Registration of mass-like objects in sequential mammograms using graph matching

by

Fei Ma, *M.Sc.*

School of Computer Science, Engineering and Mathematics,
Faculty of Science and Engineering

October 10, 2008

A thesis presented to the
Flinders University
in total fulfillment of the requirements for the degree of
Doctor of Philosophy

Contents

List of Figures

List of Tables

List of Abbreviations

CAD	Computer-Aided Detection
MLO	Medio-Lateral Oblique
CC	Cranial-Caudal
ROI	Region of Interest
ROC	Receiver Operating Characteristic
A_Z	Area under the ROC Curve
AP	Adaptive Pyramid
MI	Mutual Information
SV	Shift Variance
MST	Minimum Spanning Tree
CSI	Common Subgraph Isomorphism
RMSD	Root Mean Square Difference
Av. dif	Average Difference
STD	Standard Deviation
Mini-MIAS	Mammographic Image Analysis Society Database
ARBE	Average Right Boundary Error
FP	False Positive
FN	False Negative
RNE	Row Normalized Error
ROW	Real Orthogonal Wavelets
MRF	Markov Random Field
ICM	Iterated Conditional Modes
GT	Ground Truth
LDA	Linear Discriminant Analysis
cmr	Correct Match Rate
me	Match Efficiency

Summary

Sequential mammograms contain important information, such as changes of the breast or developments of the masses, for diagnosis of disease. Comparison of sequential mammograms plays an important part for radiologists in identifying malignant masses. However, currently computer-aided detection (CAD) programs can not use such information efficiently. The difficulties lie in the registration of sequential mammograms.

Most of current methods register sequential mammograms based on control points and image transformations. For these methods to work, extraction and correspondence of the control points is essential. This thesis presents a new approach in registering mammograms. The proposed method registers mammograms by associating mass-like objects in sequential mammograms directly. The mass-like objects appear in the images of normal breasts as well as images of breast with cancer. When the mass-like objects in sequential mammograms are accurately associated, measurements of changes in mass-like objects over time become possible. This is an important way to distinguish mass-like objects associated with cancer from cysts or other benign objects.

The proposed method is based on graph matching. It uses the internal structure of the breast represented by the spatial relation between the mass-like objects to establish a correspondence between the sequential mammograms. In this method, the mammogram is firstly segmented into separate components using an adaptive pyramid (AP) segmentation algorithm. A series of filters, based on the features of components, is then applied to the components to remove the undesired ones. The remaining components, the mass-like objects, are represented by a complete graph. The spatial relations between the remaining mass-like objects are expressed by fuzzy spatial relation representation and are associated to the edges of the graph as weights. Association of the mass-like objects of two sequential mammograms is realized by finding a common subgraph of the corresponding two graphs using the backtrack algorithm.

The segmentation methods developed in the course of this work were tested on a separate problem in computer-aided detection of breast cancer, namely the automatic extraction of the pectoral muscle.

The graph matching method was tested independently of the segmentation method on artificially distorted mammograms and the full process, including the

segmentation and the graph matching, was evaluated on 95 temporal mammogram pairs. The present implementation indicates only a small improvement in cancer detection rates but also presents opportunities for a substantial development of the basic method in the future.

Publications

1. Ma, F., Bajger, M. & Bottema, M. J. (2005), Extracting the pectoral muscle in screening mammograms using a graph pyramid, *in* B. C. Lovell & A. J. Maeder, eds, 'APRSWorkshop on Digital Image Computing', The University of Queensland, Griffith University, Brisbane, Australia.

2. Bajger, M., Ma, F. & Bottema, M. J. (2005), Minimum spanning trees and active contours for identification of the pectoral muscle in screening mammograms, *in* B. C. Lovell, A. J. Maeder, T. Caelli & S. Oursellin, eds, 'Digital Image Computing Techniques and Applications', IEEE Computer Society Conference Publishing Service, Cairns, Qld Australia.

3. Ma, F., Bajger, M., Slavotinek, J. P. & Bottema, M. J. (2006), Validation of graph theoretic segmentation of the pectoral muscle, *in* 'Digital Mammography, IWDM 2006, 8th International Workshop', Springer, Manchester, UK, pp. 642-649.

4. Ma, F., Bajger, M., Slavotinek, J. P. & Bottema, M. J. (2007), 'Two graph theory based methods for identifying the pectoral muscle in mammograms', *Pattern Recognition* **40**, 2592-2602.

5. Ma, F., Bajger, M. & Bottema, M. J. (Dec. 2007), Robustness of two methods for segmenting salient features in screening mammograms, *in* '9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications (DICTA 2007)', IEEE., Glenelg, South Australia, pp. 112-117.

6. Yu, L., Ma, F., Jayasuriya, A., Sigelle, M. & Perreau, S. (2007), A new contour detection approach in mammogram using rational wavelet filtering and MRF smoothing, in '9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications (DICTA 2007)', IEEE., Glenelg, South Australia, pp. 106-111.

7. Susukida, H., Ma, F. & Bajger, M. (2008), Automatic tuning of a graph-based image segmentation method for digital mammography applications, *in* 'Proc. IEEE International Symposium on Biomedical Imaging (ISBI08)', Paris. pp. 89-92.

8. Ma, F., Bajger, M. & Bottema, M. J. (2008a), A graph matching based automatic regional registration method for sequential mammogram analysis, *in* 'Medical Imaging 2008: Computer-Aided Diagnosis', Vol. 6915, SPIE, San Diego, CA, USA, p. 69151Z.

9. Ma, F., Bajger, M. & Bottema, M. J. (2008b), Temporal analysis of mammograms based on graph matching, *in* 'Digital Mammography / IWDM', Springer, Tucson, AZ, USA, pp. 158-165.

Certification

I certify that this thesis does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

As requested under Clause 14 of Appendix D of the *Flinders University Research Higher Degree Student Information Manual* I hereby agree to waive the conditions referred to in Clause 13(b) and (c), and thus

- Flinders University may lend this thesis to other institutions or individuals for the purpose of scholarly research;

- Flinders University may reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Signed   Dated

Fei Ma

Acknowledgements

First and foremost, I am deeply indebted to my supervisor, Associate Professor Murk Bottema for his guidance, support and encouragement all the way through my PhD studies. His extensive knowledge in mammography and excellence in mathematics have been of great help in solving many problems I confronted. The work in this thesis would not have been carried out without his profound insight and advice at every stage of my thesis research. One simply could not wish for a better or friendlier supervisor.

I am sincerely grateful to my co-supervisor Dr. Mariusz Bajger for his invaluable advice, discussion and cooperation on many issues that this thesis involved. His carefulness is mostly impressive and he shared with me a quality workstation which much of work has been done on. I am also grateful for his constructive feedback and the efforts he made in proof reading this thesis.

I wish to thank Dr. John Slavotinek for his support on this project and Dr. Ray Booth for providing me opportunities of teaching assistant work and always correcting me my spoken English.

I am grateful to the secretaries and technical staffs of school of Computer Science, Engineering and Mathematics of Flinders for always being friendly and helpful.

I would like to thank Flinders Medical Centre Foundation (FMC) and Drake Food Markets for awarding me the "2004 Drake Food Markets Award For Breast Cancer Research". I also wish to thank National Breast Cancer Foundation for providing the funds for this study.

I wish to thank my entire extended family, my brother, my sister, and my uncles for always being supportive.

Last but not least, I wish to thank my grandparents, my parents, my parents in law and my wife Limin Yu for their unconditional caring and support. I would also like to thank my 10 months old daughter Annie for being quiet during the night while I can have quiet time to write my thesis. She seems quite a keyboard lover and was always trying to take over it whenever she was close to it. This thesis is dedicated to all of them.

Chapter 1

Introduction

This chapter provides background on breast cancer and mammography, and gives the motivation of this thesis. The chapter is organized as follows. Section 1.1 describes current facts about the breast cancer, including the incidence and risk of breast cancer. Section 1.2 introduces breast cancer screening. Two important measures toward the efficiency of breast cancer screening, the sensitivity and specificity, are described in Section 1.3. Section 1.4 is about the computer-aided detection (CAD). Several studies about the efficiency of CAD programs are described in this section. Motivation and approaches are given in Section 1.5. Finally, in Section 1.6, we present an overview of this thesis.

1.1   Incidence of Breast Cancer

Breast cancer is the most common cancer among women worldwide. In 2003, breast cancer was the most common cancer in Australian female population with 11,788 new cases diagnosed, and was the most common cancer death with 2,710 deaths (AIHW (Australian Institute of Health and Welfare) and AACR (Australasian Association of Cancer Registries) 2007). According to the National Breast Cancer Foundation of Australia, more than 13,000 new cases were expected in 2007 in Australia and more than 2000 in New Zealand. In America, 184,450 new cases of invasive breast cancer (182,460 female, 1,990 male), as well as 67,770 (female) additional cases of in situ breast cancer are expected to be diagnosed and approximately 40,930 (40,480 female, 450 male) are expected to die from breast cancer in 2008, making it the second common cancer death (after lung cancer) (American Cancer Society 2008). In Europe, breast cancer was the most common cancer in 2006 with 429,900 cases expected to be diagnosed and 131,900 estimated deaths (Ferlay, Autier, Boniol, Heanue, Colombet & Boyle 2007).

The incidence of breast cancer continues to rise. The Australian Institute of Health and Welfare & National Breast Cancer Centre (2006) shows that the number of new cases of breast cancer of Australia women increased from 5,318 in

1983 to 12,027 in 2002 and it is projected that there will be 13,261 new cases in 2006 and 14,818 in 2011. According to the report, the age-standardised incidence of breast cancer in women increased from 80 per 100,000 population in 1983 to 98 per 100,000 in 1992 and 117 per 100,000 in 2002.

According to AIHW (Australian Institute of Health and Welfare) and AACR (Australasian Association of Cancer Registries) (2007), breast cancer was the highest risk in 2003 for Australian women, with a 1 in 11 chance before age 75 and a 1 in 9 chance before age 85. According to the National Breast Cancer Foundation of Australia, women aged 50 to 69 are at the highest risk of developing breast cancer, with 48 per cent of breast cancers occurring in this age group, and 75 per cent above the age of 50. In America, currently the lifetime risk of being diagnosed with breast cancer was reported as 12.3 per cent (1 in 8) (American Cancer Society 2007). In UK, according to the national statistics of UK, a rate of 121 cases per 100,000 women was recorded in 2004.

While breast cancer continues to be the first cause of cancer related death in females, more women are surviving this disease than ever before. AIHW (Australian Institute of Health and Welfare) and AACR (Australasian Association of Cancer Registries) (2007) show that deaths from breast cancer increased by 4 per cent from 1993 to 2003, but the death rate for breast cancer decreased 20 per cent from 1993 to 2003. Also, according to the National Breast Cancer Foundation of Australia, the percentage of women diagnosed with breast cancer and expected to live 5 years after the diagnosis increased from 71 per cent, in the period from 1982 to 1986, to 86 per cent. This is most likely due to the population-based screening program and more effective treatments.

1.2 Breast Cancer Screening

If breast cancer is found at an early stage, there is a greater chance of successful treatment. Early detection of breast cancer is believed to be the key to reducing the mortality. Currently, mammography is the most effective method for early breast cancer detection. Mammography can generally reveal benign and cancerous growths before the patient or the physician can feel them. Breast cancer screening has been introduced in many countries to provide intensive monitoring of breast cancer. In Australia, women, primarily those between the age of 50 to 69, are recommended and invited to screening mammograms every two years. In 2002, a working group, consisting of 24 experts from 11 countries, convened by the International Agency for Research on Cancer (IARC) of the World Health Organization (WHO), found that the mammography screening reduces the mortality from breast cancer among women between 50 and 69 by about 35 per cent. For women aged 40-49 years, the working group found only limited evidence for a reduction of mortality from breast cancer. Evidence of benefit for older women over 70 was also not strong enough to recommend general routine screening. In

Australia, all women aged 40 or more are eligible to attend screening.

A typical attending of breast cancer screening consists of following steps. The examinee positions each of her breasts in turn between two flat plates on the x-ray machine. The machine compresses the breast between the plates. This compression helps to even out the breast tissue so that the tissue can be imaged at the same thickness. It also reduces the amount of radiation needed to penetrate through the tissue and reduces the amount of scatter radiation to the rest of the body. After the positioning of the breast, low dose X-rays irradiate the positioned breast and produce pictures of the inside of the breast. For each breast, there is usually a set of two views to be taken, one from top to bottom, called cranial-caudal (CC) view, and one from the side, called medio-lateral oblique (MLO) view (Fig. 1.1).
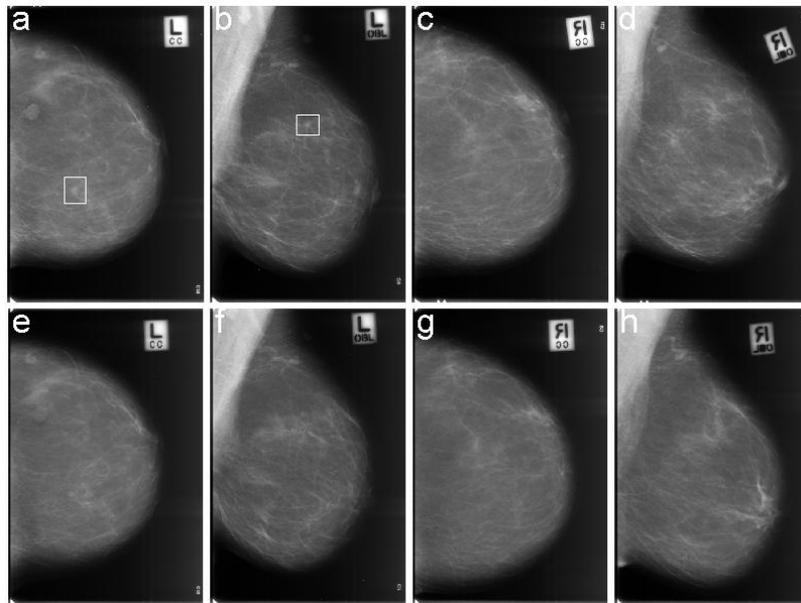
Figure 1.1: Example of temporal mammograms. (a)-(d) and (e)-(h) are two sets of mammograms taken in two consecutive screening sessions. (a)-(d) were taken in a current screening session. (e)-(h) were taken in a corresponding previous screening session. (a) and (e) are CC view mammograms of the left breast, (b) and (f) are MLO view mammograms of the left breast, (c)(g) and (d)(h) are CC and MLO view mammograms of the right breast. White boxes in (a) and (b) enclose malignant masses that were marked out by a radiologist.

The mammograms are then read by the radiologists. If any abnormalities are identified by the radiologists, the woman is called back for further assessments. In Australia, two radiologists read each case independently. A third radiologist will read the case if there is disagreement between the first two readers.

1.3   Sensitivity and Specificity of Mammography

Screening may fail to catch the disease, on the other hand, some detected cancers
are not real cancers, causing unnecessary call back. Sensitivity and specificity are
two important measures of a screening program. The sensitivity measures the
success of mammography in detecting breast cancer. The definition of sensitivity
varies between studies. In Australian Institute of Health and Welfare 2008 (2007),
the sensitivity of the breast screening program was defined as the proportion
of invasive breast cancers detected within the BreastScreen Australia Program
out of all invasive breast cancers (interval cancers plus screen-detected cancers)
diagnosed in program-screened women in the screening interval. Studies have
shown that the sensitivity of the mammography in women aged over 50 ranges
from 68 per cent to over 90 per cent, with most trials achieving sensitivities of
around 85 per cent. Australian Institute of Health and Welfare 2008 (2007) shows
that the sensitivity of the Australia screening program in 2001, 2002 and 2003
in 24 months follow-up of women aged 50-69 years in the first screening round
was reported as 79 per cent. For subsequent screening rounds, the sensitivity for
women in the same age group in the same period was 71 per cent. The specificity
measures how reliably the mammography identifies a normal mammogram as
normal. It is defined as the number of women whose mammograms are identified
as negative divided by the number of women who truly do not have breast cancer.
Studies show that the specificity of mammography ranges between 82 per cent
and 97 per cent.

1.4   Computer-Aided Detection Program

A computer-aided detection (CAD) program is a system that incorporates ad-
vanced pattern recognition and image analysis techniques to detect abnormalities
in digitized mammograms. Typically, a CAD program involves several steps. The
first step is image preprocessing. In this step, the pectoral muscle and breast
boundary are usually extracted, and the image is usually filtered or normalized
to improve the quality of the image and to facilitate the next steps. After im-
age preprocessing, isolation to the region of interest (ROI) is usually performed
using different techniques. The ROIs include lesions and suspicious regions that
resemble lesions. The next step is feature extraction. A series of features are
extracted from the ROIs. Based on the extracted features, the final step is the
classification step. The ROIs are classified as malignant or not.

The CAD programs are not designed to replace the radiologists. This is
because, on one hand the CAD programs can not promise that the disease is
always detected, on the other hand, the false positive rate of CAD detection is

normally high. The CAD programs assist radiologists in diagnosing disease by providing detection results to the radiologists as second opinion.

Many studies have shown that CAD programs are effective in detecting abnormalities on mammograms. Freer & Ulissey (2001) performed a study to assess the effect of the CAD program in interpreting mammograms. 12,860 mammograms were used in this study. Performance of radiologists with the assistance of CAD and without the assistance of CAD were compared. The effect of CAD on the recall rate, positive predictive value for biopsy, cancer detection rate, and stage of malignancies at detection were recorded. With the assistance of CAD, this study observed an increase in recall rate from 6.5 per cent to 7.7 per cent, no change in the positive predictive value for biopsy at 38 per cent, a 19.5 per cent increase in the number of cancers detected, and an increase in the proportion of early-stage malignancies detected from 73 per cent to 78 per cent. The result of this study showed that the use of CAD increases the detection of early-stage malignancies without undue effect on the recall rate or positive predictive value for biopsy.

Brem, Baum, Lechner, Kaplan, Souders, Naul & Hoffmeister (2003) evaluated radiologist detection of breast cancer with the assistance of a computer-aided detection system. This study consisted of 377 mammograms which had no malignant findings previously but had cancer diagnosed later. The mammograms were reviewed again by radiologists but with the assistance of CAD. The study found that the use of CAD program significantly improved the detection of breast cancer by increasing radiologist sensitivity by 21.2 per cent.

In the study conducted by Gur, Sumkin, Rockette, Ganott, Hakim, Hardesty, Poller, Shah & Wallace (2004), 115,571 mammograms were interpreted by 24 experienced radiologists with and without the assistance of CAD. However, the study found no significant changes in recall and breast cancer detection rates with and without the help of CAD programs.

Brem, Hoffmeister, Zisman, DeSimio & Rogers (2005) assessed the ability of a CAD system to detect biopsy-proven breast cancers, particularly tumors measuring 1 cm or less. This study included 201 breast cancer patients, with a total of 122 masses, 54 microcalcifications, and 25 mixed mass-and-microcalcification lesions. The CAD program found 84 per cent of the masses, 98 per cent of the microcalcifications and 92 per cent of the mixed mass-and-microcalcification lesions. In this study, the overall sensitivity of the CAD program was found to be 89 per cent.

Studies have shown that mammographic interpretation by two radiologists improves breast cancer detection rate by 5 to 15 per cent. However, double reading requires additional resources. Also with the increase in the complexity of mammographic interpretation and the number of images per study, the radiologists are facing increasing working load. Single reading with CAD was suggested in the aim of freeing up radiologists' time while achieving similar accuracy as

double reading. Comparisons between single reading with CAD and double reading were explored in many studies. Bennett, Blanks & Moss (2006) reviewed 8 studies toward the comparison of single reading with CAD and double reading. 6 of these 8 studies reported results of comparisons of both sensitivity and specificity. Of these 6 studies, 3 showed no difference in both sensitivity and specificity, one showed that single reading with CAD had a higher sensitivity at the same specificity and one showed a higher specificity of single reading with CAD at the same sensitivity. One study, in a real-life setting, showed that single reading with CAD had a higher sensitivity but a lower specificity. Bennett et al. (2006) concluded that most of the reviewed studies suggested that under the most difficult test conditions where interval cancers and masses only were used, there was limited evidence to show that single reading with CAD did not perform equally well as double reading. Under less stringent test conditions with screen-detected cancers seeded with normal films and with high sensitivities reported with single reading, little evidence of a difference between double reading and single reading with CAD was found. However, as most of these studies were not based on real-life conditions, Bennett et al. (2006) found that the evidence to the accuracy of a single reading with CAD in comparison with double reading was limited. In another study (Gilbert, Astley, McGee, Gillan, Boggis, Griffiths & Duffy 2006), 10267 mammograms obtained in women aged 50 years or older were used. All images were double read previously. Eight different radiologists were asked to read the mammograms again but with CAD. The study found a significantly higher cancer detection rate of single reading with CAD than double reading with 6.5 per cent more cancers detected by means of single reading with CAD. However, the recall rate of single reading with CAD was also higher than double reading (8.6 per cent vs 6.5 per cent).

1.5 Motivation and Approach

When interpreting mammograms, radiologists usually compare multiple mammograms to diagnose the abnormalities. These multiple mammograms can be bilateral mammograms, temporal mammograms or MLO-CC mammograms (Fig. 1.1). Bilateral mammograms are the same view mammograms of left and right breast taken for the same women in the same screening session. Temporal mammograms refer to the same view mammograms of the same breast taken in different screening session. MLO-CC mammograms are the MLO and CC view mammograms (see Section 1.2 for MLO and CC) taken for the same breast in the same screening session.

Comparison of the multiple mammograms provides crucial information regarding the likelihood of breast cancer. Asymmetries between the bilateral mammograms are important hints of possible abnormalities. Differences between temporal mammograms help to understand changes in the breast and help to diag-

nose malignancies. For example, changes in sizes of the masses are good signs of malignancies, without which the masses in a single mammogram might only be identified as benign. Also, a bright region appearing in current mammogram but not in previous mammogram may indicate a new malignant mass. These malignancies can only be diagnosed by comparing temporal mammograms. A mammogram is a two dimensional (2-D) projection of a three dimensional (3-D) breast. The projection of a mass in the breast may overlap with that of other objects, showing in the mammogram an area of blended density region, and making it difficult to distinguish the mass from the blended density region. Comparison of MLO-CC mammograms helps to identify these masses.

Multiple mammograms contain important information toward diagnosing the abnormalities. Automating the process of analyzing this information could play an important role in CAD. Hadjiiski, Sahiner, Chan, Petrick, Helvie & Gurcan (2001) compared the performance of CAD program with and without using the temporal mammogram information. They found that by using the temporal information, the A_Z value (area under the ROC curve) was significantly increased from 0.82 to 0.88. However, CAD programs currently have not been able to utilize the information contained in multiple mammograms effectively because of severe difficulties in registering the multiple mammograms. The registration problem is compounded by natural changes of the breast over time not related to cancer, differences in equipment used for image acquisition over time, differences in the positioning of the breast, and inconsistent distortion of breast morphology at acquisition.

