
- [Can83b] John Francis Canny. Finding edges and lines in images. Master's thesis, Massachusetts Institute of Technology, June 1983.
- [CBRE88] Jennifer Caseldine, Roger Blamey, Eric Roebuck, and Christopher Elston. *Breast Disease for Radiographers*. Wright, Toronto, 1988.
- [CD86] Roland T. Chin and Charles R. Dyer. Model-based recognition in robot vision. *ACM Computer Surveys*, 18(1):67–108, March 1986.

- [Chi88] William A. Chilcote. Screening for breast cancer. In Sharon Grundfest-Broniatowski and Caldwell B. Esselstyn, Jr., editors, *Controversies in Breast Disease: Diagnosis and Management*, pages 181–200. Marcel Dekker, Inc., 1988.
- [CK72] C. K. Chow and T. Kaneko. Automatic boundary detection of the left ventricle from cineangiograms. *Computers and Biomedical Research*, 5:388–410, 1972.
- [Cla89] James J. Clark. Authenticating edges produced by zero-crossing algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(1):43–57, January 1989.
- [CR88] Robert E. Coleman and Robert D. Rubens. Management of advanced breast cancer. In Sharon Grundfest-Broniatowski and Caldwell B. Esselstyn, Jr., editors, *Controversies in Breast Disease: Diagnosis and Management*, pages 401–424. Marcel Dekker, Inc., 1988.
- [DBG86] Atam P. Dhawan, Gianluca Buelloni, and Richard Gordon. Enhancement of mammographic features by optimal adaptive neighborhood image processing. *IEEE Transactions on Medical Imaging*, MI-5(1):8–15, March 1986.
- [Ega70] Robert L. Egan. *Mammography and Breast Diseases*. Golden's Diagnostic Roentgenology. Williams and Wilkins Company, Baltimore, 1970.
- [Gal83] H. Stephen Gallager. Problems in the classification of breast cancer. *Radiologic Clinics of North America*, 21(1):13–26, March 1983.
- [GB83] Richard H. Gold and Lawrence W. Bassett. Mammography: History and state of the art. In Stephen A. Feig and Robert McLelland, editors, *Breast Carcinoma: Current Diagnosis and Treatment*, pages 95–98. MASSON Publishing USA, Inc., New York, 1983.
- [GBB88] Sharon Grundfest-Broniatowski and Thomas W. Bauer. Benign breast disease: A clinical and pathological classification. In

- Sharon Grundfest-Broniatowski and Caldwell B. Esselstyn, Jr., editors, *Controversies in Breast Disease: Diagnosis and Management*, pages 3–42. Marcel Dekker, Inc., 1988.
- [GBC87] Richard H. Gold, Lawrence W. Bassett, and Walter F. Coulson. Mammographic features of malignant and benign disease. In Lawrence W. Bassett and Richard H. Gold, editors, *Breast Cancer Detection: Mammography and Other Methods in Breast Imaging 2nd Edition*, pages 15–66. Grune & Stratton, Inc, Toronto, 1987.
- [GBKS87] Richard H. Gold, Lawrence W. Bassett, and Carolyn Kimme-Smith. Introduction to breast imaging: State of the art and future directions. In Lawrence W. Bassett and Richard H. Gold, editors, *Breast Cancer Detection: Mammography and Other Methods in Breast Imaging 2nd Edition*, pages 3–14. Grune & Stratton, Inc, Toronto, 1987.
- [GPP88] Paolo Grattoni, Fabrizio Plllastri, and Amedeo Premoli. A contour detection algorithm based on the minimum radial inertia (MRI) criterion. *Computer Vision, Graphics and Image Processing*, 43:22–36, 1988.
- [HKD⁺71] Ernest L. Hall, Richard P. Kruger, Samuel J. Dwyer, David L. Hall, Robert W. McLaren, and Gwilym S. Lodwick. A survey of preprocessing and feature extraction techniques for radiographic images. *IEEE Transactions on Computers*, C-20(9):1032–1044, September 1971.
- [Hom87] Marc J. Homer. Imaging features and management of characteristically benign and probably benign breast lesions. *Radiologic Clinics of North America*, 25(5):939–951, 1987.
- [Hor86] Berthold K. P. Horn. *Robot Vision*. MIT Press, Cambridge, Massachusetts, 1986.
- [HSAA79] William Hand, John L. Semmlow, Laurens V. Ackerman, and Frank S. Alcorn. Computer screening of xeromammograms: A