Many researches have investigated the use of multiple mammograms to facilitate the computer aided detection of breast abnormalities. According to different types of multiple mammograms that are focused on, these studies branch into temporal, bilateral and MLO-CC mammogram studies. However, the division is not strict, some methods claim that they can be applied to different types of multiple mammograms.

Many attempts have been made in registering multiple mammograms. Some early traditional methods register multiple mammograms by trying to simulate rigid or affine deformations between multiple mammograms. However, both natural changes occurred in the breast and unpredictable motion of breast tissue under varying amount of compression put on the breast during screening cause nonrigid changes between multiple mammograms. Also many methods use simple subtraction to analyze the difference between multiple mammograms. For these methods to work, the mammograms need to be highly consistent in both geometry and intensity. Many methods register multiple mammograms by transforming one to the other by using interpolation based on selection and matching of a set of points of interest (also called control points). Many of these methods select points of interest both from the breast boundary and interior breast. The success of these methods heavily depends on the selection of the set of points of interest. However, reliable and consistent points of interest are difficult to extract.

This thesis presents a new approach in registering temporal mammograms. The proposed method registers temporal mammograms by establishing correspondence directly between the mass-like objects in temporal mammograms. The novelty of this method lies on the use of spatial relations between the mass-like objects of single mammogram to establish correspondence between temporal mammograms and the use of graph matching to establish the correspondence. Under compression, motion of breast tissue is unpredictable and thus the relative positions of objects of interests are not necessarily consistent in consecutive images. However, in most cases spatial relationships between salient regions remain fairly consistent (Novak 1988). In this method, during the graph matching, constraints put on the difference of spatial relations allows a certain degree of variance of spatial relations of mass-like objects between temporal mammograms. The breast boundary is also used in this method to provide global reference for the mass-like objects. Unlike most of the methods that rely on the alignment of breast boundaries of multiple mammograms to provide global registration, boundaries are treated as mass-like objects and the spatial relations between the boundary objects and the other mass-like objects are used. This is made possible by using a fuzzy representation of spatial relation.

In this method, the mammograms were segmented by using an adaptive pyramid (AP) algorithm. The segmentation produced both segmented components and a graph whose nodes representing components of the mammogram and edges connecting the nodes representing neighborhood relationships between the corresponding components. The components were filtered to remove unwanted ones based on the breast boundary and a series of component features. Spatial relations between the remaining components were expressed using fuzzy spatial relation representation. A weighted complete graph was constructed for each mammogram with each node in the graph representing a remaining component, a mass-like object. Fuzzy spatial relations between the mass-like objects were used as weights that associated with the edges connecting corresponding mass-like objects. Correspondence between the mass-like objects in temporal mammogram pairs was established by finding the common subgraph isomorphism of the corresponding graphs using graph matching.

As only spatial relations between mass-like objects are used during the graph matching, rigid changes of mammogram, slight differences in shape and size of masses and different size of breast do not affect the matching results. Thus the proposed method accommodates both rigid and certain degree of nonrigid deformation of mammogram.

Although this study focus mainly on temporal mammograms and the proposed method was mainly developed based on the temporal mammograms, there is no particular restrictions on the type of mammogram pairs. Thus this method can also be applied to both bilateral and MLO-CC mammograms. However, as the mass-like objects in MLO-CC mammograms may have very different spatial relations due to the projection of 3D objects to 2D images in different direction,

the MLO-CC mammograms might not be suitable for the proposed method, since it was based on the spatial relations of mass-like objects.

Registration of multiple mammograms was the first step toward multiple mammogram analysis. Based on this work, it might be possible to track the changes of masses or reveal new masses by further analyzing the matched mass-like objects and unmatched mass-like objects. However, this part of work was not included in this thesis. As a more direct application, in this study, the correspondence of mass-like objects was used to reduce the false positive detections. When comparing temporal mammograms, radiologists usually try to find some corresponding similar structures in temporal mammogram pairs first. The correspondence established in this work can be used as a second opinion for the radiologists.

1.6 Overview of the Thesis

This thesis is organized as follows. The current chapter introduces some background of breast cancer and mammography. Motivation and approach of this thesis are also included in this chapter. Chapter 2 provides technique background and literature review on temporal mammogram registration, graph matching and pyramid based segmentation algorithms. An implementation of the AP segmentation algorithm is explained in Chapter 3. Robustness of the AP algorithm in terms of image deformation is also studied in this chapter. Extraction of the pectoral muscle and the breast boundary are introduced in Chapters 4 and 5. Chapter 6 introduces mass features that are used in this thesis. Details of the graph matching method is presented in Chapter 7. Finally, experiments, discussion and conclusions are made in Chapters 8 and 9.

Chapter 2

Technical Background and Literature Review

This chapter provides technique background and reviews research in the field of temporal mammogram registration, graph matching and pyramid based image segmentation algorithms. Many valuable methods have been proposed in literature for registering sequential mammograms. Most of these techniques register one mammogram to the paired temporal mammogram by means of interpolation or warping based on a set of selected and matched points of interest. One method matches salient regions between temporal mammogram pairs directly. These methods are reviewed in section 2.1. In this thesis, weighted complete graphs are used to represent the mass-like objects of the mammograms and the relations between them. Correspondence between mass-like objects of mammograms are established by finding a common subgraph between two graphs of two mammograms using graph matching. Section 2.2 reviews the literature of graph matching. In mammogram analysis, segmentation of mammograms always plays one of the most important parts and is the basis of mammogram analysis. In this thesis, the AP algorithm is implemented to segment the mammograms. A literature review of pyramid methods is given in section 2.3.

2.1 Temporal Mammogram Registration

Mammogram registration is a challenging task. The difficulty of mammogram registration sources from the high complexity of the images themselves. There are usually no clear landmarks in a mammogram. Signs of disease in a mammogram are usually subtle. Temporal mammograms, even taken in a short period of time, may differ considerably. The difference is compounded by natural changes of the breast over time not related to cancer, differences in equipment used for image acquisition over time, differences in the positioning of the breast, and inconsistent distortion of breast morphology at acquisition.

Many methods have been proposed for registering and comparing mammograms. Some early traditional methods register multiple mammograms by trying to simulate rigid or affine deformations between multiple mammograms. However, both natural changes occurred in the breast and motion of breast tissue under varying amount of compression put on the breast during screening cause nonrigid change between multiple mammograms. Many methods analyzed the difference of multiple mammograms by simple subtraction. For these methods to work, the mammograms need to be highly consistent in both geometry and intensity. As mammograms are highly complex images, consistency of geometry and intensity seldom exists in multiple mammograms. Many methods register multiple mammograms by transforming one image to the other using interpolation and based on selecting and matching of a set of points of interest. The success of these methods heavily relies on the selection of points of interest. Points of interest are often selected both from the breast boundary and interior breast. The boundary of the breast is one of the distinct landmarks of the mammogram and it is relatively easy to extract. The breast boundary provides important information about the deformation of the breast. Alignment of breast boundaries of the mammograms provides partial global registration and thus was adopted as an important step in many methods. To imitate the nonrigid deformations between the multiple mammograms, points of interest from the interior of the breast are necessary when registering one mammogram to the other using interpolation. However, it is difficult to extract reproducible or consistent points of interest. This is because, on one hand, reproducible points of interest may not exist due to the changing organic nature of the breast, the varying conditions in obtaining the mammogram, and the 2-D projection nature. On the other hand, the mammograms are complex images and it is sometimes difficult to extract representative points of interest. The following paragraphs summarize some work that has been done in registering and analyzing temporal mammograms.

Vujovic & Brzakovic (Oct. 1997) extracted points of interest by analyzing elongated structures in the breast. Correspondences between mammograms were then established by matching the extracted points of interest of mammograms using an accumulator matrix and based on identified reference points and signatures that capture local pattern characteristics. Twenty-nine mammogram pairs were used in this study to test the proposed method. To evaluate the results, 3 observers, one of whom was a radiologist, were asked to visually identify the points in the newer mammogram corresponding to the highlighted points in the older mammogram. Performance of the algorithm was measured based on the Euclidean distance between the observer identified points and the algorithm matched points. An average of 9 pixel distance was found between the points determined by the radiologists and the algorithm.

Sallam & Bowyer (1999) proposed a method to register mammograms in two stages. In the first stage, points of interest were extracted along the breast boundaries and were aligned. An initial unwarping function was calculated by

interpolating aligned points based on thin-plate spline. In the second stage, a feature selector based on steerable filters was used to select interest points from interior of the breast region. These points of interest were matched by imposing a locality constraint on the points based on the image alignment done in the first stage. The matched points of interest were combined with those formed in the first stage and were together used to construct a thin-plate spline interpolation which unwarped one image to the other to produce a registered image. Finally a difference image was generated by simply subtracting the registered image from the compared image. The difference image was thresholded to extract suspicious regions.

Kok-Wiles, Brady & Hignam (1998) used a saliency metric to pick up the salient regions and used a nested tree structure to organize the salient regions. Correspondence between the regions of the compared mammograms were made by using a constraint-based region matching algorithm. The algorithm used 4 constraints; a smooth motion constraint; a topological constraint; an order constraint; a monotonicity of geometry constraint; and proceeded coarse to fine. The method was tested on 2 temporal pairs and further tested on 26 bilateral pairs. The results were claimed to be successful.

Sanjay-Gopal, Chan, Wilson, Helvie, Petrick & Sahiner (1999) proposed a regional registration technique. In this method, breast boundaries of the temporal mammograms were first aligned based on the nipples manually identified by the radiologist. The centroid of breast region was then estimated and a nipple-centroid axis was defined for each mammogram. Next the suspicious regions were automatically segmented in the current mammogram using a density-weighted contrast enhancement (DWCE) technique. A polar coordinate system was defined using the nipple as the origin and the previously defined nipple-centroid axis as the $0°$ axis on both images. For each suspicious region in current mammogram, a search for the matching region was performed in a fan-shaped region in the previous mammogram. This fan-shaped region was defined based on the established polar coordinate systems. The method was tested on 74 temporal mammogram pairs. Evaluation of the accuracy of the registration was based on a bounding box that was identified by the radiologist on the previous mammogram which enclosed the corresponding region. The results were that 85% of the cases had the centroid of the extracted matching masses inside the bounding boxes, and the average distance between the centroid of the extracted matching masses and the center of the bounding boxes was 2.8 ± 1.9 mm.

Mutual information (MI) is a measure of how much information a image contains about a second image. Concretely, MI is a cost function used to measure the similarity of images. MI has been used extensively in image registration. Typically, by optimizing MI, a transformation is obtained that optimally aligns the images. Wirth, Narhan & Gray (2002) investigated the usage of MI in mammogram registration. An initial global registration was first performed by using MI. Usually only rigid or affine transformations was used in optimizing MI. To simu-

late nonrigid changes between multiple mammograms, Wirth et al. divided the mammogram into smaller sub-images and matched each sub-image independently to imitate localized nonrigid change. Points of interest were then extracted from the matched sub-images and used with thin-plate splines to register one mammogram to the other. The method was tested on mammograms from the MIAS digital mammogram database (Suckling, Parker, Dance, Astley, Hutt, Boggis, Ricketts, Stamatakis, Cerneaz, Kok, Taylor, Betal & Savage 1994) and the results were reported as promising.

Engeland, Snoeren, Hendriks & Karssemeijer (Nov. 2003) compared 4 mammogram registration methods, which were an alignment method based on nipple location, an alignment method based on the center of mass of the breast area, a method based on MI and a method based on warping. Performance of these 4 methods was measured based on the distance between the center of manually identified abnormalities in the previous and current view before and after registration. The MI based method was found superior to the others. The benefit of this method over the other methods was due to the use of the internal structure of the breast, instead of just one or several points in the breast contour, and the capacity for allowing rotation, scaling, and shearing. The method based on the alignment of mass center worked reasonably well, especially when the pectoral muscle was excluded when calculating the center of the mass. The method based on the alignment of the nipple was found to work only when the nipple was visible in the profile. The warping based method was found to be the worst.

Marias, Behrenbruch, Parbhoo, Seifalian & Brady (2005) proposed a method to register sequential mammograms using a similar schema as used in Sallam & Bowyer (1999). The method was also based on a set of points of interest which were also selected both from the boundaries of breast and from interior of the breast. In this method, the points of interest from the breast boundaries were selected based on curvature analysis of the boundaries. The boundaries of the mammograms were then aligned based on the selected points of interest using thin-plate interpolation. To select the points of interest from interior of the breast, first, ROIs were isolated using a nonlinear wavelet scale-space. The isolated regions of the mammograms were matched based on the aligned breast and a "search" window. Interior points of interest were then defined as the centroid of the matched regions. All boundary points of interest and interior points of interest together controlled a thin-plate spline approximation technique that was used to register one mammogram to the other. Results of this method were evaluated on 50 temporal pairs and an additional 16 temporal pairs that had annotated masses both quantitatively and visually. The results were visually judged by an expert clinician and were found "good" in 70% of the cases, "average" in 25% of the cases and "poor" in 5% of the cases. The Euclidean distance between clinician selected landmarks in a mammogram and its registered paired mammogram was calculated to quantitatively measure the performance of the method. The average and standard deviation of the distance after the final registration were found to

be 3 and 2.7 mm.

Timp & Karssemeijer (2006) combined global and regional registration to analyze the interval changes between temporal mammograms. The global registration was realized by aligning the images based on the center of the breast area. After the global registration, regional registration was performed by finding a corresponding region in the previous mammogram for each suspicious region in the current mammogram. The coordinates of the suspicious regions were used as initial locations of the corresponding regions in the previous image, since the two images were globally registered. Analysis of the interval changes was made by analyzing the corresponding regions based on a series of features. The method was tested on 2873 temporal mammogram pairs and two measures, a correlation measure from Sanjay-Gopal et al. (1999) and a cost measure, were used to assess the matched regions. A highest number of 72% was reported for the correctly matched regions.

Martí, Raba, Oliver & Zwiggelaar (2006) proposed a points of interest based registration method that used similar process as used by Sallam & Bowyer (1999) and Marias et al. (2005). In this method, points of interest were also extracted both from breast boundary and internal regions. The points of interest were matched using a point matching algorithm based on a cost matrix. The points were then transformed with an affine transformation followed by a thin-plate spline based transformation. The method was evaluated both qualitatively and quantitatively. The results were also compared with 2 other methods and was reported being better.

Most of the methods described above are pixel level registration. Although Sanjay-Gopal et al. (1999) and Timp & Karssemeijer (2006) find matched regions between mammogram pairs, they are based on pixel level searching in the previous mammogram for a region to match the region in the current mammogram. Because of the three dimensional deformation of the tissue at acquisition and all the other differences in image acquisition mentioned previously, the same object in one mammogram may be seen at a different angle, superimposed on a different background, and have different size or shape than in another image, even if the second image is nominally the same view (both MLO, for example). Accordingly, the object cannot be expected to be represented by the same set of pixel intensities in sequential images. For this reason, the registration in this thesis is based entirely on matching mass-like objects.

2.2   Graph Matching

Graphs are versatile data structures that are used to represent structural information. Graph representation has been widely used in different domains such as computer vision, scene analysis, chemistry, molecular biology, image analysis, pattern recognition, and many others. In image analysis, typically a graph can

be used to represent the content of an image with the vertices of the graph representing regions of the image and the edges connecting the vertices representing the relations between the regions.

To analyze and compare the information, an important task is to find a similar or common structure between them. When graphs are used to represent the information, the task becomes that of finding a similar or common subgraph between the graphs (For a definition of subgraph, see Section 3.1). In image analysis, finding a common subgraph of two graphs representing two corresponding images helps to compare and recognize the objects in the images. The process of finding a common subgraph within two graphs is called **graph matching**.

According to the difference in the possibly matched subgraphs, there are two categories of graph matching, **exact graph matching** and **inexact graph matching**. When a graph isomorphism exists between two graphs, and the goal is to find this graph isomorphism, the corresponding problem is called **exact graph matching**. Here a **graph isomorphism** is a one-to-one mapping between the vertices of two graphs such that if there is an edge connecting two vertices in one graph, then the vertices in the other graph corresponding to these two vertices are also connected by an edge (see Section 3.1). Due to the difference in the size, edges or other factors, two graphs may not be isomorphic. If, however, the smaller graph is isomorphic to a subgraph of the larger graph, the corresponding graph matching is called **exact sub-graph matching**. Exact sub-graph matching is also classified as exact graph matching. In many applications, it is not possible to identify an isomorphism between two graphs to be matched. This is because of the different number of vertices, different edges or different attributes. In these applications, the aim of graph matching is not to find an exact isomorphism, but to find the best matches. This type of graph matching is called **inexact graph matching**. More specifically, the goal of inexact graph matching is to find two subgraphs of two graphs that are isomorphic to each other. Here the best match is defined according to particular application. It can be the common subgraph that contains the maximal number of vertices, or the common subgraph that maximizes or minimizes an objective function.

Graph matching is known to be a computationally expensive process. In fact, both exact sub-graph matching and inexact graph matching have been proved to be NP complete problems (Abdulkader 1998, Garey & Johnson 1979). The complexity of exact graph matching has not been classified yet and remains to be an opening theoretical problem. However, for some type of graphs, the complexity of exact graph matching has been proved to be of polynomial complexity (Hopcroft & Wong 1974). Although some methods were developed to deal with some particular large graphs, it is generally difficult or even impossible to find solutions for large graphs. Most of proposed graph matching algorithms aim at small or medium size graphs (no more than 1000 nodes).

Many algorithms have been developed focusing on different types of graph matching problems (Ullmann 1976, Krissinel & Henrick 2004, McGregor 1982).

Some of these algorithms aim at finding the optimal solutions, they are optimal algorithms. These algorithms usually require exponential running time and space due to the NP completeness of the problem. Other algorithms target the nearly-best solutions with relatively low computational costs. They are suboptimal, or approximative algorithms. Running time for these algorithms is usually polynomially bounded. However, these algorithms may fail to find the optimal solutions.

A well known algorithm, addressing the exact graph matching problems, was proposed by Ullmann (1976) (UA) in 1976. It is widely accepted that the exact graph matching was much more effectively solved due to this algorithm. Although the algorithm mainly focused on the exact graph matching, the author also suggested a way to employ it for the maximum common subgraph matching. Briefly, the algorithm used the backtracking algorithm to enumerate all possible matches to find the best solution. A binary vertex matching matrix M was used in the algorithm with each entry $m_{ij} = 1$ meaning that vertex i of the first graph is matchable to vertex j of the second graph. To prune unnecessary searching branches during the enumeration, a so-called *refinement procedure* was used to refine the matrix M in each step. This procedure was based on the following condition: if a vertex v_α of the first graph is mappable to a vertex v_β of the second graph, then for each vertex $v_{\alpha i}$ in the first graph that is adjacent to v_α, there must exist a vertex $v_{\beta j}$ of the second graph which is adjacent to v_β, such that $v_{\alpha i}$ and $v_{\beta j}$ are mappable. Although proposed in 1976, the algorithm is still widely used today and is probably the most popular algorithm in graph matching.

Many modifications have been made to the backtracking schema provided in UA aiming at finding "better" solutions and saving more computational time by rejecting more unsuitable search branches of the search tree. Among them, McGregor (1982) proposed an efficient modification which was used to match the structure of molecules. In his method, each vertex i of the first graph G_1 was associated with a "priority subset" of vertices of the second graph G_2 to which vertex i can correspond. During the graph matching, vertices in the "priority subset" were always considered first to match the corresponding vertex of G_1. Whenever a match was formed during the graph matching, all subsets of unmatched vertices of G_1 were updated. The update was based on the condition that vertex j of G_2 was in the subset of vertex i of G_1, if in the current partially formed matches, all the vertices of G_1 that are adjacent to vertex i were mapped to the vertices that are adjacent to vertex j. The method was found effective when matching chemical molecules present before and after a reaction.

Krissinel & Henrick (2004) also proposed a version of modification to UA. The proposed method also uses the backtracking schema to recursively enumerate all possible mappings that satisfy the subgraph isomorphism condition. In each step of the process, a procedure called *Extendable* was first performed to see if the currently formed solution can still be extended. The process was continued if *Extendable* returned true. A *PickVertex* procedure was then used to

pick a vertex from G_1 for matching and a *GetMappableVertices* was used to select a set of vertices from G_2 to map the selected vertex of G_1. In many graph matching algorithms, an earlier formed "better" solution can help to cut off many search branches and thus improves the efficiency of graph matching sharply. The processes *PickVertex* and *GetMappableVertices* enable combinations of different factors to be taken into account when picking vertices to match. In the algorithm, *PickVertex* picked the vertex that has minimal number of possible matches in G_2 and *GetMappableVertices* returned all the vertices in G_2 that were matchable to the picked vertex of G_1. Krissinel & Henrick also analyzed the complexity of the algorithm in terms of both time and space. The time complexity of the algorithm was found to be between $o(mn)$ and $o(m^{n+1}n)$ and the space complexity was found to be $O(mn^2)$. The algorithm was found superior to the traditional maximal clique approaches which have time complexity between $o(mn)$ and $o((mn)^n)$ and space complexity $O((mn)^2)$.

For more information about graph matching algorithms, the readers are referred to Jolion (2003), Bunke (2000) and Conte, Foggia, Sansone & Vento (2004).

In multiple mammogram analysis, graphs representing multiple mammograms are not expected to be of the same size. It is also unlikely that a graph representing a mammogram is isomorphic to a subgraph of a graph representing another mammogram. Thus graph matching in this field can only be inexact graph matching. Graph matching in mammogram analysis is further compounded by the fact that, unlike other graph matching applications, the maximum common subgraph might not be the best solution, since it may contain false matches, which is not better than a shorter solution with more true matches.

2.3 Pyramid Based Image Segmentation

Graph pyramids were introduced by Tanimoto & Pavlidis (1975) and have been applied widely in image processing since then. A graph pyramid is a stack of successively reduced graphs. At each level in the stack, the graph is a reduction of the graph at the previous level. A vertex of a graph at one level is connected to a number of vertices at the previous level. The vertex in the higher level is called the ***parent*** of the vertices in the previous level and the vertices to which the parent is connected in the previous level are called the ***children*** of the vertex. By transitivity of the "parent-children" relation, a vertex of a graph at any higher level is connected to a set of vertices in the bottom level graph. This set of vertices is called the ***receptive field*** of the vertex. If the bottom level graph is an image with each vertex in the graph corresponding to a pixel of the image, then the receptive fields of the vertices of a graph in a higher level constitute a segmentation of the image, with a receptive field of a vertex comprising a segmented component. Thus the collection of graphs forms a multi-resolution description of the image, but unlike multi-resolution representations

via wavelets or filter banks, the connectivity between layers provides a vehicle for tracking information from disparate components of the image. By tracing along the "parent-children" relations, as Jolion & Montanvert (1992) described, "a global interpretation is obtained by a local evidence accumulation".

The bottom graph is not necessary an image, in fact, it can be anything, for example, it can be symbolic information or derived image properties.

A pyramidal structure is defined if the following three steps are specified which together define how a new $l+1$ level graph $G_{l+1} = (V_{l+1}, E_{l+1})$ is derived from the l level graph $G_l = (V_l, E_l)$:

1. The selection of vertices V_{l+1} from V_l.
   The selected vertices from V_l are named the surviving vertices while the unselected vertices are named non-surviving vertices.

2. The connection of non-surviving vertices to the surviving vertices.
   This step defines the parent-children relationships between the corresponding two levels of graph pyramid.

3. The definition of the adjacency relationships between vertices of V_{l+1}.
   This step defines E_{l+1} of G_{l+1}.

A graph pyramid is formed by iteratively executing these three steps. Step 1 defines how the vertices of a new level graph are selected out from the previous level. The ratio between the numbers of vertices in level l and level $l+1$ is called **decimation ratio**. If the decimation ratio is large, this means few vertices are contained in $l+1$ level, and thus the image loses its information fast and also the height of the graph pyramid is low. On the contrary, a small decimation ratio results in more levels of the pyramid.

Depending on the structures used in the pyramids, the graph pyramids can roughly be classified as **regular pyramids** and **irregular pyramids** (Fig. 2.1). In a regular pyramid, the structures of the pyramids are rigid. The survivors of a new level graph and the adjacency relations among them are predetermined. The decimation ratio is constant. The size and number of vertices of each level in the graph are fixed and known beforehand. Processing a regular pyramid is relatively convenient. However, drawbacks of regular pyramids have been exposed. Due to the rigid structure, the receptive field of any vertex can only be in a relatively regular shape (The whole region might not be regular, but it is composed by smaller regular shaped regions). Thus regular pyramids can not express irregular shaped objects well. Regular pyramids are also reported to suffer from shift variance (SV, see Chapter 3). Irregular pyramids, on the other hand, have flexible structures. Receptive fields of vertices can be irregular shapes and thus can be closer to the real shapes of objects. Survivors of a new level graph in an irregular pyramid are not fixed and different factors can be combined into the selection of the survivors. For irregular pyramids, the decimation ratio is not constant,

neither is the height of the pyramids. As survivors are not fixed, the adjacency relations among the survivors need to be defined and computed. The unfixed height of an irregular pyramid and not well-defined adjacency relations increases the computational complexity of irregular pyramids.

Except the normal simple graphs used in irregular pyramids, there are also dual graphs and combinatorial maps (Kropatsch 1998, Brun & Kropatsch 2003). Simple graphs used in irregular pyramids do not allow detection of boundaries between the receptive fields. They also do not allow differentiation of an adjacency relationship between receptive fields. Dual graphs are proposed to resolve the drawbacks of simple graphs. With dual graphs, each level of pyramid is represented by dual pair of graphs. Any operations on the pyramid are computed on the dual graphs. Dual graphs can keep the boundary information of the image, however, with dual graphs, computational complexity is increased. Combinatorial maps can realize similar functionality as dual graph, but improve the efficiency of computation. Combinatorial maps use planar graphs to represent each level of the pyramids. The planar graph encodes the orientation of edges around the vertices.

Many algorithms have been proposed to construct graph pyramids, each having different ways of realizing the three steps. The following paragraphs review work in this field.

(a)   (b)
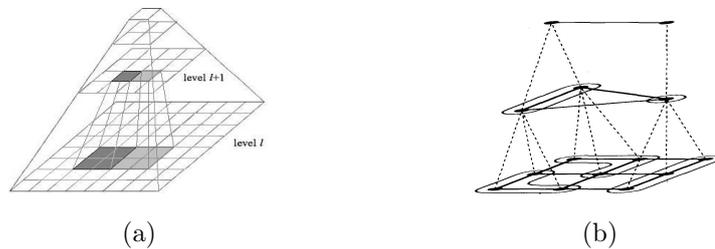
Figure 2.1: Regular and irregular pyramids. (a) depicts a regular pyramid and (b) shows an irregular pyramid.

To overcome the drawbacks of the regular pyramids, Meer (1989) introduced the stochastic pyramid. In the construction of a stochastic pyramid, to maintain a proper decimation ratio, two constraints were imposed on the selection of survivors.

- Two neighbors at any given level can not both be survivors.

- For any non-surviving vertices, there exists at least one survivor in its neighborhood.

To construct a new level graph G_{l+1}, three variables, two binary state variables p_i and q_i, and a random variable x_i uniformly distributed between [0,1], were

used for each vertex v_i in G_l. At the beginning, variable $p_i^{l+1}(1)$ was initialized according to the following rule:

$p_i^{l+1}(1) = 1$ if random variable x_i of v_i was the maximum among v_i's neighbors,
$p_i^{l+1}(1) = 0$ otherwise.

All the state variables were updated iteratively as follows. At the beginning of the kth iteration, variable $q_i^{l+1}(k)$ was initialized as:

$q_i^{l+1}(k) = 1$ if all neighbors of v_i having variable $p^{l+1}(k-1) = 0$,
$q_i^{l+1}(k) = 0$ otherwise.

$p_i^{l+1}(k)$ was then set as follows for all v_i having $q^{l+1}(k) = 1$.

$p_i^{l+1}(k) = 1$ if random variable x_i of v_i was the maximum among v_i's neighbors
         having $q^{l+1}(k) = 1$,
$p_i^{l+1}(k) = 0$ otherwise.

These steps were iterated until no $q_i^{l+1}(n) = 1$ remained. As the whole process relied on random variables x_i, the stochastic pyramid was not reproducible. The stochastic decimation process used in this stochastic pyramid makes it possible to build irregular pyramids. In fact, Montanvert, Meer & Rosenfeld (1991) proposed a method to generate irregular tessellations which was believed to be the first irregular pyramid.

Jolion & Montanvert (1992) proposed a framework to construct an adaptive pyramid (AP). In this framework, a support set was defined for each vertex. The support set S_{ij} of vertex (i, j) is the set of all the neighbors of vertex (i, j). S_{ij} is initialized as the 3×3 neighborhood centered on (i, j). The same two constraints, as used in stochastic pyramid, were also imposed in selection of survivors. A process similar to the one used in the stochastic pyramid, was used to select the survivors of a new level graph. But unlike the stochastic pyramid which used the outcome of stochastic variable x_i, in this adaptive pyramid, Jolion & Montanvert used an interest operator to decide the value of each vertex which was used for comparison in survivor selection. The interest operator is not fixed. Any image characteristic, global or local, can be incorporated into the interest operator. For example, Jolion & Montanvert used the variance of the intensity values within the receptive fields as the interest operator. After the selection of survivors, all non-survivors were linked to one of the survivors in their support sets who were most similar to the non-survivors. During the construction of the pyramid, a component in the original image may have all its pixels identified as non-survivors, thus in the final (highest) level of the pyramid, there is no representative for this component. In order to extract components during the construction of the adaptive pyramid and to leave a representative for each component, a root extraction process was introduced. A non-survivors was identified as a root if it was very different to the closest survivors in its support set and the size of it was large

enough. If a non-survivor was identified as a root, it was retained as a survivor and was kept in the final level of the pyramid.

Many algorithms have been developed based on the stochastic decimation process and the adaptive pyramid. Strategies about survivor selection, linking of non-survivors to survivors, pyramid height reduction, computational efficiency, control of region growing, have been proposed. Among them, Lallich, Muhlenbach & Jolion (2003) proposed a test to control the decimation process both locally and globally. The test was based on Moran's spatial autocorrelation coefficient.

Marfil, Molina-Tanco, Bandera, Rodríguez & Sandoval (2006) systematically reviewed pyramid based segmentation algorithms, including both regular and irregular pyramids. A comparison between two regular pyramid based algorithms and 6 irregular pyramid based algorithms was also performed in this work. The SV, a F function and a Q function were used to evaluate the performance of these selected pyramid algorithms. The comparison verified that the classic regular pyramids suffer from shift, rotation and scale variance.

Application of graph pyramids has been extensive and graph pyramids have proved to be useful in analyzing images. However, Bister, Cornelis & Rosenfeld (1990) found that when images were slightly shifted, rotated or re-scaled, segmentation results produced by regular pyramids were changed. To verify the shift, rotation and scale variance of regular pyramids, Bister et al. tested several regular pyramid based segmentation algorithms on eight images and their different shifted, rotated and re-scaled versions and verified the difference of the segmentation results.

Chapter 3

Mammogram Segmentation using Adaptive Pyramid Algorithm

In this chapter we describe our implementation of the AP algorithm (Jolion & Montanvert 1992) to segment mammograms. A segmentation method based on graph theory was chosen for this study for three reasons. First, graph theory methods form a flexible class of techniques capable of incorporation both local and global image information. Second, the graph structure obtained during segmentation is conjectured to provide information useful for registration. Third, while several mature methods exist for detecting masses or clustered microcalcification in mammograms, the objective here is not to identify specific manifestations of cancer but to summarize the information content of the image. This information content should include masses and calcifications but also structures such as the nipple, the pectoral muscle, regions of dense tissue, patches of anomalous texture, etc.

Graph pyramids are powerful tools for representing images. Based on graph theory, the flexible structures of graph pyramids make it convenient to manipulate the graph pyramids, and thus to manipulate the content of the images. AP provides a way to construct irregular pyramids and is relatively convenient to implement. Objects, such as masses, often appear as fuzzy bright regions and many times there are no clear boundaries around these bright regions. Thus dual graphs and combinatorial maps, which are capable of preserving boundary information, are not considered in this thesis.

A certain degree of deformation normally happens between temporal mammograms. This can be caused by different positioning of breasts during examination and different pressure put on the breasts. A shift variant segmentation algorithm, when used to segment similar but slightly deformed images, such as temporal mammograms, may produce very different segmentation results. When registering multiple images based on these results, the same object in the multiple images may be registered as different objects. Thus the SV of a segmentation algorithm affects the registration. In this chapter, we also investigate the robust-

ness of the primary algorithm in terms of image distortions. Since salient objects are of interest in mammogram analysis, we especially investigate the SV of the segmentation algorithm on the salient objects. For comparison, a minimum spanning tree (MST) based segmentation, previously developed by our research team (Ma, Bajger, Slavotinek & Bottema 2007, Bajger, Ma & Bottema 2005), is also used.

This chapter starts by first introducing some basic notation and terminology of graph theory in Section 3.1. Details of the implementation of AP and the tunning for mammogram segmentation are described in Section 3.2. Section 3.3 presents a component merging process to merge over-segmented components. The robustness of the segmentation algorithm with respect to image shifts, rotations, and warping is evaluated in Section 3.4.

3.1 Basic Notation and Terminology

A **graph** is a 2-tuple $G = (V, E)$, where $V = \{v_i, i = 1 \ldots n\}$ is a non-null set of vertices, $E = \{e_{ij}\} \subset V \times V$ is a set of edges with edge e_{ij} connecting vertex v_i and v_j. An edge e connecting vertices u and v can also be denoted as $e = (u, v)$. Two vertices u and v are said to be neighbors or adjacent if they are connected by an edge. The size of a graph G is defined as the number of vertices of G and is denoted as $|G|$.

If edges of a graph G have no direction, G is called **undirected graph**. In an undirected graph, edge $(u, v) = (v, u)$. When edges in a graph are directed, thus edge (u, v) and (v, u) are distinguished, the corresponding graph is called a **directed graph**.

A graph is called **complete** if every two vertices in this graph are connected with an edge.

When using a graph to represent information, vertices and edges of the graph are usually associated with information. A graph having attributes attached to its vertices or edges is called **attributed graph** or **weighted graph**.

A **subgraph** of graph $G = (V, E)$ is a graph $H = (W, F)$, denoted as $H \subseteq G$, if $W \subseteq V$ and $F = E \cap (W \times W)$. A graph is a subgraph of itself.

A **graph isomorphism** is a bijection (a one-to-one and onto mapping) between the vertices of two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$,

$$f : V_1 \to V_2,$$

with the property that $e_{u,v} \in E_1$ if and only if $e_{f(u),f(v)} \in E_2$. A **common subgraph isomorphism** (CSI) between graphs G_1 and G_2 is a pair of graphs H_1 and H_2, such that $H_1 \subseteq G_1$, $H_2 \subseteq G_2$ and H_1 and H_2 are isomorphic to each other.

3.2 Segmentation by AP

The AP algorithm segments an image into many layers which together form a pyramid structure. In this pyramid structure, any higher layer is a blurred version of the lower layer. The bottom layer is usually the original image. A vertex in a higher layer is connected to one or several vertices in the lower layer and is called the parent of these vertices. The vertices to which the parent is connected in the previous layer are called the children of the vertex. Following the "parent-children" relation, a vertex in the top layer connects to a set of vertices in the bottom layer and so represents a component in the original image.

The AP algorithm implemented in this thesis is based on the framework proposed by Jolion & Montanvert (1992). In this framework, a support is first defined for each pixel. The support of v_i is the set $S_i = \{v_j : (v_i, v_j) \in E\}$. At the base layer, the support of a vertex is its immediate 3×3 neighborhood. Supports evolve with the construction of each layer graph in the pyramid and the supports of the non-survivors are merged into the supports of their selected parents.

Based on the support, an interest operator, h, is used to determine survivor selection (step 1 of the 3 steps listed in Section 2.3). In principle, the interest operator can be chosen to exploit a wide variety of image characteristics. In this study, $h(v_i)$ is the variance of the image intensities computed over the receptive field of v_i (Jolion & Montanvert 1992). At the base level of the pyramid, the receptive field is not defined and $h(v_i)$ is the variance over S_i.

Three variables are involved in selecting surviving vertices; two binary state variables p, q, and the outcome of the interest operator, h. To simplify the notation, h_i, p_i, q_i will denote $h(v_i)$, $p(v_i)$, and $q(v_i)$ respectively. The selection process works in two steps. In the first step, the state variable p_i is set as

$$p_i = \begin{cases} 1, & \text{if } h_i = \min\{h_j : v_j \in S_i\}, \\ 0, & \text{otherwise.} \end{cases} \tag{3.1}$$

In the second step, the state variable q_i is set and some of the p_i are updated according to the following rules:

$$q_i = \begin{cases} 1, & \text{if } p_j = 0 \ \forall \ v_j \in S_i, \\ 0, & \text{otherwise.} \end{cases} \tag{3.2}$$

$$p_i = 1, \qquad \text{if } h_i = \min_{v_j \in S_i}\{h_j : q_j = 1\}. \tag{3.3}$$

A vertex v_i is retained for the next level if $p_i = 1$. The condition in Eq. (3.1) insures that two vertices sharing an edge at one level cannot both survive to the next level. The conditions in Eq. (3.2) and (3.3) insure that for each non-surviving vertex, there is at least one surviving vertex in its support.

To make the connection between non-surviving vertex and the surviving vertex (step 2), a contrast operator is used. A non-surviving vertex v_i will be connected to its surviving neighbor v_j, if and only if

$$|\mu_i - \mu_j| = \min_{v_m \in S_i \text{ and } p_m = 1} |\mu_i - \mu_m|,$$

where μ_i denotes the mean intensities of the receptive fields of v_i.

Whenever a non-surviving vertex v_i is connected to a surviving vertex v_j, S_j is updated by $S_j = S_i \cup S_j$. Thus the new adjacency relationships are formed (step 3).

In the framework proposed by Jolion & Montanvert (1992), a root extraction process is used to detect roots of the components of the original image during the construction of the pyramid. A *root* is a non-surviving vertex that is sufficiently different from its neighbors. More specifically, a non-surviving vertex v_i is called a root if and only if

$$|\mu_i - \mu_j| > F_{io}(A_i) \text{ whenever } v_j \in S_i \text{ and } p_j = 1,$$

where A_i is the number of pixels in the receptive field of v_i and the function F_{io} is defined by

$$F_{io}(x) = \begin{cases} \text{min\_contrast} & \text{if } x > \text{min\_size}, \\ \text{min\_contrast} * e^{\alpha(\text{min\_size}-x)} & \text{otherwise.} \end{cases} \tag{3.4}$$

The purpose of the function F_{io} is to compensate for the fact that in small sets, intensity statistics are not necessarily robust. The parameter min_contrast controls the minimum contrast for assigning roots, while min_size and α determine details of root assignment for vertices with small supports.

When applying AP to mammograms, it was found that no single value of min_contrast was satisfactory. Low values of min_contrast resulted in masses in bright regions with high variance being segmented into several components. High values of min_contrast resulted in subtle masses in uniform regions of the breast being incorporated into large background components. Accordingly, the function F_{io} was replaced in this study by

$$F_{io}(x) = A(x)B(x)F_{io}(x) \tag{3.5}$$

where $A(x)$ and $B(x)$ are defined by

$$A(x) = \begin{cases} 1 & \text{if } x < 1000, \\ (\frac{1}{a})^{x-1000} & \text{otherwise,} \end{cases}$$

and

$$B(x) = \begin{cases} b^{\text{Var}(S_i)} & \text{if } x < 300 \text{ and } \text{Var}(S_i) > 5, \\ 1 & \text{otherwise,} \end{cases}$$

where $\mathrm{Var}(S_i)$ is the variance of S_i and a and b are empirically chosen parameters (in this study, a = 1.0024 and b = 1.1040 for mammograms from Mini-MIAS database (see Sections 3.4.1), and a = 2.1041 and b = 2.0024 for mammograms from a local screening archives (see Sections 8.2.1)). This new $F_{io}(x)$ is designed for the purpose of better segmenting the mass components and at the same time preventing components belonging to the uniform and relatively low contrast areas being merged together. The numbers used in $A(x)$ and $B(x)$ are set based on the characteristics of the masses. Although sizes of mass components vary greatly, they are usually relatively small, and most of them are smaller than 1000. When the size of a component is bigger than 1000, $A(x)$ decreases the $F_{io}(x)$ sharply, preventing large components belonging to relative low contrast areas from being merged. ROIs usually have high intensity contrast and are thus more likely to be segmented into several pieces. These pieces usually have sizes less than 300 and have intensity variances larger than 5. When a component meets these conditions, $B(x)$ increases the $F_{io}(x)$ to facilitate the merging of the components.

If a non-surviving pixel v_i is identified as a root, it will be retained to be a survivor and will appear in the highest level graph. The root extraction process prevents components of the original image from disappearing during the construction of the graph pyramid and insures that each significant component of the original image has a representative vertex in the highest level graph.

The highest level graph is reached when no non-survivor can be selected. All the remaining vertices are roots. Each vertex in this level graph represents a component of the original image. The component represented by a vertex can be recovered by tracing the receptive fields down to the base level.
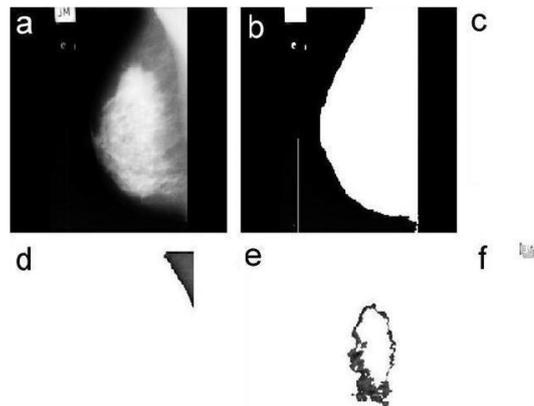
Figure 3.1: Examples of components after initial segmentation by AP. The image used is mdb003 from the Mini-MIAS database (Suckling et al. 1994). (a) The original image. In panels (b) - (f), the shaded region is the segmented component. (b) a non-breast component (background), (d) a component belonging to the pectoral muscle, (f) the image label. (c) and (e) are two components within the breast.

Examples of the components segmented by AP are listed in Fig. 3.1. Examples of mass segmentation can be found in Fig. 3.2 and the following figures.
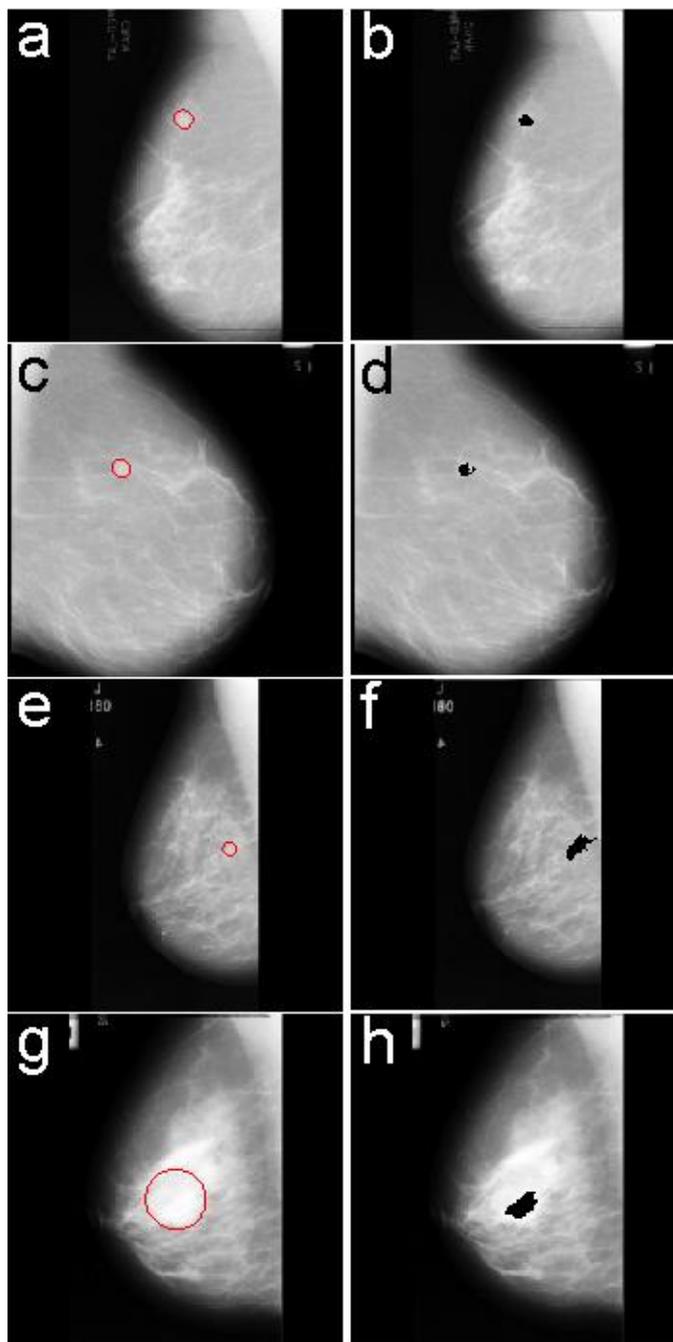
Figure 3.2: Examples of mass segmentation: mini-MIAS images. The images in the left column are the original mammograms with masses marked out by circles. In the corresponding images in the right column, components corresponding to the true mass were selected from all segmented components and were marked out in black. In the first two images ((b) and (d)), the segmented components well covered the masses. In (f), the mass was under segmented and so the corresponding segmented component was bigger than the true mass. In (g), a large area was enclosed in the circle. Compared with the mass in (g), the mass segmented by AP in (h) was over segmented.