- technique for defining suspicious areas of the breast. *Computers and Biomedical Research*, 12:445–460, 1979.
- [KHDL71] R. P. Kruger, E. L. Hall, S. J. Dwyer, and G. S. Lodwick. Digital techniques for image enhancement of radiographs. *Bio-Medical Computing*, 2:215–238, 1971.
- [Kop87] Daniel B. Kopans. Nonmammographic breast imaging techniques: Current statues and future developments. *Radiologic Clinics of North America*, pages 961–971, September 1987.
- [KOS75] Carolyn Kimme, Bernard J. O’Loughlin, and Jack Sklansky. Automatic detection of suspicious abnormalities in breast radiographs. In A. Klinger, K.S. Fu, and T.L. Kunii, editors, *Data Structure, Computer Graphics and Pattern Recognition*, pages 427–447. Academic Press, New York, 1975.
- [KW87] Michael Kass and Andrew Witkin. Analyzing oriented patterns. *Computer Vision, Graphics and Image Processing*, 37:362–385, 1987.
- [Lai88] Shuk-Mei Lai. A technique for automated detection of breast tumors in mammograms. Master’s thesis, The University of Alberta, Fall 1988.
- [Les84] Richard G. Lester. The contributions of radiology to the diagnosis, management, and cure of breast cancer. *Radiology*, 151:1–7, 1984.
- [LLB88] S. Mei Lai, Xiaobo Li, and Walter F. Bischof. Automated detection of breast tumor. In *Proceedings of Vision Interface’ 88*, pages 35–40. Canadian Image Proceedings and Pattern Recognition Society, June 1988.
- [LZ85] Xiaobo Li and Jan M. Zytkow. A survey of methods for object detection from images. Technical Report #WSU-CS-85-2, Department of Computer Science, Wichita State University, Wichita, Kansas, March 1985.
- [Mar82a] D. Marr. *Vision*, pages 54–79. Freeman, San Francisco, 1982.

- [Mar82b] John E. Martin. *Atlas of Mammography, Histologic & Mammographic Correlations*, pages 65–101. Williams & Wilkins, Baltimore, London, 1982.
- [MB88] Siva Murthy and Ronald M. Bukowski. Systemic therapy for advanced breast cancer. In Sharon Grundfest-Broniatowski and Caldwell B. Esselstyn, Jr., editors, *Controversies in Breast Disease: Diagnosis and Management*, pages 425–438. Marcel Dekker, Inc., 1988.
- [MSE83] Marjorie B. McSweeney, Perry Sprawls, and Robert L. Egan. Enhanced image mammography. *American Journal of Roentgenology*, 140:9–14, January 1983.
- [RK82] Azriel Rosenfeld and Avinash C. Kak. *Digital Picture Processing*, volume 2. Academic Press, 1982.
- [RS88] Foster Roger S., Jr. Breast cancer epidemiology: Controversies in the identification and interpretation of risk factors. In Sharon Grundfest-Broniatowski and Caldwell B. Esselstyn, Jr., editors, *Controversies in Breast Disease: Diagnosis and Management*, pages 111–122. Marcel Dekker, Inc., 1988.
- [Sk176] J. Sklansky. Boundary detection in medical radiographs. In K. Preston and M. Onoe, Jr., editors, *Digital Processing of Biomedical Images*, pages 309–322. Plenum Press, New York and London, 1976.
- [SSA⁺80] John L. Semmlow, Annapoorin Shaagopappan, Laurens V. Ackerman, William Hand, and Frank S. Alcorn. A fully automated system for screening xeromammograms. *Computers and Biomedical Research*, 13:350–362, 1980.
- [SSWC88] R. K. Sahoo, S. Soltani, A. K. C. Wong, and Y. C. Chen. A survey of thresholding techniques. *Computer Vision, Graphics and Image Processing*, 41:233–260, 1988.
- [SWGS77] Kennan T. Smith, Sheldon L. Wagner, Ronald B. Guenther, and Donald C. Solmon. The diagnosis of breast cancer in mammo-

- grams by the evaluation of density patterns. *Radiology*, 125:383–386, November 1977.
- [TD85a] László Tabár and Peter B. Dean. Basic principles of mammographic diagnosis. *Diagn. Imag. Clin. Med.*, 54:146–157, 1985.
 - [TD85b] László Tabár and Peter B. Dean. *Teaching Atlas of Mammography*. Georg Thieme, Stuttgart, second edition, 1985.
 - [TD87] László Tabár and Peter B. Dean. The control of breast cancer through mammography screening. *Radiologic Clinics of North America*, pages 961–971, September 1987.
 - [WB87] Christopher J. Williams and Roger B. Buchanan. *The Medical Management of Breast Cancer*. The Johns Hopkins University Press, Baltimore, 1987.
 - [WEM⁺67] Fred Winsberg, Milton Elkin, Josiah Macy, Victoria Bordaz, and William Weymouth. Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis. *Radiology*, 89:211–215, 1967.
 - [WH78] Robert Y. Wong and Ernest L. Hall. Sequential hierarchical scene matching. *IEEE Transactions on Computers*, c-27(4):359–366, 1978.
 - [YP86] Alan L. Yuille and Tomaso A. Poggio. Scaling theorems for zero crossings. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(1):15–25, January 1986.