3.3   Merging Segmented Components

The modified root extraction process (Eq. 3.5) helps to isolate the high contrast components with more accurate boundaries and at the same time prevents large components belonging to relatively low contrast or uniform areas from being merged together. In our experiments, boundaries obtained by the algorithm for true masses were closer to the true boundaries that were annotated by the radiologists with the use of the modified root extraction process (Fig. 3.2, 3.3). However, conditions of mass components, such as the size, the variance of intensity, are complex and vary greatly. In our experiments, some mass components with high intensity contrast or large size were segmented into several pieces (Fig. 3.4). On the other hand, in this thesis, we select the parameters involved in the segmentation so that ROI with low intensity contrast can be isolated (Fig. 3.3). Thus, the segmentation results tend to be over-segmented. To compensate for this problem, we use a merging process to merge the over-segmented components.
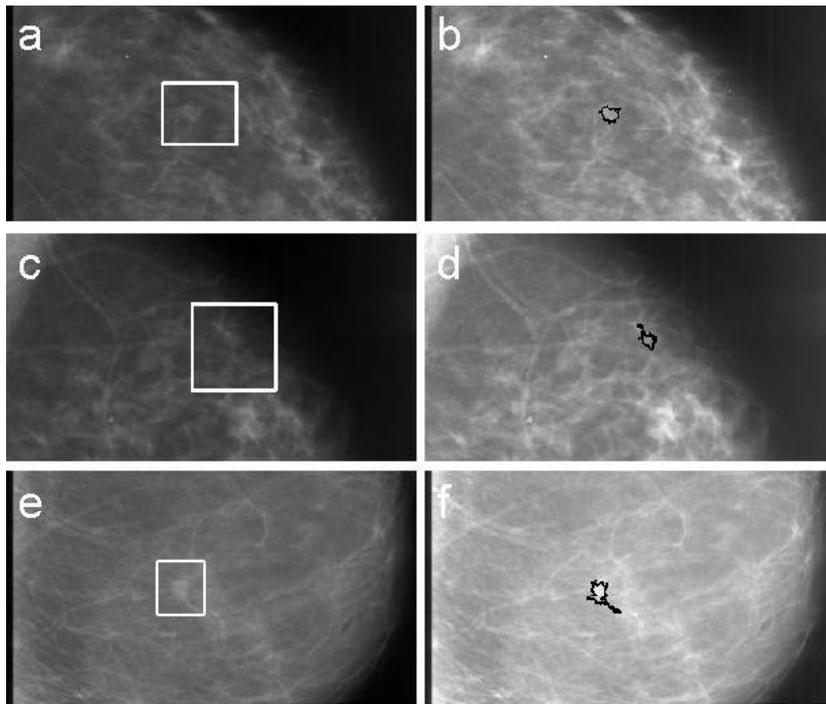
Figure 3.3: Examples of mass segmentation: local images. White boxes in the left column images indicate the radiologist's annotation, black outlines in the right column images indicate the boundaries of components found by the AP algorithm.

The merging process involves a set of criteria which are based on the size, intensity and shape of the components. Masses are usually bright and high contrast components and so are more likely to be segmented into several subtle components. The merging process aims to merge the subtle components belonging to
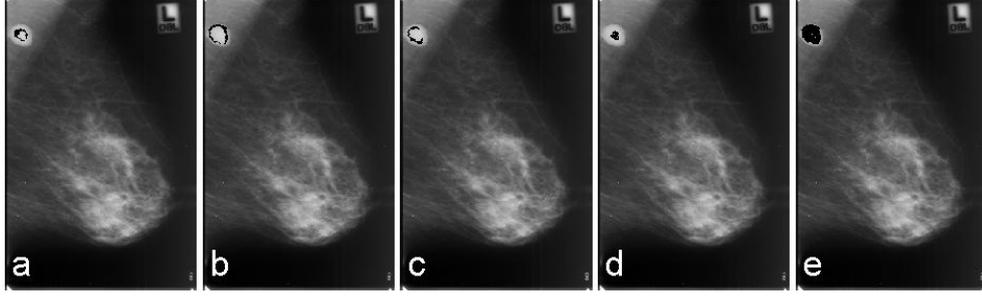
Figure 3.4: Examples of merged components. (a)(b)(c)(d) are four segmented components. After the merging process, (a)(b)(c)(d) are merged together and (e) is the resulting component.

a single ROI. For this purpose, at the beginning of the merging process, a set of filters is used. Components meeting one of the following conditions

- area > 4400

- solidity < 0.56

- $\frac{\text{major axis}}{\text{minor axis}} < 5.6$

are excluded from merging process. These filters are based on the characteristics of the masses of a training dataset (see Section 8.2.1).

Components R_1 and R_2 are merged if they meet the following 4 conditions.

- $\text{area}(R_1) + \text{area}(R_2) < 4400$,

- $|\frac{\text{Var}(R_1)}{\text{area}(R_1)} - \frac{\text{Var}(R_2)}{\text{area}(R_2)}| \leq 0.09$,

- $|\text{E}(R_1) - \text{E}(R_2)| \leq 20$,

- $circu(R_1, R_2) \leq 0.1 * \log(2(\text{area}(R_1) + \text{area}(R_2)))$.

Here $\text{Var}(R)$ is the intensity variance of component R and $\text{E}(R)$ is the mean intensity of R and $circ(R_1, R_2)$ is defined as the proportion of pixels in $R_1 \cup R_2$ that lie in the intersection of $R_1 \cup R_2$ and the disk centred at the centroid of $R_1 \cup R_2$ and having the same area as $R_1 \cup R_2$. Again, the conditions were developed based on a training dataset (see Section 8.2.1). Fig. 3.4 gives an example of merged components.

Beside the 4 conditions above, components satisfying a "surrounded" relation and meeting the area condition are further merged. That is, if component R_1 is surrounded by another component R_2, and $\text{area}(R_1) + \text{area}(R_2) < 600$, then R_1 and R_2 are merged. In the graph naturally generated from AP segmentation, a vertex corresponding to the component surrounded by another component has

only one neighbor which is the surrounding component. Thus based on the graph, the "surrounded" relation can be easily extracted. Fig. 3.5 gives examples of surrounded components.
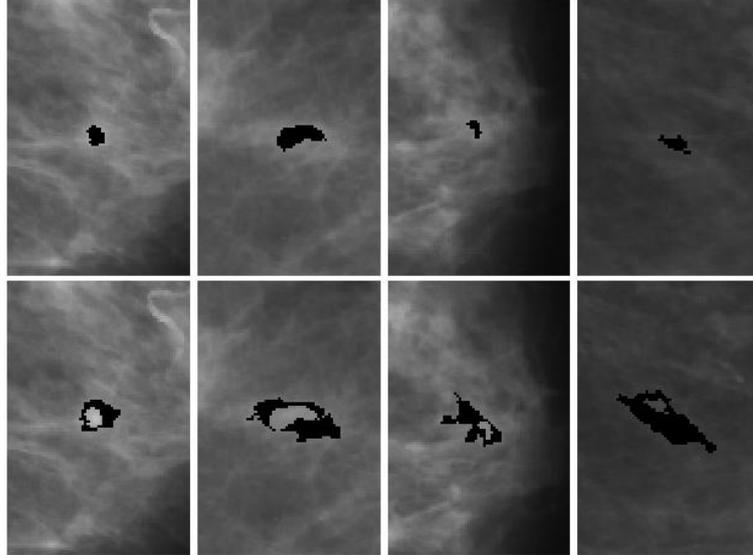
Figure 3.5: Examples of surrounded components. Components in the first row are surrounded by the corresponding components in the second row. The components in the first row are merged to the corresponding components in the second row by the merging process.

3.4 Robustness of AP

SV of regular pyramid based algorithms has been reported and verified (Bister et al. 1990, Marfil et al. 2006). The comparison experiments carried out by Marfil et al. (2006) showed that SV also slightly affects six irregular pyramid based algorithms, though the SV measures are much higher with regular pyramid based algorithms. This section investigates the robustness of the AP algorithm to image shifts, rotations, and warping. The performance of the segmentation method to image distortion is crucial for temporal analysis of screening mammograms where natural changes in the breast plus inherent deformation of soft tissue during image acquisition result in severe image registration problems. In this study, the robustness of the segmentation is measured according to how similarly the same object is segmented in the two images, not how well boundaries of segmented objects match true boundaries or boundaries drawn by experts.

3.4.1 Methods

Twenty MLO view images from the Mammographic Image Analysis Society (Mini-MIAS) database of mammographic images (Suckling et al. 1994) were selected for this study. Images were subsampled by $4 \times 4 \rightarrow 1$ resulting images of size 256×256 pixels and spatial resolution of 800 μm per pixel. For this experiment, the same root extraction method, provided in the framework proposed by Jolion & Montanvert (1992) (Eq. 3.4), was used. The merging process (Section 3.3) is not used for this experiment. The parameters min_size and min_contrast were set as 100 and 5 respectively. Parameter α was set so that $F_{io}(1) = 64$. For each image, three derived images were computed: a shifted version, a rotated version and a warped version.

The shifted image was obtained by removing 100 rows from the top of the image and 100 columns from the side of the image away from the chest wall, and then adding 100 rows of zeros to the bottom of the image and 100 zero columns to the chest wall side of the image.

The rotated image was obtained by resampling the original image at a grid oriented 10 degrees with respect to the rows and columns in the original image using nearest neighbor interpolation.

The warped image was obtained by automatically selecting 4 equally spaced control points along the pectoral muscle boundary, 6 equally spaced control points from the breast boundary, and 5 control points from the chest wall edge of the image. The 5 points from the chest wall edge were kept fixed in order to retain a straight edge in the warped image. The control points along the pectoral muscle were moved 10 pixel units into the breast in the direction normal to the pectoral muscle boundary. The first two and last two control points along the breast boundary were moved 20 pixel units away from the breast in the direction normal to the breast boundary and the middle two control points along the breast boundary were moved 25 pixels outward (Fig. 3.6). These distances were adjusted in images with large breast if some of the points would otherwise fall outside the image. Thin plate splines (Bookstein 1989) based on these 15 pairs of control points was used to form the warped image.

For a comparison, a minimum spanning tree (MST) based segmentation algorithm was used (Ma et al. 2007). All 80 images (20 original + 20 shifted + 20 rotated + 20 warped) were segmented twice, once using the MST method and once using the AP. The segmented original version of each image was then shifted, rotated and warped as before so that for each deformation type (shift, rotate, warp), each algorithm (MST and AP) and each image, a pair of segmentations was available for which the true alignment of components was known.

For each true component pair, C_0 and C_d, a match score was assigned as $m(C_0, C_d) = |C_0 \cap C_d|/|C_0 \cup C_d|$. A pair of true components was viewed as matched correctly if $m(C_0, C_d) > 0.25$ (Fig. 3.7).
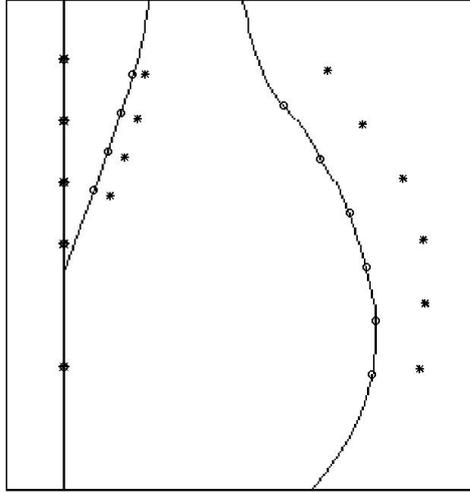
Figure 3.6: Control points for image warping. The open circles represent the initial locations of the control points and stars (∗) represent the final positions. The control points along the left edge (chest wall edge) are stationary and so the initial and final positions for these points coincide.

In addition, a salience score was assigned to each component in each image as $S(C) = |E(C) - E(B_C)|/E(C)$, where $E(C)$ is the mean intensity of set C and B_C is the "background" set of C obtained by dilating the component C by D_5, the disk of radius 5, and then removing C. Hence $B_C = (C \oplus D_5) \setminus C$.

The SV was computed for each original and distorted image pair as follows. For each component in the segmentation of the original image, the pixel values were set equal to the mean of the original image intensity values within that component. Similarly, for each component in the distorted image, the pixel values were set equal the mean of the image intensities of that component in the distorted image. The original image with these adjusted pixel values was then distorted (shifted, rotated, or warped) in the same way as was used to obtain its distorted pair. Thus one image was obtained by segmentation followed by distortion and the other was obtained by distortion followed by segmentation. In each image, pixel values within segmented components were set to the mean intensity of the component. The SV for the pair was summarized in two ways; the root mean square difference (RMSD) and the average difference (Av. dif).

$$\text{RMSD} = \sqrt{\frac{\sum d_{ij}^2}{\text{size of the image}}} \, ,$$

$$\text{Av. diff} = \frac{\sum |d_{ij}|}{\text{size of the image}},$$
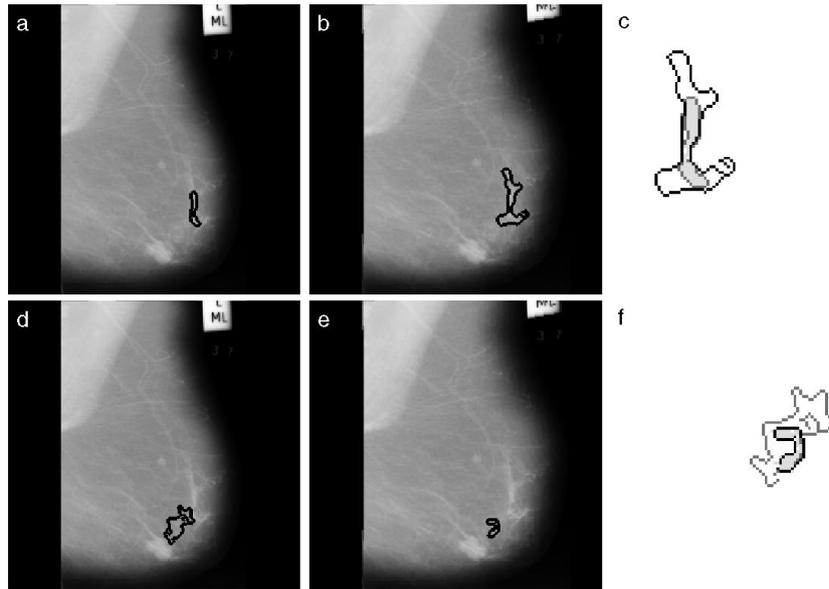
Figure 3.7: Matched and unmatched component pairs. (a) A component of image mdb005 (b) A corresponding component in warped image mdb005 (c) Original image mdb005 was warped again and was compared with warped image. Component of original mdb005 is in light boundary and component of warped mdb005 is in dark boundary. The shaded region is the overlapped part of these two components. This component pair has a match score of 0.32 and thus was identified as a true pair. (d)(e)(f) are the same as (a)(b)(c) but for a different component. In this case, the match score was 0.23 and hence this example was identified as unmatched.

where d_{ij} is the difference of the value of the (i, j)th pixel between the segmented image of the original image and the segmented image of the distorted image. Thus the SV for an image with respect to a distortion is zero if and only if the segmentation of the original image and the segmentation of the distorted image are identical both in terms of the location of the boundaries of all the components and in terms of the mean intensity of each component.

3.4.2 Results

In general, the number of components in the segmentation of the original is not the same as the number of components in the segmentation of the distorted image. The maximum number of correct matches is therefore equal to the minimum number of components found in two images. Hence the matching rate (Table 3.1) for each image is the number of correct matches divided by the minimum of the number of components found in the original and the number of components found in the distorted image.

	mean		std		
	AP	MST	AP	MST	p-value
shift	0.668	0.996	0.146	0.005	5.52×10^{-9}
rotate	0.523	0.659	0.049	0.039	8.54×10^{-10}
warp	0.486	0.820	0.048	0.043	3.53×10^{-17}

Table 3.1: Mean and STD of the proportion of possible components matched for AP and MST. For each type of distortion (shift, rotation, warp) the table shows the mean and the standard deviation (std) of the proportion of possible components matched over all 20 images for both the AP and MST algorithms. The last column shows the p-value for the hypothesis that the difference of the means is zero (pairwise test).

To test the relationship between the salience of a component pair and the matching score, the salience of a true matching pair C_0, C_d was taken to be the mean of their salience scores $S(C_0, C_d) = (S(C_0) + S(C_d))/2$. The proportion of true pairs with $m(C_0, C_d) < 0.25$ was recorded for pairs with $S(C_0, C_d) \leq 1$ and for pairs with $S(C_0, C_d) > 1$ (Table 3.2).

	shift		rotate		warp	
	$S \leq 1$	$S > 1$	$S \leq 1$	$S > 1$	$S \leq 1$	$S > 1$
AP	0.098	0	0.200	0.022	0.240	0.074
MST	0.0003	0	0.096	0	0.024	0

Table 3.2: The proportion of pairs in each group with match scores less than 0.25.

3.4.3 Discussion and Conclusion

The proportions of correctly matched pairs ($m(C_0, C_d) > 0.25$) are difficult to interpret on their own (Table 3.1). First, the fact that the proportion of successfully matched pairs is greater for warped images than for rotated images for MST is largely a consequence of the level of each distortion. The rotation by 10 degrees was chosen because, in practice, the variation of rotation due to positioning of the breast during acquisition is within this amount. The level of warping was chosen to loosely imitate distortions seen in practice but represents a qualitative judgement that may not be realistic. Second, the proportions include all components that together tessellate the image. By necessity, components are included that comprise components within roughly uniform background tissue and provide little information content regarding temporal image analysis (Fig. 3.8 - 3.10).

		shift	rotate	warp
AP	RMSD	4.45 ± 1.22	4.69 ± 0.71	6.43 ± 0.71
	Av. diff	1.95 ± 0.73	1.61 ± 0.27	3.44 ± 0.44
MST	RMSD	1.47 ± 0.87	5.58 ± 0.82	6.39 ± 1.33
	Av. diff	0.13 ± 0.10	1.72 ± 0.31	2.48 ± 0.62

Table 3.3: Mean and STD of RMSD and Av. diff: measurements of SV. Mean and STD of RMSD and Av. diff for original and distorted image pairs for 20 images and three distortions (shift, rotate, warp).

Useful segmentation in the context of temporal analysis of mammograms only requires robustness with respect to shifts, rotation, and warping for salient features. Accordingly, greater interest lies in the proportion of true pairs with high salience scores but low match values (Table 3.2). From this perspective both algorithms are quite successful. For true pairs with mean salience measure $S(C_0, C_d) > 1$, there were no pairs with $m(C_0, C_d) > 0.25$ for MST and very few for AP.

The proportion of total matches was higher for the MST algorithm than the AP algorithm (Table 3.1). The MST algorithm also performed better in terms of proportion of high salience features with low match scores (Table 3.2). The average RMSD measure of SV was significantly greater for AP than MST for shifts but was similar for AP and MST for rotations and warping. The Av. diff measure of SV was significantly greater for AP than MST for shifts and warping but was marginally less for MST than AP for rotation (Table 3.3).

No shift-variant has been reported with MST based segmentation algorithms. By comparing a SV immunized segmentation algorithm of MST, this experiment shows that irregular pyramids based AP is also slightly affected by SV, which has been shown in Marfil et al.'s (2006) work. However, the experiment also shows that SV affects of AP on salient features are little (Fig. 3.8 - 3.10). Performance of AP on salient components is quite successful. Although MST performs better than AP in terms of image distortion, in a previous study on segmenting the pectoral muscle in screening mammograms, the AP algorithm outperformed the MST algorithm (Ma et al. 2007). In this case, the criterion for good segmentation was based on expert drawn boundaries and so concentrated more on the detailed shape of the segmented component instead of the robustness to distortion.

Graph theory methods for segmentation have the advantage of providing both a segmentation of the image and a graph that encodes relationships between components. Accordingly, the output comprises more information than just the segmentation and hence the graph can be used, in principle, to improve subsequent tasks such as image registration, detection of anomalies, classification of tissue types, and detection of changes in the breast over time. In each of these
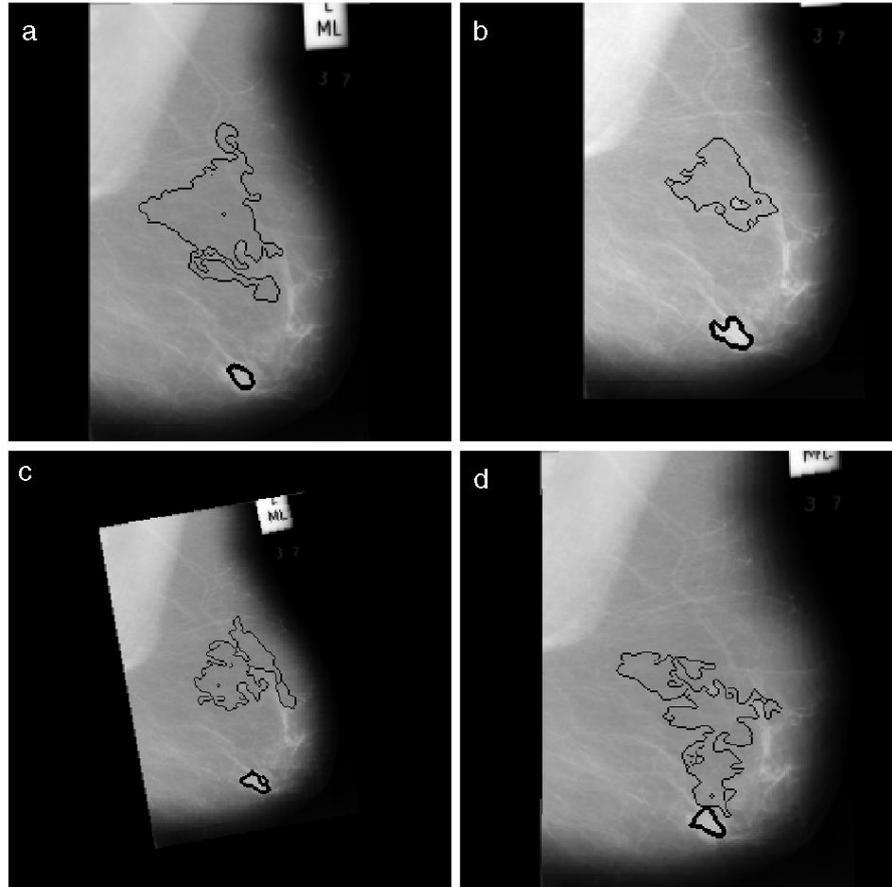
Figure 3.8: Examples of robust and non-robust segmentation. (a) Image mdb005. The thick line indicates a salient component, C_1, of the mammogram and the thin line indicates a poorly defined component, C_2, found by the AP algorithm. The salience values are $S(C_1) = 0.131$ and $S(C_2) = 0.0268$. (b) The shifted image with the same components. The match scores are $m(C_{1,O}, C_{1,d}) = 0.414$ for the salient component and $m(C_{2,O}, C_{2,d}) = 0.410$ for the poorly defined component. (c) The rotated image. Here the match scores are $m(C_{1,O}, C_{1,d}) = 0.517$ and $m(C_{2,O}, C_{2,d}) = 0.373$. (d) The warped image with match scores $m(C_{1,O}, C_{1,d}) = 0.588$ and $m(C_{2,O}, C_{2,d}) = 0.299$. Although both components are identified correctly under all three distortions, C_1 maintains high match scores while the match scores for C_2 deteriorate under rotation and warping.

cases, robustness of segmentation is only required for salient features and not all components. This observation allows the use of pyramid based methods such as the AP algorithm, though regular pyramid based algorithms have previously been shown not to be robust in general. Although runtime is generally not an issue in mammogram application, the average runtime of AP on a Pentium 4,

Figure 3.9: **Examples of segmentation of salient components. These five components of mdb019 segmented by AP are from relatively salient area, they are consistently segmented out by 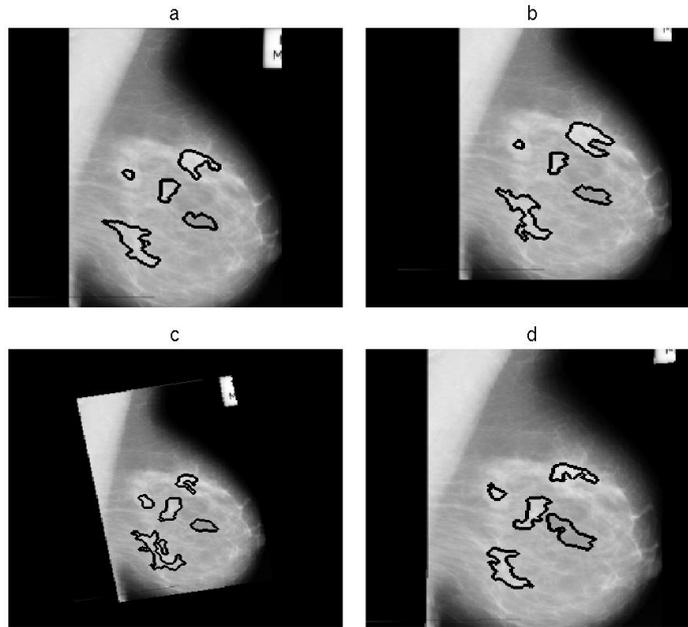AP in original mammogram and its shifted, rotated and warped version. (a)(b)(c)(d) are the original, the shifted, rotated and warped mdb019.**

2.80 GHz computer with 1GB memory for segmenting subsampled 256×256 size Mini-MIAS database mammogram was approximately 3 seconds.

To achieve reliable correspondence between the components of temporal mammograms and effective analysis for the detection of malignant masses, a mammogram segmentation method needs to be both robust and effective. Effectiveness of the segmentation means that the boundaries of the isolated objects are close to their true boundaries. An accurately segmented boundary of an object allows better extraction of object features and thus facilitates recognition of manifestations of malignancy in masses. For an accurate evaluation of the effectiveness of the mammogram segmentation, true boundaries of the masses need to be identified and compared to the segmented boundaries. However, identification of the true boundaries of the masses is difficult, even for an experienced radiologist. In fact, a member of our research team, who is an experienced radiologist, was asked to mark the boundaries of the masses. However, he can only give a box that covers the whole mass, as he believes that in many cases, masses do not have perceptible boundaries. In this thesis, the number of the successfully isolated masses in the experimental dataset was used as a rough evaluation of the effectiveness of the segmentation. A mass was identified as unsuccessfully segmented if the corresponding segmented component was too big (under-segmented, containing too much non-mass area) or too small (over-segmented, the component was only a
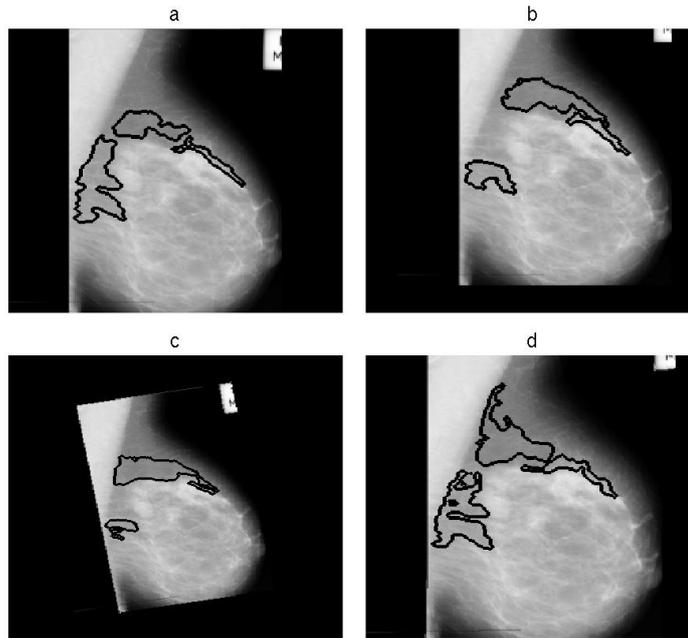
Figure 3.10: Examples of segmentation of non-salient components. Three non-salient components of mdb019 segmented out by AP are shown in (a)(b)(c)(d), the original, shifted, rotated and warped mdb019. Segmentation on these three components differ significantly.

small part of the true mass). The AP algorithm successfully segmented all of the masses in the selected Mini-MIAS mammograms (see Section 3.4.1, 8.1.1) and in the training dataset (see Section 8.2.1), and only missed two in the testing dataset (see Section 8.2.1). In terms of the number of successfully segmented masses in the data sets used, the AP algorithm outperformed the MST algorithm. In this thesis, the AP algorithm was used to segment the mammograms.

Chapter 4

Automatic Pectoral Muscle Extraction on MLO View Mammograms

In the previous chapter, the robustness of the AP segmentation was evaluated in terms of image shifts, rotations, and warping. The effectiveness of the AP segmentation was roughly estimated based on the number of successfully isolated masses. However, as stated before, there is no easy way to measure the effectiveness of the segmentation for masses because of the difficulty of defining the true boundary. In order to quantify the accuracy of the segmentation method and to compare it with other segmentation routines one has to choose a segmentation task with the following properties. (i) The task should be the detection of some class of objects in mammograms (as opposed to the other class of images) so that the image properties are similar to the task of segmenting masses. (ii) Expert drawn boundaries should be available or possible to obtain and should be somewhat reliable (iii) Data sets should be used that have been used in other studies so that direct comparisons can be made. The task of pectoral muscles segmentation satisfies all these criteria.

The pectoral muscle is one of the few anatomical features that appears clearly and reliably in most MLO view mammograms (see e.g. Fig. 4.4, 4.7 and 4.8). The pectoral muscle is an important landmark both for providing contextual information regarding anatomies and for image registration. Accurate segmentation of the pectoral muscle is among the many tasks that is needed to improve CAD for mammography. Exclusion of the pectoral muscle saves processing time and boosts performance of many applications.

To a first approximation, the pectoral muscle appears as a bright triangular patch in the upper left or upper right corner (depending on right or left breast) of the image. This motivated initial algorithms based on the Hough transform (Karssemeijer 1998, Ferrari, Rangayyan, Desautels, Borges & Frère 2004a). The pectoral muscle is usually not exactly triangular and more accurate segmentation

was achieved by using Gabor wavelets to segment the pectoral muscle without assuming straight boundaries (Ferrari et al. 2004a). Aside from incorporating the general shape and location assumptions of the pectoral muscle, these methods rely only on local image information.

This chapter describes a method to extract the pectoral muscle boundary automatically. In this method, the AP segmentation algorithm is used to segment the mammogram into many components. The same settings of the AP algorithm, as introduced and used in the previous chapter (see Section 3.2), is used in this method. After the segmentation, components belonging to the pectoral muscle were identified based in three steps. Section 4.1 presents these three steps. An initial pectoral muscle boundary is extracted from these components. An adaptive deformable contour model is then used to refine the initial boundary. The adaptive deformable contour model is described in Section 4.2. An experiment is designed to measure the performance of the method. Section 4.3 introduces the database used for the experiment. Results from the experiment are presented in Section 4.4. For comparison, the results of this method are compared to the results obtained by the Hough transform method, the Gabor wavelets method and an MST based method. Finally, discussion and conclusion are presented in Section 4.5.

4.1 Identification of Pectoral Muscle Components

Typically, the AP algorithm segments the mammogram into one to three big components belonging to the non-breast area and dozens of small and medium components belonging to the breast area (Fig. 3.1). The pectoral muscle is generally separated into a number of components. Accordingly, it is necessary to develop rules for assigning the components as forming part of the pectoral muscle or not. This is accomplished by using prior geometric information in three steps.

Step 1: Since the pectoral muscle always occupies the upper left corner of the image, the pixel at position $(1, 1)$ will be part of pectoral muscle. (In digitized mammograms, the edges of the mammogram may not be flush with sides of the digitized image. Also, in some data sets, like e.g. the Mini-MIAS database (see Section 4.3), the digitized mammograms are embedded into larger images in order to maintain fixed image sizes. Accordingly, it is necessary to distinguish between the first and the last column of the overall image and the first and the last column of the mammogram. The reference to the position $(1, 1)$ above is with respect to the mammogram and not the image.) The root of this pixel is taken to represent the seed component for the pectoral muscle. In cases where this first candidate component, C_1, is too small (less than 100 pixels in our experiments), the root of the pixel at location $(I + 1, J + 1)$ is used as the seed component, where $J = \max\{j : (i, j) \in C_1\}$ and $I = \max\{i : (i, j) \in C_1\}$.

Step 2: Other components are considered as candidates for forming part of the pectoral muscle if they meet the following intensity, size and location restrictions. Numbers used in these restrictions were selected based on the mammograms from the Mini-MIAS database (Suckling et al. 1994).

1. Intensity restriction. The mean intensity of the component is within 80 units of the seed component.

2. Size restriction. The size is in the range [15,4500].

3. Location restriction. Here two location restrictions are used. First, the root of the component is inside the "triangle of interest" (The triangle of interest is the upper left triangle formed by cutting the smallest box containing the breast component by the diagonal joining the upper right corner to the lower left corner). Second, the distance between the root of the component and the root of the seed component is less than the distance between the root and the diagonal boundary of the triangle of interest.

The above restrictions used the mean intensity, size and location of the root of each component, which were naturally produced by the segmentation and associated with each root, and thus requires little computation.

Step 3: The previous step disqualifies most of the candidate components. In this step, the candidate components are further investigated and a candidate components is disqualified if the location, shape, orientation, or right boundary is unrealistic. The measurements in this step involves all vertices of each component. The location is considered unrealistic if the geometric center of the component is located more than 30 pixels from the edge of the seed component. The shape is considered unrealistic if the area of the component divided by the area of its smallest bounding box is less than 5. To judge the orientation and right boundary, best straight line fits (in the least square sense) are found for both the left and right boundaries of the candidate component. For each row, the error between the best straight line fit to the right boundary and the right boundary is recorded. The average right boundary error (ARBE) is found by summing these errors and dividing by the number of rows containing pixels from the candidate component. If the slope of either the left or right best fitting line has sign opposite to the sign of the slope of the hypotenuse of the triangle of interest, the orientation is considered unrealistic. The right boundary is considered unrealistic if ARBE is greater than four. Again, the numbers used in this step are developed for images from the Mini-MIAS database.

4.2   Adaptive Deformable Contour Model

The AP provides estimates of the pectoral muscle that corresponds well to the true pectoral muscle in terms of location and general shape. However, boundaries

are ragged and generally appear somewhat to the left of the visually apparent boundary (Figs. 4.4, 4.7, 4.8, 4.9 and 4.10). To improve the detail of the segmentation, the boundaries found by AP are used to initialize a local segmentation method based on active contours. Active contours were introduced by Kass, Witkin & Terzopoulos (1987) and have evolved in a number of directions. The version used here is based on work by Lobregt & Viergever (1995), Williams & Shah (1992), and Ferrari, Rangayyan, Desautels, Borges & Frère (2004b).

The algorithm for implementing active contours will be described as it applies to the left breast. Let $V = \{v_1, v_2, ..., v_N\}$ denotes the set of pixels forming the current pectoral muscle boundary. Let (x_i, y_i) denote the coordinates of $v_i, i = 1, 2, ..., N$. The active contour model works by moving the boundary through the spatial domain of the image to minimize a measure of energy based on the following formulas.

$$E_i = \alpha E_{in,i} + \beta E_{ex,i},$$

where α, β are two weights controlling the internal and external energies $E_{in,i}$ and $E_{ex,i}$. The internal and external energies are given by

$$E_{in,i} = a_1 V'(v_i) + a_2 V''(v_i)$$
$$E_{ex,i} = -\left|I_x(v_i)\right| / \max_I(I_x),$$

where $V'(v_i)$ and $V''(v_i)$ are the first and second derivatives of the contour V at v_i, I is the image, and

$$I_x = \frac{\partial I}{\partial x}.$$

The weights a_1 and a_2 are used to control the relative contributions of $V'(v_i)$ and $V''(v_i)$ and are fixed for this study at $a_1 = 1$ and $a_2 = 2$.

The internal energy serves to reduce the curvature of the contour. This is important since the pectoral muscle has a general smooth straight shape. The external energy drives the contour toward strong edges in the image. This is also important since the pectoral muscle generally appears much brighter in the image than surrounding tissue.

The active contour is an open curve that only moves horizontally in the image. This simplification takes advantage of prior information regarding the general appearance and location of the pectoral muscle in the image. The advantages are a reduction of complexity and the convenience of using row numbers as a fixed index for points at which the active contour is evaluated.

At every point v_i, the energies are computed on an asymmetric neighborhood (Fig. 4.1)

$$\Omega_i = [(x_i - 5, y_i), \ldots, (x_i + 3, y_i)] \quad \text{(left breast)}. \tag{4.2}$$

Figure 4.1: Example of asymmetric neighborhood used in adaptive deformable contour model. This example shows the domain Ω of v_i. In this case, the chest wall is left hand side positioned. The v_i is modified to e_j if the $\min E_i$ is reached in e_j.

Asymmetric neighborhoods are used since the initial pectoral muscle boundary usually appears closer to the chest wall than the true boundary.

Unlike in other implementations of deformable contour models, the weights for internal and external energy, α and β, are adjusted automatically as follows.

$$\alpha = |x_i - x_{i-1}| + |x_{i+1} - x_i| - 2*d, \quad \text{where} \quad d = (x_1 - x_N)/N$$
$$\beta = \exp((\max_{\Omega_i} |I_x| - \min_{\Omega_i} |I_x|)/\operatorname*{mean}_{\Omega_i} |I_x|)$$

The parameter α measures the local deviation of the slope from the global average slope. When this value is small, the external forces are allowed to dominate in order to push the active contour toward boundary. When α is large, the internal forces dominate to straighten the active contour. Similarly, β forces the active contour to converge to the boundary quickly when the active contour is far away, but favors local straightening of the active contour over the rate at which it approaches the boundary when it is close. As a result the active contour attains an anatomically realistic local shape as it moves toward the boundary.

The elements of Ω_i will be denoted by $e_j, j = 1, 2, .., 9$, and the internal and external energies at these points will be denoted by $E_{in,i}^j$ and $E_{ex,i}^j$ respectively. Thus

$$E_{in,i}^j = a_1 V'(e_j) + a_2 V''(e_j)$$
$$E_{ex,i}^j = - |I_x(e_j)| / \max_I (I_x),$$

where $V'(e_j)$ and $V''(e_j)$ are the derivatives along the curve obtained by replacing v_i by e_j.

To allow comparison between the different energy terms, it is necessary to rescale them to the range $[0, 1]$.

$$\hat{E}_{in,i}^j = \frac{E_{in,i}^j - E_{in,i}^{\min}}{E_{in,i}^{\max} - E_{in,i}^{\min}},$$
$$\hat{E}_{ex,i}^j = \frac{E_{ex,i}^j - E_{ex,i}^{\min}}{E_{ex,i}^{\max} - E_{ex,i}^{\min}},$$

where the superscripts min and max denote the minimum and maximum of the respective quantities over the domain Ω_i. Thus the contour is driven to minimise

$$\hat{E}_i = \alpha \hat{E}_{in,i} + \beta \hat{E}_{ex,i}.$$

The energy of the contour is minimised iteratively. Each iteration consists of minimising $\hat{E}_i$ for $i = 1, \ldots, N$ consecutively. At a given step, the point v_i will be replaced by the point e_j, if

$$\hat{E}_i^j = \min \hat{E}_i^k, k = 1, 2, ..., 9.$$

In this study the number of iterations is fixed at 30 though experiments show that a stable contour is generally reached in fewer than 8 iterations.

4.3 Database

To evaluate and compare the performance of the proposed method with state-of-the-art methods reported in the literature, 84 images were obtained from the Mini-MIAS database of mammographic images (Suckling et al. 1994). The same images were selected as those used in the study on identifying the pectoral muscle conducted by Ferrari et al. (2004a). The spatial resolution of these images is $200 \ \mu m$ and depth resolution is 8 bit. The images in the database are 1024×1024 pixels in size. For this study, the images were further subsampled to 256×256 pixels. In the work by Ferrari et al. (2004a), boundaries found by the algorithm were compared to boundaries drawn by a radiologist. The coordinates of the lines drawn by the radiologist in that study were kindly provided by Ferrari and Rangayyan so that our results could be compared to the same standard.

4.4 Results

Area between the algorithm identified pectoral muscle boundary and the boundary drawn by the radiologists is used as segmentation error to measure the performance of the method. To compensate for the variation of the size of the pectoral muscle in different mammograms, the error area is normalized in two different ways; both by the area of the radiologist identified pectoral muscle and by the number of rows in the radiologist identified pectoral muscle. According to these two different normalization methods, the measurements are called the area normalized error and the row normalized error, respectively.

Results are compared with Hough transform method, Gabor wavelet method and the MST based method. Pectoral muscle extraction using Hough transform method and Gabor wavelet method were reported by Ferrari et al. (2004a). The MST based method uses a similar framework as used by the method described

in this chapter but based on the MST segmentation results. An initial pectoral muscle boundary is also extracted from the MST segmentation results and the same active contour model, as described in Section 4.2, is used to refine the initial boundary. In the following paragraph, the MST based method will be referred to as MST and the method presented in this chapter will be referred to as AP.

Since the manually identified boundaries are obtained from the original full-size images (1024×1024 pixels), while the results of this study are extracted from the down-sampled images of size 256×256, the detection results are transferred back to the original size by interpolation.

4.4.1 Area Normalized Error

The area normalized error is the same as the one used in the work (Ferrari et al. 2004a) to measure the performance of Hough transform method and Gabor wavelet method. By using the area normalized error, results obtained in this thesis can be compared to the results achieved by using Hough transform method and Gabor wavelet method (Ferrari et al. 2004a).
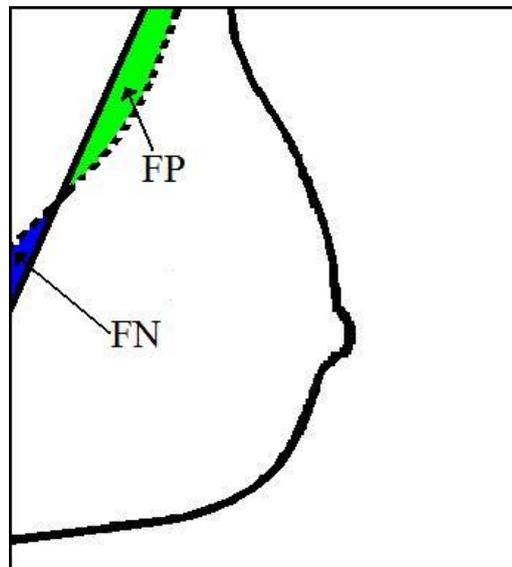
Figure 4.2: False positive and false negative. The solid straight line represents the radiologist drawn pectoral muscle boundary and the dashed line is the algorithm detected boundary. The green region is outside the radiologist identified pectoral muscle area but inside the algorithm detected pectoral muscle area and is FP area. The blue area is FN area.

In the area normalized error, the error area is separated into false positive (FP) area and false negative (FN) area (Fig. 4.2). The FP area is the area

inside the algorithm detected pectoral muscle but outside the radiologist identified pectoral muscle. On the other side, the FN area is the area included by the radiologist drawn pectoral muscle boundary but excluded by the algorithm detected boundary. The use of FP and FN in this context is slightly contrary to the standard use of these terms (In ROC analysis, for example) but is used here to reflect several papers in the literature that address this specific problem. With the definition of FP and FN area, for a left breast, the FP rate (FP_I) and FN rate (FN_I) for an image I were calculated as

$$FP_I = \frac{1}{A(I)} \sum_{i=1}^{p} \max\{0, B_{alg}(i) - B_{rad}(i)\},$$

$$FN_I = \frac{1}{A(I)} \sum_{i=1}^{p} \max\{0, B_{rad}(i) - B_{alg}(i)\},$$

where $B_{rad}(i)$ and $B_{alg}(i)$ are horizontal coordinates of the ith points in the boundaries identified by the radiologist and determined by the algorithm, respectively. $A(I)$ is the area of the pectoral muscle identified by the radiologist and p is the maximum of the number of rows in radiologist identified pectoral muscle and the number of rows in algorithm detected pectoral muscle. $B_{alg}(i)$ and $B_{rad}(i)$ are set to 0 if i exceeds the total number of points in algorithm detected boundary and radiologist identified boundary. The average rate of FP and FN for the collection of images $I_i, i = 1, ..., N$ in the data set are computed by

$$FP_{ave} = \frac{1}{N} \sum_{i=1}^{N} FP_i,$$

$$FN_{ave} = \frac{1}{N} \sum_{i=1}^{N} FN_i.$$

The MST algorithm did not identify any components belonging to the pectoral muscle for images mdb098 and mdb109 (Fig. 4.3). Accordingly, the FP_{ave} and FN_{ave} rates for the MST method are based on 82 images.

Table 4.1 presents FP/FN rates for Hough transform method, Gabor wavelet method, MST and AP. From the table, MST shows comparable results to the Hough transform method in terms of FP_{ave}, but AP did not performs as well as either the Hough transform or Gabor wavelets. In terms of FN_{ave}, AP outperformed the Hough transform and is comparable to the Gabor wavelets.

The FP_{ave} and FN_{ave} are greatly influenced by few difficult cases. For example, the radiologist drawn pectoral muscle boundary of image mdb055 seems incomplete (Fig. 4.4). Although AP and MST both found visually acceptable boundaries for this image, FP_I is very huge for both MST and AP on this case. In the case of AP, when mdb005 is excluded from the analysis, the value of FP_{ave} drops from 0.0371 (Table 4.1) to 0.0270.
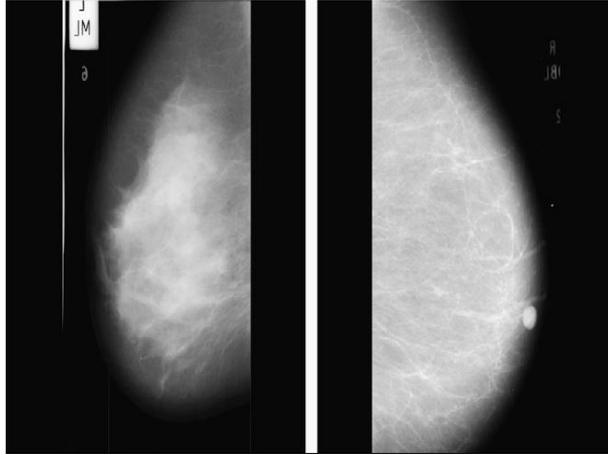
Figure 4.3: Two images with pectoral muscle absence. Unlike most MLO view mammograms the pectoral muscle is essentially absent in these two images (mdb109 on the left and mdb098 on the right).

	Hough	Gabor	MST	AP
FP_{ave}	0.0198	0.0058	0.0255	0.0371
FN_{ave}	0.2519	0.0577	0.1168	0.0595
$FP_I < 0.05, FN_I < 0.05$	10	45	40	50
$\min(FP_I, FN_I) < 0.05$	0	0	20	18
$0.05 < \max(FP_I, FN_I) < 0.10$				
$\min(FP_I, FN_I) < 0.05$	0	0	18	11
$\max(FP_I, FN_I) > 0.10$				
$0.05 < FP_I < 0.10$	8	22	0	0
$0.05 < FN_I < 0.10$				
$0.05 < \min(FP_I, FN_I) < 0.10$	0	0	1	0
$\max(FP_I, FN_I) > 0.10$				
$FP_I > 0.10, FN_I > 0.10$	66	17	3	5

Table 4.1: **FP and FN proportion and distribution according to the area method:** pectoral muscle boundary extraction results of 4 methods. The first two rows report the proportion of FP and the proportion of FN pixels averaged over all the images. The next six rows indicate the distributions of FP and FN proportions. For example, the third row lists the number of images for which the proportion of FP pixels and the proportion of FN pixels were both less than 0.05. Values for the Hough transform and Gabor wavelet method were reported previously by Ferrari et al. (2004a).
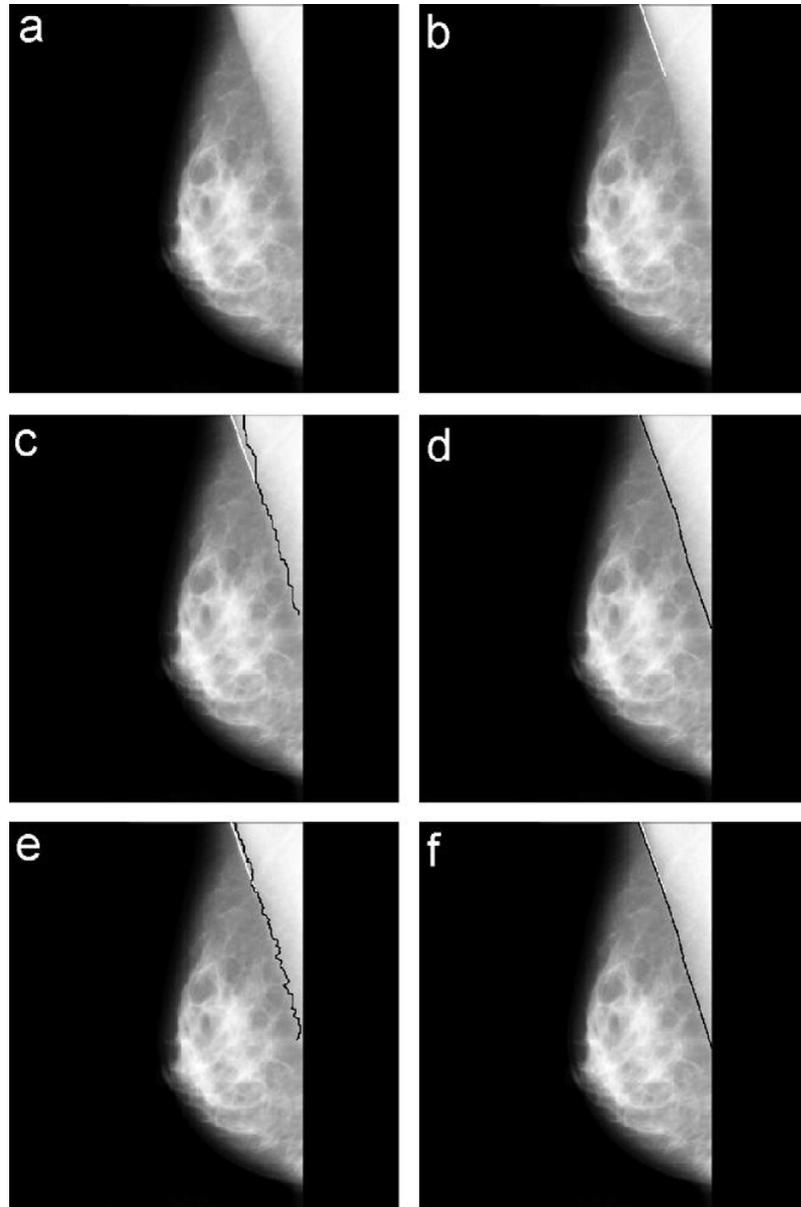
Figure 4.4: False source of error. (a) Image mdb055, (b) the radiologist drawn boundary (white), (c) the boundary found by MST before applying active contours (black), (d) the boundary after active contours (black), (e) and (f) the same as (c) and (d) but for the AP method. The radiologist's boundary terminates prematurely but both MST and AP find reasonable boundaries. Although the raw boundaries found by MST and AP are nominally correct, the raggedness of the lines is not anatomically reasonable.

From this view point, the number of images for which both FP_I and FN_I are small (row 3 of Table 4.1) provides a more robust measure of segmentation performance. In this case the AP shows slightly better performance than Gabor wavelets and MST and all three of these are vastly superior to the Hough transform. Both the MST and AP perform well according to the number of images for which at least one of FP_I or FN_I is less than 0.05 and both are less than 0.01 (rows 3 and 4 of Table 4.1). The number of images for which both FP_I and FN_I are large (Table 4.1, last row) is much smaller for AP and MST than for Gabor wavelets and Hough transforms.
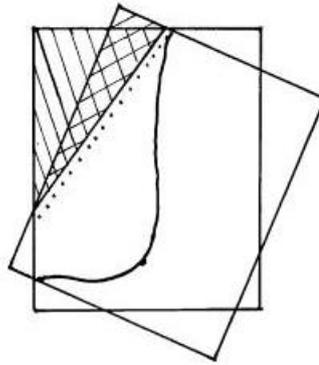
Figure 4.5: Area normalized error affected by different breast positionings. The sketch shows two different images (two squares) of a same breast which is caused by different positionings of the breast. The dotted line is a hypothetical boundary found by the algorithm and the parallel solid line is the hypothetical true pectoral muscle boundary. The two shaded regions are the hypothetical pectoral muscles. In these two images, the area between the algorithm detected boundary and the true boundary is exactly the same. However, as the areas of the pectoral muscle in these two images are very different, both FP_I and FN_I rates computed by the area normalized error will be very different. The RNE value (Eq. 4.3) is the same for the two images.

4.4.2 Row Normalized Error

The area normalized error heavily depends on the amount of the pectoral muscle appearing in the mammogram, which is greatly affected by the positioning of the breast during the examination and so is not well controlled (Fig. 4.5). The row normalized error, on the other hand, normalizes the error area by the number of rows included in the radiologist identified pectoral muscle and thus is independent of the positioning of the breast at image acquisition. The row normalized error

(RNE) for image I is defined as

$$RNE(I) = \frac{1}{p} \sum_{i=1}^{p} |B_{rad}(i) - B_{alg}(i)| \qquad (4.3)$$

The total error RNE is a single quantity. Dividing the total error area into FP and FN area provides insight into the nature of the errors encountered and is clearly useful during the development of the algorithm. However, there is no clear medical reason for greater for FP or FN and so RNE is a suitable measure of error.

For MST, $RNE(I)$ is only computed on 82 images as it did not detect any boundary for images mdb098 and mdb109 (Fig. 4.3).

For both the AP and MST algorithms, more than 80% of the images had RNE scores less than 10 pixels (2.0 mm) per row (Fig. 4.6). In the case of the AP algorithm, only two images had mean error greater than 25 pixels (5.0 mm).
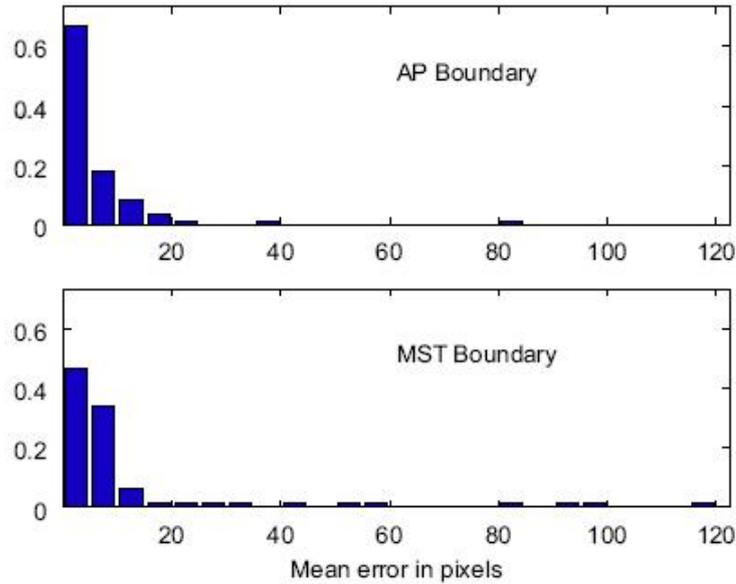
Figure 4.6: Histogram of RNE of AP and MST segmentation. Histograms showing the proportion of images with error values (as measured by RNE) in various ranges. The first bar represents images with average errors in the range (0, 5] pixels, the second bar represents the range (5, 10], and so on. Five pixels correspond to 1mm actual size. Hence, approximately 60% of the images have errors of less than 1.0mm per row and 97% have errors less than 5.0mm per row.

4.5 Discussion and Conclusion

The proposed AP method does not clearly perform better than the method based on Gabor wavelets studied by Ferrari et al. (2004*a*) in segmenting the pectoral muscle in mammograms. However, in the case of the number of images with small error, the proposed method surpasses other methods (Table 4.1). In a previous study conducted by our research team (Ma, Bajger, Slavotinek & Bottema 2006), 4 radiologists drawn pectoral muscle boundaries of the same set of mammograms as used in this study were compared with the boundaries obtained by both AP and MST methods, the study found that the performance of the AP method is not statistically different from radiologists.

Factors that cause difficulties in segmenting pectoral muscle include dense tissue appearing near the pectoral muscle, very small pectoral muscle (occupies an extremely small portion of the image) and strong lines parallel to the true boundary existing in the pectoral muscle region. In this study, the presence of dense components near pectoral muscle boundary does not cause serious errors to the method proposed in this chapter. This is due to the implementation of the AP. The steps described in Section 4.1 avoid dense components outside pectoral muscle to be picked up. A small pectoral muscle also does not cause trouble to the proposed method. The main source of significant error is the presence of strong lines within the pectoral muscle region parallel to the true boundary. Such lines appeared in approximately 10% of the images. Although the proposed method identified the correct boundary in some of these examples (Fig. 4.7, 4.8 and 4.9), it failed in others (Fig. 4.10). AP identified the wrong line in two such images and these were the only two images for which it produced a boundary with RNE greater than 25 pixels (5 mm). In real mammogram diagnosis, both of these two images are not of diagnostic standard. The strong lines within the pectoral muscle region are due to skin folds. If these appear, the mammogram is usually rejected and a new image is acquired.

Figure 4.7: Pectoral muscle boundary extraction for mdb033. (a) Original image (b) Hand-drawn pectoral muscle edge (c) and (d) Pectoral muscle edge detected by AP method and adaptive deformable contour model, respectively.
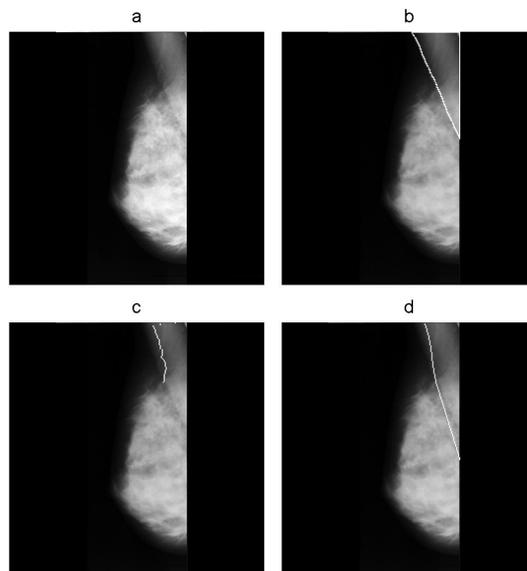

Figure 4.8: Pectoral muscle boundary extraction for mdb110. (a) Original image (b) Hand-drawn pectoral muscle edge (c) and (d) Pectoral muscle edge detected by AP method and adaptive deformable contour model, respectively.
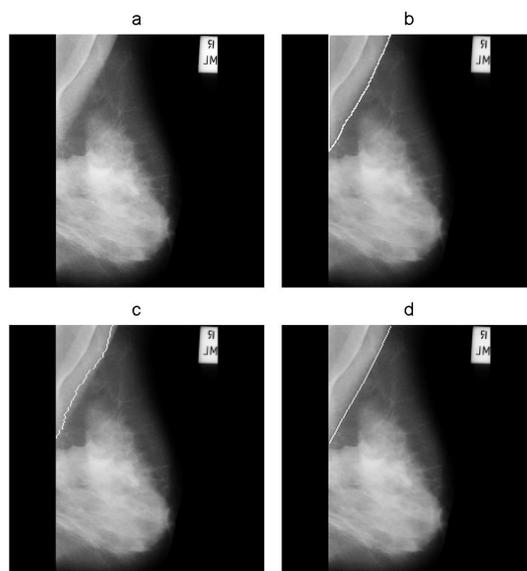
Figure 4.9: Example of pectoral muscle boundary extraction with bad initial boundary. (a) image mdb123, (b) the radiologist drawn boundary (white), (c) the boundary found by MST before applying active contours (black), (d) the boundary after active contours (black), (e) and (f) the same as (c) and (d) but for the AP method. The strong cleft parallel to the true boundary did not prevent the MST and AP algorithms from finding the correct components associated with the pectoral muscle. In this case, both algorithms could not find components that extended to the full length of the pectoral muscle, but the active contour step provided reasonable extension, smoothing, and fine tuning of location.
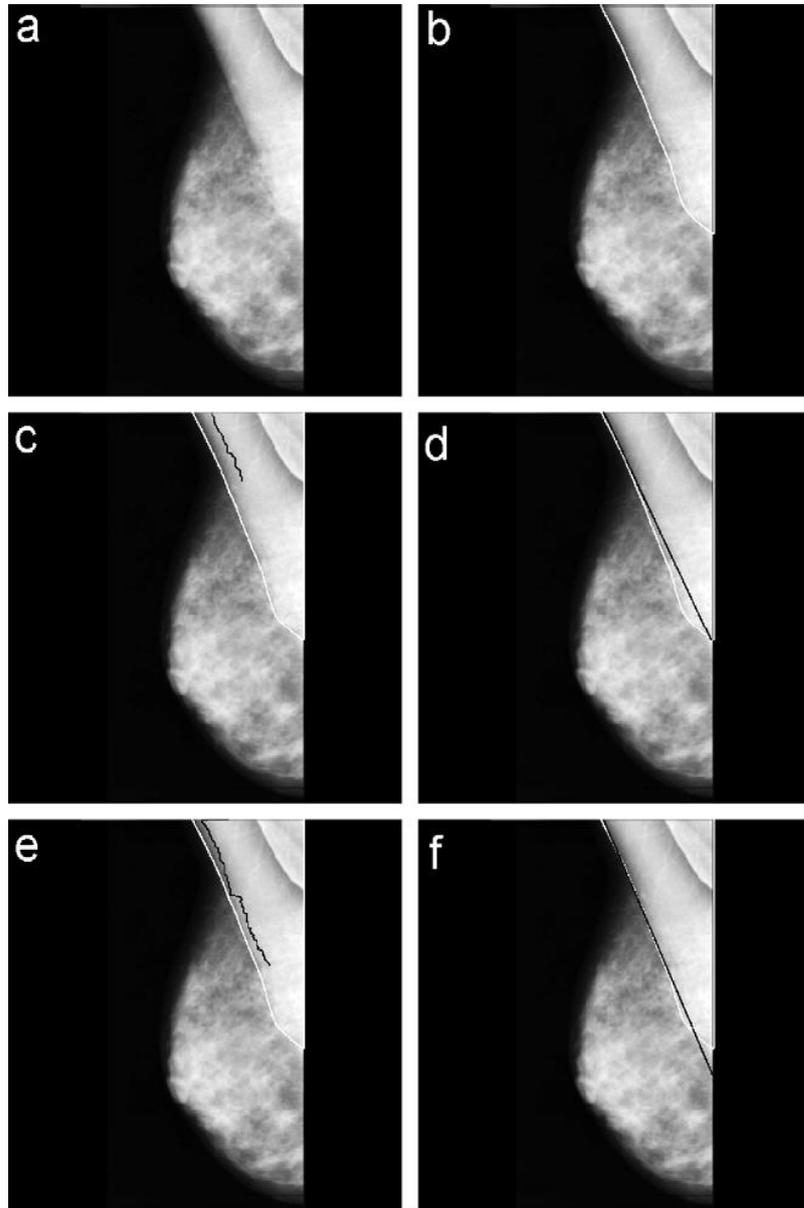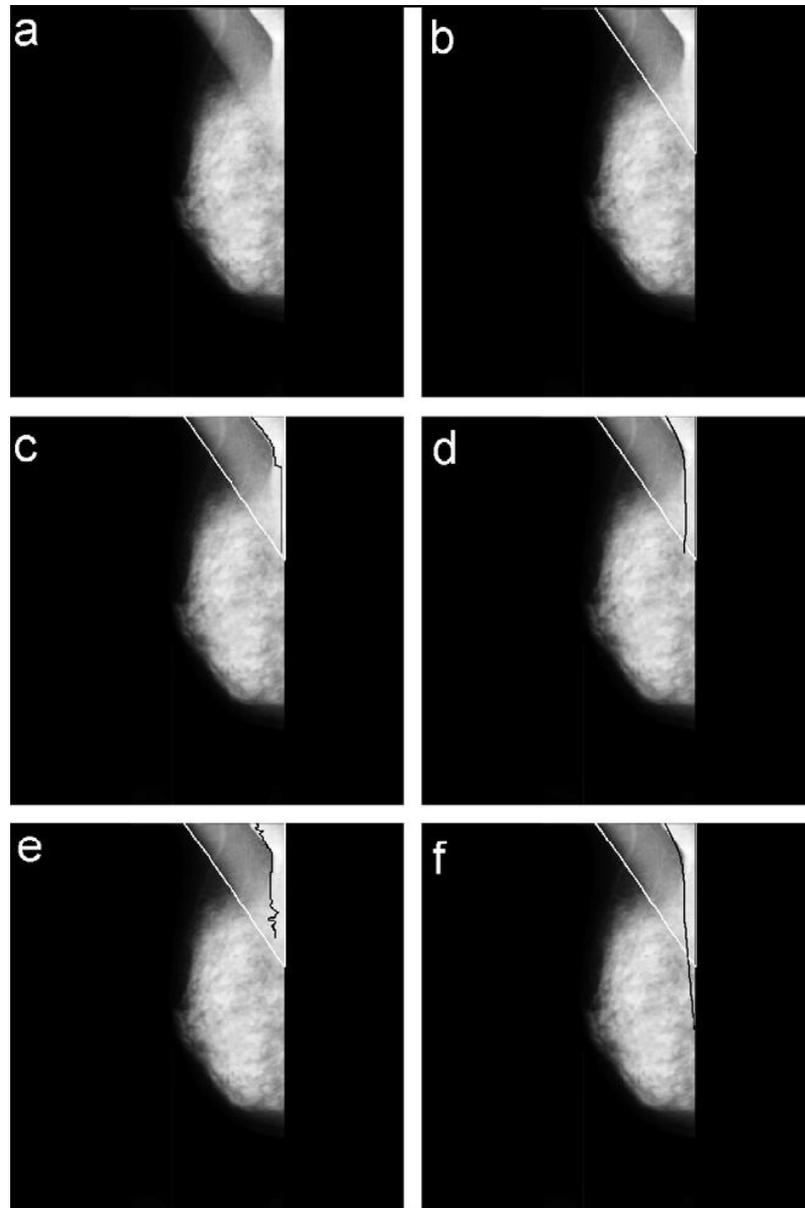
Figure 4.10: Example of poor pectoral muscle boundary extraction. (a) image mdb039, (b) the radiologist drawn boundary (white), (c) the boundary found by MST before applying active contours (black), (d) the boundary after active contours (black), (e) and (f) the same as (c) and (d) but for the AP method. In this example, the strong edge parallel to the true boundary fooled both the MST and AP algorithms. Similar failure occurred in image mdb068, but these were the only two images for which the boundary found by the AP algorithm was not essentially correct.

Chapter 5

Automatic Breast Boundary Segmentation

Accurate detection of the breast boundary is an important task in CAD systems. An accurate extracted breast boundary allows removal of the background of the image and thus avoids unnecessary processing time. Artifacts, such as patient and imaging information labels, are normally present in mammograms and may affect the performance of the CAD systems. By removing the background subsequent to breast boundary detection, these artifacts may be removed.

The breast boundary is also used in many mammogram registration methods (Section 2.1). In these methods, global registration of the multiple mammograms is realized by aligning the corresponding breast boundaries. For these methods, accurate extraction of breast boundaries is essential. In many other tasks of the mammography, the breast boundary is used as additional information. In this thesis, the breast boundary is used to remove the segmented components that belong to the background (Section 6.4). During graph matching (Section 7.2), global reference is provided by using the fuzzy spatial relations between the breast boundary and the mass-like objects.

As one of the few reliable anatomic landmarks on the breast, the localization of the nipple is of great importance. It serves as an alignment pivot in many mammogram analysis algorithms (Chandrasekhar & Attikiouzel 1997, Marias et al. 2005, Sanjay-Gopal et al. 1999). A breast boundary with the nipple well preserved is a prerequisite of nipple localization algorithms applied to the detected boundary.

Although detection of the breast boundary may seem quite straightforward, it has been shown not to be an easy task. Difficulties of breast boundary extraction include the presence of artifacts, which may present near or overlap the breast boundary, and noise. Many methods have been proposed to detect the breast boundary. Thresholding and histogram analysis have been used in many of these algorithms (Bick, Giger, Schmidt, Nishikawa, Wolverton & Doi 1995, Abdel-Mottaleb, Carman, Hill & Vafai 1996, Ojala, Näppi & Nevalainen 2001). Recent

advances in breast boundary identification include active contour models based methods (Ferrari et al. 2004*b*) and dependency approach (Sun, Suri, Desautels & Rangayyan 2006).

This chapter describes two methods for automatic extraction of breast boundary. An initial method, based on the AP segmentation results, is presented in Section 5.1. A more robust method, based on real rational orthogonal wavelets (ROW) and Markov Random Field (MRF) smoothing is described in Section 5.2. Robustness of this method comes from the ROW filtering. Another advantage of this method is its ability of preserving the nipple. Performance of these two methods is analyzed in Section 5.3. Discussion and conclusion are made in Section 5.4.

5.1 Breast Boundary Extraction Based on AP Segmentation

The AP algorithm segments the image into components. The background of the mammogram is usually represented by one or several components. A breast boundary detection method was developed based on the AP segmentation results. The first postprocessing step is to identify the components belonging to the background. Based on the experiments on the mammograms both from Mini-MIAS database and from a local archive, it was observed that the components belonging to the background usually have large size and low average intensity. Simple thresholding on the size and average intensity of the components is sufficient to identify the components belonging to the background.

After the identification of the background components, the whole breast region is isolated by simply removing the identified background components from the mammogram. An initial breast boundary is then extracted from the isolated breast region.

The initial boundary is generally close to the true boundary but somewhat inside the true boundary. Due to the noise that may be present near the boundary, the initial boundary may have discontinuities. A simple procedure is used to fix the discontinuities and to draw the initial boundary to the true boundary.

The procedure detects discontinuities in the initial breast boundary based on the first derivative of the y coordinates of the boundary with respect to the x coordinates. The first derivative of the boundary at (x_i, y_i) is calculated as

$$y_i' = \frac{y_{i+1} - y_i}{x_{i+1} - x_i}.$$

Here the coordinate system is defined with the upper left corner $(1, 1)$ as origin and the top to bottom direction as x-coordinate and left to right direction as y-coordinate (Fig. 5.1). Any point in the initial boundary with the first derivative

bigger than 4 is identified as a discontinuity. Discontinuities are corrected by replacing the points with the mean of its corresponding nearest points. Thus a discontinuity at (x_i, y_i) is replaced by the point nearest to $(\frac{x_{i-1}+x_{i+1}}{2}, \frac{y_{i-1}+y_{i+1}}{2})$. The scanning for the discontinuity ends when no discontinuity can be detected.
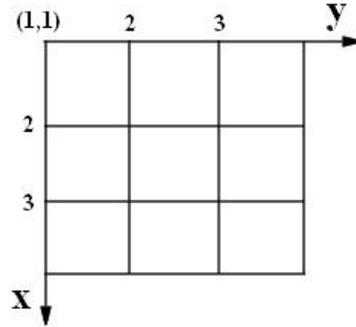

Figure 5.1: Coordinate system defined for the mammogram. The upper left corner $(1,1)$ is the origin and the x-coordinate from top to bottom and y-coordinate from left to right.

After discontinuity correction, a simple procedure is used to push the initial boundary out to the true boundary. The push-out procedure is based on the intensity and pushes the boundary out both horizontally and vertically. For the horizontal push-out, for a boundary point v_i at (x_i, y_i), a neighborhood point set of v_i, S_i, is firstly identified. S_i consists of the 30 pixels to the right of v_i and the 80 pixels to the left of v_i (for left breast). Thus $S_i = \{(x_{i-30}, y_i), (x_{i-29}, y_i), \ldots, (x_i, y_i), (x_{i+1}, y_i), \ldots, (x_{i+80}, y_i)\}$. Let S_i^{mean}, S_i^{min} be the mean and minimum intensity of S_i, a new pixel $v_i' \in S_i$ is selected to replace v_i according to the following judgements.

1: **if** $S_i^{min} < 3$ **then**
2:     **if** $S_i^{mean} < 25$ **then**
3:         v_i' be the first pixel in S_i having intensity less than or equal to $S_i^{min} + 2$,
4:     **else**
5:         v_i' be the first pixel in S_i having intensity less than 10,
6:     **end if**
7: **else**
8:     **if** $S_i^{mean} \geq 10$ **then**
9:         v_i' be the first pixel in S_i having intensity equal to S_i^{min},
10:     **else**
11:         v_i' be the first pixel in S_i having intensity less than or equal to S_i^{mean}
12:     **end if**
13: **end if**

Depending on the physical profile of the breast boundary, for some breasts,

the slope of the boundary in the bottom part is big (according to the coordinate system defined in Fig. 5.1). For these breasts, the vertical push-out is applied to push the bottom part of the boundary out. In the vertical push-out, S_i consists of the 50 pixels above v_i and the 300 pixels below v_i. Thus $S_i = \{(x_i, y_{i-50}), (x_i, y_{i-49}), \ldots, (x_i, y_i), (x_i, y_{i+1}), \ldots, (x_i, y_{i+300})\}$. The criteria for the selection of the new v_i' is also different. The first pixel in S_i having intensity value less than 5 is selected as v_i'. In case no pixels in S_i have intensity value less than 5, then the first pixel having intensity value less than 10 is selected.

This push-out procedure was developed for the Mini-MIAS mammograms. Although very simple, in the experiments, it works well in drawing the initial breast boundary to the true boundary (Table 5.1).

5.2 Breast Boundary Extraction Based on ROW Filters and MRF Smoothing

In the initial method described in the previous section, both filters used to identify the background components and the thresholds used in the push-out procedure were determined empirically. The thresholds used in the filters for identification of background components were set according to the experimental database. In fact, in experiments, thresholds of 1800 and 15 are used for the size and average intensity for the Mini-MIAS database and 4500 and 60 for our local archive. For the threshold set used in the push-out procedure, the same set, chosen for the Mini-MIAS database, was also used for the local archive mammograms and produced satisfactory results. Another drawback of the initial method is its dependency on the AP segmentation. Poor segmentation of the background components may result in a poor breast boundary extracted.

To overcome the drawbacks of the initial method, a new method was developed, based on ROW wavelets filtering and MRF smoothing. The method uses ROW wavelets to pre-filter the image. A Canny edge detector is used on the filtered image, resulting in a binary image. An initial boundary is extracted by scanning the binary image. The initial boundary is then smoothed using MRF model and iterated conditional modes (ICM) relaxation.

5.2.1 Pre-filtering with ROW Filters

The ROW were derived by Auscher (1992) under the framework of rational multiresolution analysis (MRA). Baussard, Nicolier & Trucheter (2004) generalized this class of wavelets by permitting more general types of roll-off in the transition bands of the frequency spectrum. Yu & White (2005) derived explicit formulas for a special class of rational wavelets with rational dilation factor of $a = \frac{q+1}{q}$ and whose spectrum has no constant passband between two transition bands.

The ROW has recently found its application in communications (Yu & White 2007, Yu & White 2006). Application of ROW in image processing has not been explored. In this method, ROW is utilized to construct image filters which are highly adaptive to varying statistics of mammograms regarding the pixel intensity.

The corresponding ROW basis function $\psi(t)$ is defined in the frequency domain by

$$\Psi(\omega) = \begin{cases} (2\pi)^{-\frac{1}{2}} e^{j\frac{w}{2}} \sin(\frac{\pi}{2}\beta(\frac{q}{\omega_1}|\omega| - q)), & \omega_1 \leq |\omega| \leq \omega_2 \\ (2\pi)^{-\frac{1}{2}} e^{j\frac{w}{2}} \cos(\frac{\pi}{2}\beta(\frac{q}{\omega_2}|\omega| - q)), & \omega_2 \leq |\omega| \leq \omega_3 \\ 0, & |\omega| \notin [\omega_1, \omega_3], \end{cases} \tag{5.1}$$

and the corresponding ROW scaling function $\phi(t)$ is defined by

$$\Phi(\omega) = \begin{cases} (2\pi)^{-\frac{1}{2}}, & |\omega| \leq \omega_1 \\ (2\pi)^{-\frac{1}{2}} \cos(\frac{\pi}{2}\beta(\frac{q}{\omega_1}|\omega| - q)), & \omega_1 \leq |\omega| \leq \omega_2 \\ 0, & |\omega| > \omega_2, \end{cases} \tag{5.2}$$

where

$$\omega_1 = (q - \frac{q}{2q+1})\pi,$$
$$\omega_2 = a\omega_1,$$
$$\omega_3 = a\omega_2 = a^2\omega_1, \tag{5.3}$$

and $\beta(t)$ is the construction function which is not unique. One construction function that leads to fast decay property of the wavelet is defined by $\beta(t) = t^4(35 - 84t + 70t^2 - 20t^3)$. The variables ω, ω_1, ω_2 and ω_3 are in radians. The functions $\Phi(\omega)$ and $\Psi(\omega)$ are the Fourier transform (FT) of $\phi(t)$ and $\psi(t)$ respectively. The waveforms and spectra of three ROWs are illustrated in Fig. 5.2.

The two-dimensional (2-D) image filter is constructed based on the waveform of the ROW scaling function $\phi(t)$ (Eq. 5.2). More detail about the ROW can be found in Yu & White (2005), Yu & White (2006) and Yu & White (2007).

After the ROW wavelet filtering, a Canny edge detector is applied to the filtered image to detect the breast boundary. The low sensitivity and high sensitivity of the canny edge detector were set 0.18 and 0.26 empirically. Several ROW wavelet filtered results are shown in Fig. 5.3 - 5.5. The images in Fig. 5.4 are examples of difficult boundaries. The image in (a) has a label close to the boundary. The images in (b) and (d) have noisy backgrounds and images (b) and (c) include a bright line artifact. With ROW filter, the breasts in these images clearly emerged. However, ROW gave poor results in some cases (Fig. 5.5). In (a) of Fig. 5.5, the ROW filter causes a bright area in the top of the image. The corresponding piece of the breast boundary is missed by the Canny edge detector. In (b), the Canny edge detector failed to give a clear breast boundary, due to the dark band near the true boundary in the filtered image.
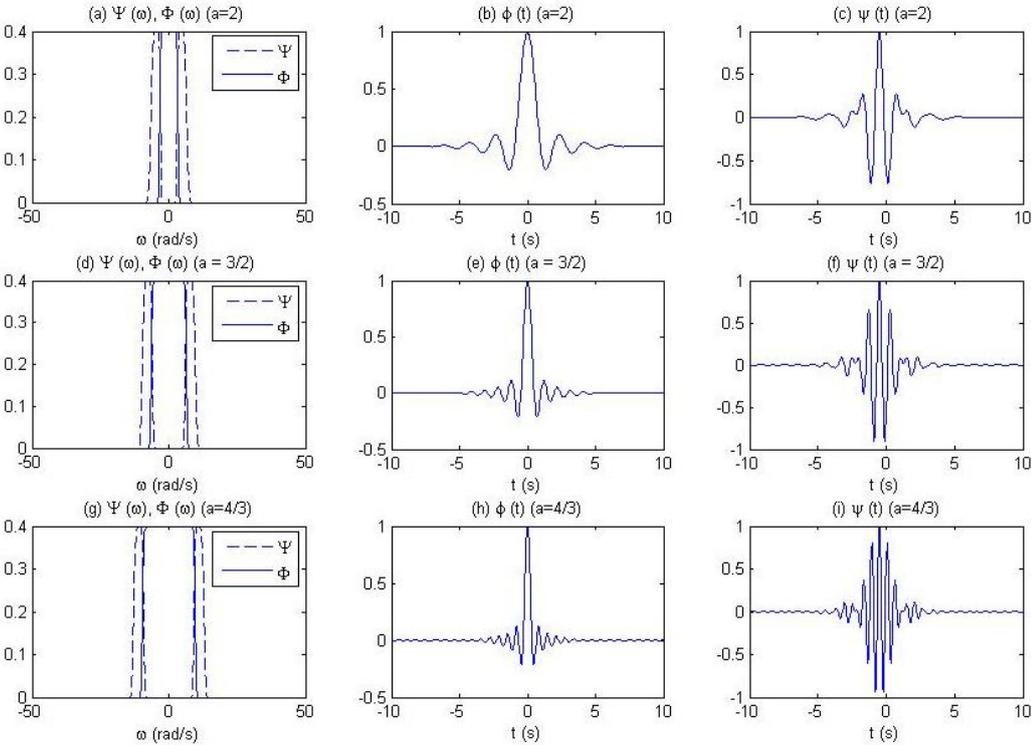
Figure 5.2. Real rational orthogonal wavelets with q = 1, 2 and 3.

(a) (b) (c)
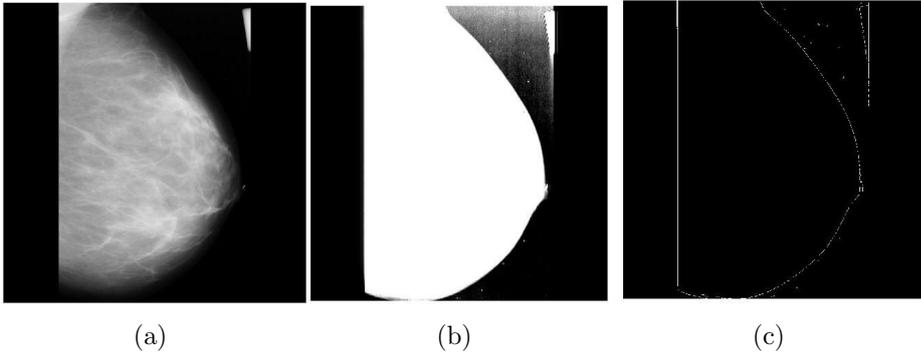
Figure 5.3: Example of ROW filtering and the Canny edge detection (a) original image, (b) ROW filtered image, (c) the Canny edge detected image.
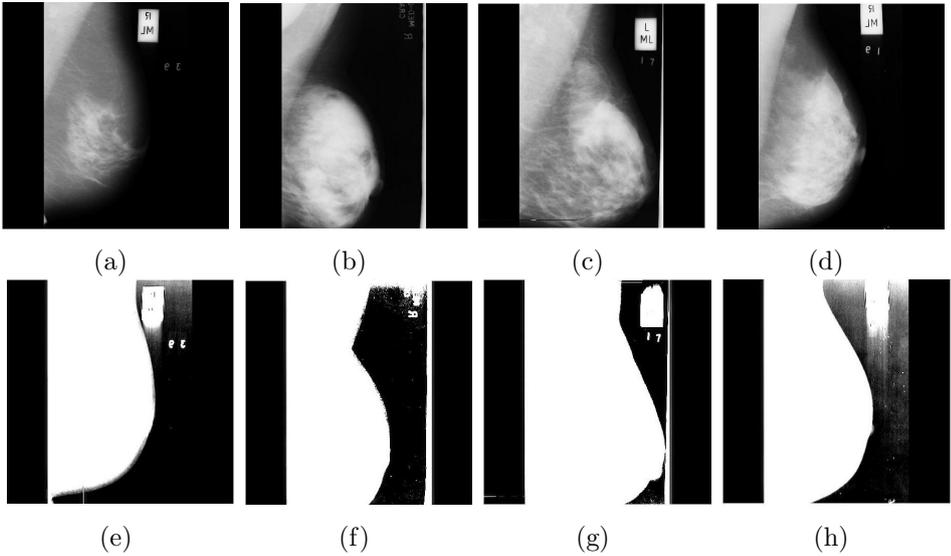
Figure 5.4: Examples of ROW filtered images. (a)(b)(c)(d) are the original images (mdb042, mdb068, mdb115, mdb004) and (e)(f)(g)(h) are the corresponding filtered images.

Figure 5.5: Examples of poor ROW filtered images. (a)(c) are the ROW filtered images for mdb034 and mdb090 and (b)(d) are the corresponding Canny edge detected results.

5.2.2 Breast Boundary Modelling and Smoothing via 1-D MRF and ICM

The initial breast boundary obtained by the canny edge detector has high fidelity compared with the true boundary marked by the radiologists, except for some discontinuities and noise introduced in conjunction with the binarisation and the scanning procedure. To smo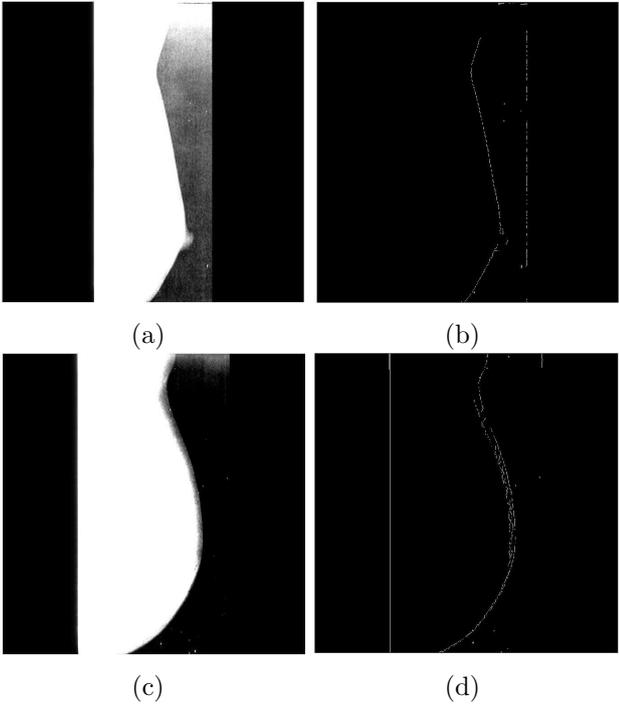oth the initial boundary, a 1-D MRF model is developed to model the breast boundary, and an iterated smoothing algorithm, the ICM algorithm, is applied to smooth the initial breast boundary. The MRF model and ICM algorithm were based entirely on the work by the first author of the study (Yu, Ma, Jayasuriya, Sigelle & Perreau 2007). For details of these two parts, readers are referred to this study (Yu et al. 2007). Compared with many other smoothing methods, which have reported good results in boundary smoothing, the advantage of the proposed smoothing algorithm is its capability in preserving nipple while smoothing the whole boundary.

The initial breast boundary obtained from the ROW filtered binary image is saved as a 2-D signal (x_i, y_i), $i = 1, \cdots, N$, where (x_i, y_i) is the x- and y-coordinates of the ith point in the boundary, and N is the length of the boundary. The physical profile of the breast determines that at the top and bottom parts of the boundary, one of the coordinates is single valued with respect to the other coordinates. The smoothing of the 2-D signal is simplified into a 1-D signal smoothing problem.

According to the physical profile of the boundary, the mammograms can be classified into two categories. In one case the y-coordinates are single valued along the whole boundary as shown in Fig. 5.6(a) and (b). In this case, the smoothing problem is a straightforward 1-D signal smoothing problem. The signal to be smoothed is $\{y_i\}$, $i = 1, \cdots, N$ while $\{x_i\}$ takes on the natural order. The other category, as shown in Fig. 5.6(c) and (d), the y-coordinates are not single valued along the whole boundary. For this category, we divide the boundary into two pieces, each of which is single valued at one coordinates with respect to the other coordinates, by finding a turning point of the boundary (Fig. 5.7). Two 1-D signals are generated accordingly and are smoothed separately. The signal regarding the top piece of the contour is $\{y_i\}$, $i = 1, \cdots, N_1$ and $\{x_i\}$ takes on the natural order. For the bottom piece, the signal to be smoothed is $\{x_i\}$, $i = 1, \cdots, N_2$ and $\{y_i\}$ takes on the natural order. The smoothing problem in the first category can be treated as a special case of the second category where only the top piece smoothing is required.

The two pieces are smoothed separately both using 1-D MRF model and ICM algorithm. The two smoothed boundaries are then combined to form the final whole breast boundary. To treat the possible discontinuity caused by 2-piece smoothing, a small portion of the initial boundary is retained near the connection point. This may slightly degrade the smoothness of the boundary near the connection point but ensures a smooth connectivity of the two pieces.
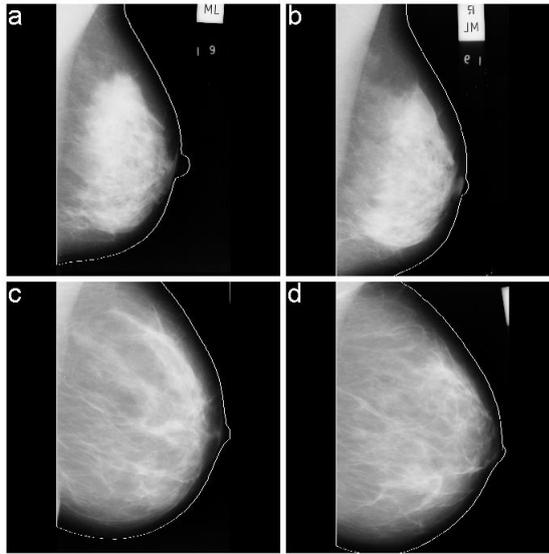
Figure 5.6: Two categories of the profile of the breast boundary. In (a)(mdb003) and (b)(mdb004), the y-coordinates are single valued with respect to the x-coordinates along the whole boundary. In (c)(mdb049) and (d)(mdb050), y-coordinates of the boundary are not single valued along the whole boundary.
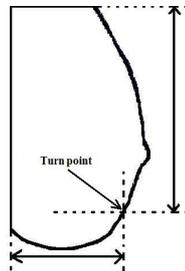
Figure 5.7: Division of the breast boundary into two single valued parts. The y-coordinates of the breast boundary with respect to the x-coordinates are not single valued along the whole boundary. A turn point is located and the whole boundary is divided into two parts, one coordinates of each of which are single valued with respect to the other coordinates. The two parts are smoothed separately by 1-D MRF smoothing.

5.3 Performance Analysis

To evaluate the performance, both of the methods were applied to 82 MLO view mammograms from the Mini-MIAS database. The same images were selected and used in Ferrari et al. (2004b). However, the images selected in Ferrari et al. (2004b) included 84 mammograms. Two cases are excluded in this analysis.

Analysis of these two cases is given in Section 5.4.

The detected boundaries are compared to the ground-truth (GT) boundaries. The same GT boundaries, used in Ferrari et al. (2004*b*), traced by a radiologist, were used in this thesis. These manually drawn boundaries were kindly provided by Rangaraj M. Rangayyan. Pixels assigned by the algorithm to the breast but assigned by the radiologist to background are labelled false positive (FP) and the pixels assigned by the algorithm to background but assigned by the radiologist as part of the breast are labelled false negative (FN). Table 5.1 presents the results obtained by both of the methods together with the results obtained in Ferrari et al. (2004*b*).

	ACM	AP boundary	RM boundary
FP_{ave}	0.41 ± 0.25	0.70	0.59 ± 0.78
FN_{ave}	0.58 ± 0.67	0.81	0.69 ± 0.67
$FP_I < 0.05, FN_I < 0.05$	33	22	46
$0.05 < FP_I, FN_I < 0.10$	38	24	14
$\max(FP_I, FN_I) > 0.10$	13	36	22

Table 5.1: FP and FN proportion and distribution: breast boundary extraction results of three methods. Here ACM refers to the active contour model reported by Ferrari et al. (2004*b*). AP boundary refers to our initial method and RM boundary refers to the ROW and MRF based method. The first two rows report the proportion of FP and the proportion of FN pixels averaged over all images. The next three rows indicate the distributions of FP and FN proportions. Standard deviations for the initial method (AP boundary) are not provided as the results were obtained earlier and no standard deviations were obtained at that time.

Two examples of the detection results obtained by the ROW and MRF based method are illustrated in Fig. 5.8 and Fig. 5.9. In Ferrari et al. (2004*b*), these two mammograms (mdb003 and mdb068) were singled out because of the bad results. The reason for mdb003 is the non-detection of the nipple region which leads to poor FN statistics. For image mdb068, both the FP and FN percentages are greater than 1% because the boundary was attracted to high density region in breast. With the ROW and MRF based method, we achieved satisfactory results for both cases. The FP and FN percentages are $FP = 0.32\%$ and $FN = 0.46\%$ for mdb003. The percentages are $FP = 0.15\%$ and $FN = 0.79\%$ for mdb068.

To investigate the performance of the second method in preserving nipples, we selected the mammograms whose nipples can be identified from the GT boundaries. Among the 82 mammograms, thirty four mammograms had visible nipples in the radiologist drawn boundary. For these thirty four mammograms, the average percentage of FP and FN were $FP = 0.50 \pm 0.58\%$ and $FN = 0.53 \pm 0.24\%$, respectively. Nipple information is preserved in the detected breast boundary in
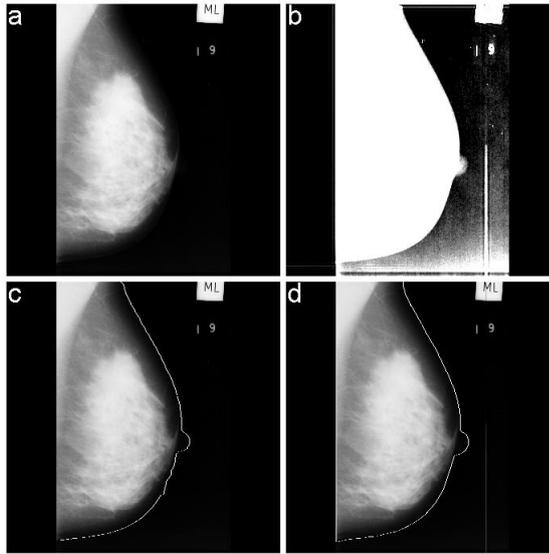
Figure 5.8: Breast boundary extraction for mdb003 (a) original mammogram, (b) filtered image, (c) detected contour, (d) contour drawn by radiologist
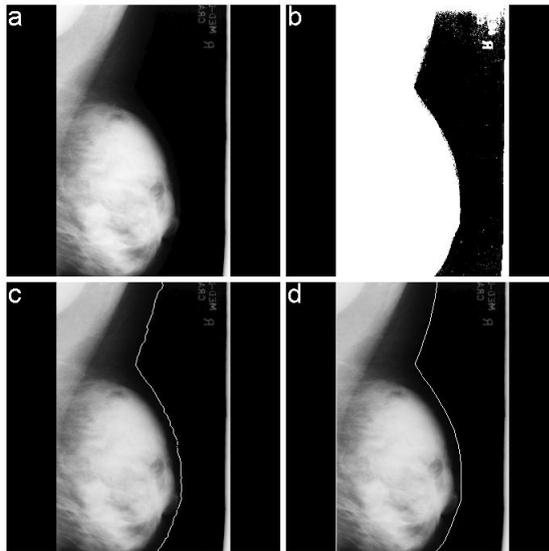
Figure 5.9: Breast boundary extraction for mdb068 (a) original mammogram, (b) filtered image, (c) detected contour, (d) contour drawn by a radiologist

31 images. In the other 3 cases, the contours of the nipple are weak with nipple information lost in only 1 case (Fig 5.10). An illustration of the filtering and ICM smoothing on the nipple area is shown in Fig. 5.11. The smoothness parameter $\beta = 100$ is selected in the analysis.

The results reported in this section are based on the image filter constructed
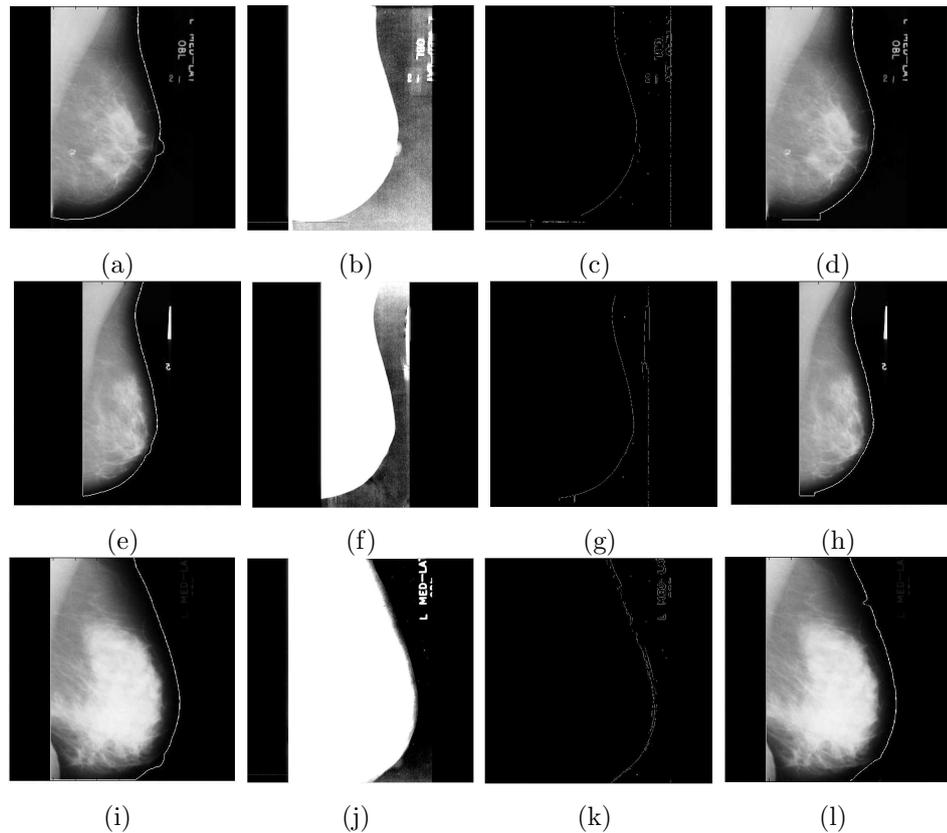
Figure 5.10: Examples of bad nipple preservation. (a)(e)(i) are the original images of mdb119, mdb044 and mdb105 with superimposed GT boundaries, (b)(f)(j) are the corresponding filtered images, (e)(g)(k) the canny edge detected images and (d)(h)(l) the original images with algorithm identified boundaries.
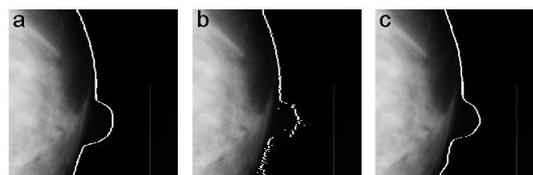
Figure 5.11: Example of contour smoothing around nipple area (mdb003) (a) GT contour, (b) initial contour, (c) smoothed contour

by ROW ($q = 3$). Compared with other ROWs with different dilation factors, such as ROW ($q = 2$) and ROW ($q = 4$), the performance of ROW ($q = 3$) is the best in statistics while ROW ($q = 2$) or ROW ($q = 4$) achieves better results for certain mammogram cases.

(a)   (b)

(c)   (d)

Figure 5.12: Dropped image case: mdb097. (a) Original image with superimposed GT boundary, (b) ROW filtered image, (c) the Canny edge detected image. (d) Original image with algorithm detected boundary. A piece of the boundary in the bottom is missed, causing misleading FP and FN. This image is excluded in the performance analysis.

5.4   Discussion and Conclusion

Images mdb097 and mdb106 were removed from this analysis. Fig. 5.12 and Fig. 5.13 show these two cases. Image mdb097 is removed because a piece is missed in the final boundary, which totally biased the result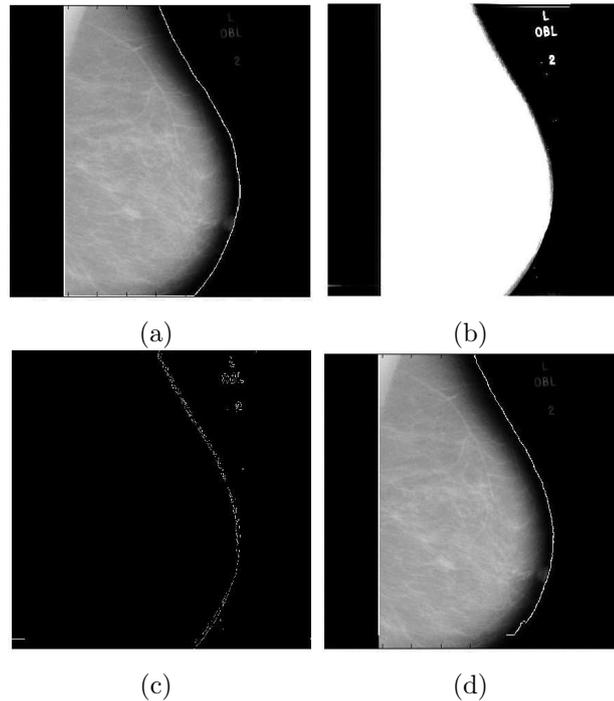s. The filtered image of mdb106 ((b) of Fig. 5.13) is very unclear and the canny edge detector can only capture part of the breast boundary, causing a very bad result. The histogram of the intensity of this image is very different from the other images. For this special case, we developed a simple histogram adjustment procedure to correct the histogram distribution. (d)(e)(f) of Fig. 5.13 show the results of histogram adjusted image mdb106. A good result is obtained with the histogram adjustment.

The robustness of the ROW and MRF based method lies in the robust ROW filtering owing to the flexible spectrum control via rational dilation factor specification, high regularity and fast decay properties of ROW (Yu & White 2006). The ROW is adaptive to varying statistics of mammograms regarding the pixel intensity. In fact, the method was applied to the mammograms from our local archive without changing any settings and good results were obtained. In comparison with techniques relying on the calculation of varying thresholds based on histogram analysis, a single fixed ROW image filter is sufficient for all mammo-
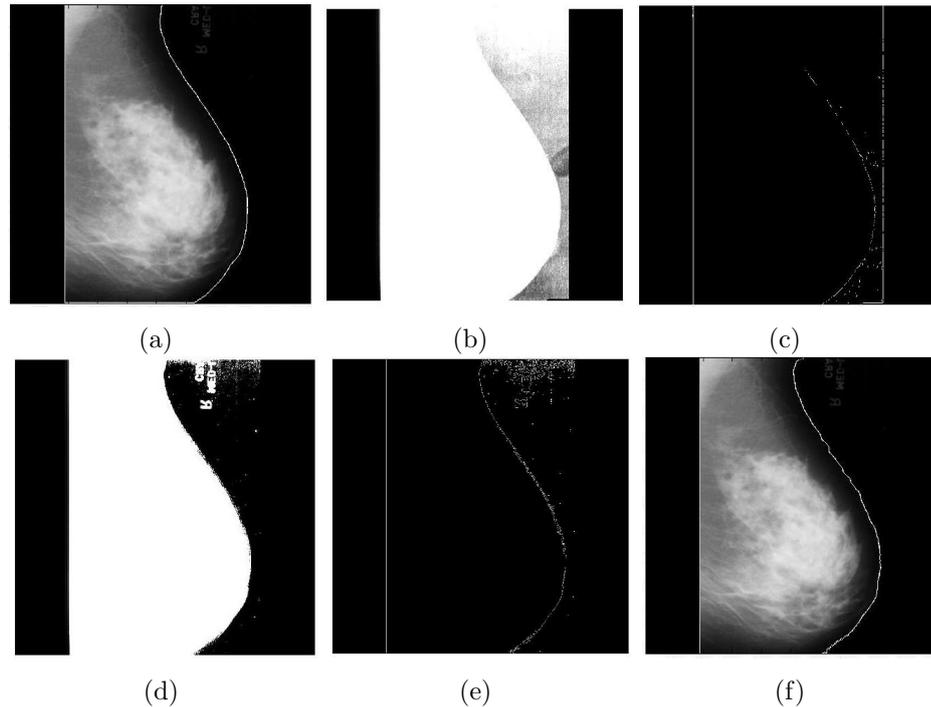
Figure 5.13: Dropped image case: mdb106. (a) Original image with superimposed GT boundary, (b) ROW filtered image, (c) the Canny edge detected image, (d) Histogram adjusted and ROW filtered image, (e) the Canny edge detected image after histogram adjustment and ROW filtering, (f) Image with final breast boundary.

grams being analyzed.

MRF theory and the ICM algorithm have been widely used in image processing (Geman & Geman 1984, Besag 1986, Dubes & Jain 1989). Based on the MRF modelling for breast contour, the ICM smoothing algorithm developed here smoothes the initial contour with fast convergence. The piece-wise smoothing technique introduced in section 5.2.2 transforms a 2-D MRF modelling and relaxation problem to a 1-D MRF modelling and therefore the complexity of the relaxation algorithm is greatly reduced.

Results prove the robustness of the proposed detection algorithm for 82 mammograms. We notice that the main reason for the degradation of the FN and FP statistics of the proposed algorithm is the gap along the top and bottom edges of the detected contour. This gap is related to the edge scan procedure after ROW filtering. The performance of the proposed method can be further improved by refining the edge scan step in the algorithm to mitigate the gaps in the initial contour.

Chapter 6

Mass Features

Features play the essential role in the classification of malignant masses. Various features have been explored in the literature. Among them, shape and texture features are the two common and important categories.

The shape of an object usually contains important information about the object. The shape features have been found important and useful in mass classification in mammography. Masses may have circumscribed, spiculated or stellate shapes. Cancerous lesion usually have a more irregular shape than benign lesions. Compared to the malignant masses, benign masses generally have better defined boundaries. Shape based features are expected to be effective in distinguishing these differences. Ackerman & Gose (1972) used spicularity, roughness and shape features extracted by computers to classify masses or clusters of microcalcifications as benign or malignant. Shen, Rangayyan & Desautels (1993, 1994) developed several position-, size-, and orientation-invariant shape features for classification of calcifications in mammograms. Brzakovic, M. & Brzakovic (1990) used area, shape, edge distance variation, and edge intensity variation determined by computer to differentiate benign from malignant lesions. In Sahiner, Petrick, Chan, Hadjiiski, Paramagul, Helvie & Gurcan's (2001) study, normalized radial length, standard deviation, entropy, area ratio and convexity are used to classify malignant masses.

The texture of an object contains information about the structural arrangement of the surfaces of the object and their relations to the surrounding environment. Many texture based features have been created and investigated in the literature. Haralick, Shanmugam & Dinstein (1973) used many texture features in category identification tasks in three different image data. Chan, Wei, Helvie, Sahiner, Adler, Goodsitt & Petrick (1995) and Wei, Chan, Helvie, Sahiner, Petrick, Adler & Goodsitt (1995) investigated the use of image texture features for classification of masses and normal tissue on mammograms. Their results indicated that image texture contains useful information that can be used to effectively distinguish masses from normal tissue. Mudigonda, Rangayyan & Desautels (2000) use gradient-based features and texture measures based on gray-level

co-occurrence matrices (GCMs) for the classification of mammographic masses as benign or malignant.

In this thesis, 7 shape based features and 5 texture based features are used for mass characterization. This chapter describes these features. A short summary of these features is given in Table 6.1. Section 6.1 and Section 6.2 give detailed definitions of these shape based features and texture based features respectively. Location of a ROI is an important feature for malignant mass identification. Section 6.3 describes a location feature based on a specially defined coordinate system. Based on the extracted features, a filter is set up to remove uninteresting components. Section 6.4 describes the details of the filter. To indicate how likely a component is a true mass, a mass like score is calculated for each component based on the component features. Calculation of the mass like score is described in Section 6.5.

The following notations are used in this chapter. The ith component of a mammogram is represented by R_i. $|R|$ represents the number of pixels in component R. The set of pixels outside component R but within t pixels distance is represented by O_t. $E(R)$ is the mean intensity value of the pixels in the component R and $\sigma(R)$ is the standard deviation of the intensity of pixels in R. $P(R, i)$ represents the proportion of pixels in component R with intensity i.

6.1 Shape Based Features

In this section, we introduce seven shape based features used in this thesis for characterizing components.

area. The number of pixels in the component (Brzakovic et al. 1990).

$$area = |R|$$

s1. The proportion of pixels in the convex hull that are also in the component (Sahiner et al. 2001). Here the convex hull is the smallest convex polygon that can contain the component. When an object is convex, this feature would approach its maximum value. On the contrary, the value of this feature would be small when the shape of the object is highly nonconvex, as is the case with many malignant masses.

$$s1 = \frac{\text{area}}{\text{area of convex hull}}$$

s2. The ratio of the area of the component to the area of the bounding box which is the smallest rectangle containing the component.

$$s2 = \frac{\text{area}}{\text{area of bounding box}}$$

ratio. Ratio between the length of the major and minor axis of the ellipse that has the same normalized second central moments as the component. Fig. 6.1 illustrates the axis of the ellipse.

$$ratio = \frac{\text{major axis length}}{\text{minor axis length}},$$
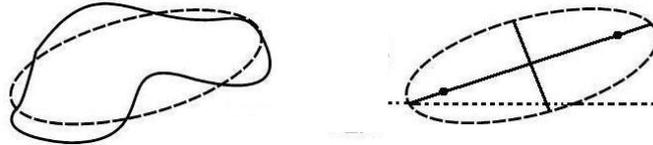
Figure 6.1: Axis of the ellipse having the same normalized second central moments as the component. Figure in the left side shows a segmented component with black boundary and the ellipse with dashed boundary having the same normalized second central moments as the component. The right side figure shows the same ellipse with the axis in solid lines. The two black dots are the foci of the ellipse.

dradi. Standard deviation of the radial distance (Sahiner et al. 2001). Here the radial distance is the Euclidean distance of edge points to the centroid of the component. Fig. 6.2 illustrates the radial distance.
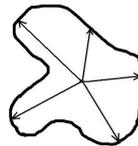
Figure 6.2: Radial distance of component. The radial distance is the Euclidean distance of edge points to the centroid of the component.

radi. The average value of the radial distance (Sahiner et al. 2001).

circularity. Circularity measures to what extent a component is circularly shaped (Gonzalez & Wintz 1987). The smallest value of this feature is 12.56. When the circle deviates towards more complicated shape, the value of this feature becomes larger. This feature is invariant to the translation, rotation, and scale.

$$circularity = \frac{p^2}{\text{area}}.$$

Here p is the perimeter of the component. A larger value of circularity describes a irregular and elongated object while a smaller value represents a more symmetric object.

6.2 Texture Based Features

This section introduces 5 texture based features, *int*, *relint*, *pri*, *c2* and *c3*, that are used in this thesis. The same features *c2* and *c3* were used in Timp & Karssemeijer (2006).

int. The average intensity value of the component.

$$int = E(R).$$

relint. The average intensity value of the component divided by the average intensity value of the whole breast.

$$relint = \frac{E(R)}{E(\text{Bst})},$$

here Bst represents the region of the whole breast.

pri. Intensity variance along the boundary of the component.

$$pri = \frac{|E(R) - E(O_2)|}{E(R) + E(O_2)},$$

c2. A measure of intensity contrast.

$$c2 = \frac{(E(R) - E(O_d))^2}{\sigma(R) + \sigma(O_d)},$$

here d is defined as

$$d = \sqrt{\frac{area}{\pi}}.$$

c3. A measure of intensity contrast.

$$c3 = \sum_i |P(R, i) - P(O_d, i)|,$$

here d is the same as defined in *c2*.

6.3 Location Based Feature

The location of a ROI within the mammogram can provide information useful for classifying the component as malignant mass. As reported by Caulkin, Astley, Asquith & Boggis (1998), the upper outer quadrant of the breast is more likely to develop malignant masses. Registration of the temporal mammogram might also benefit from the location information of the segmented components. However, the absolute locations of the components can not be used, due to the different compression and positioning of breast involved in image acquisition. A

Feature Name	Description
area	area of the component
s1	solidity of the component
s2	solidity of the component
ratio	ratio between length of major and minor axis
radi	average radial distance
dradi	standard deviation of the radial distance
circularity	ratio between perimeter and area
int	mean intensity of the component
relint	relative intensity
pri	intensity variance along the boundary
c2	intensity contrast
c3	intensity contrast

Table 6.1. List of component features used in this thesis.

coordinate system, based on the fitted line to the pectoral muscle boundary (for MLO view mammogram) or chest wall boundary (for CC view mammogram) and the longest distance from the fitted line to the breast boundary points (see Fig. 6.3), as also reported by Timp, Varela & Karssemeijer (Jul. 2007), may be helpful in providing location information. However, our experiments showed that this location feature is very sensitive to the pectoral muscle boundary (for MLO view mammogram) and breast boundary (results not shown). A little deviation in the boundaries may change the coordinate system substantially. In temporal mammogram registration, this sensitivity may cause components of temporal mammograms, that correspond to the same part of breast, having very different coordinates. For this reason, this location feature is not used in this study.
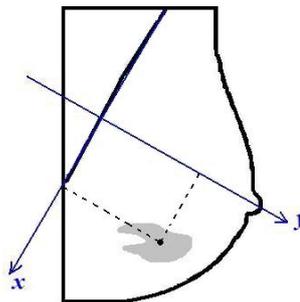
Figure 6.3: The coordinate system for MLO view mammogram. The coordinate system is defined with a fitted line to the pectoral muscle boundary as x axis and the line perpendicular to the pectoral muscle boundary fitted line and having longest distance from the fitted line to the breast boundary points as y axis. Positions of any components under this coordinate are determined regarding to the positions of the centroids of the components.

6.4   Feature Filter

The AP algorithm results in a tessellation of the breast by components (Chapter 3). In our experiments on 95 temporal mammogram pairs (Section 8.2.1), the number of segmented components per breast varies from the minimum 114 to the maximum 533. As the processing complexity of graph matching is highly related to the number of vertices, a large number of components increases the processing time to become prohibitively large. On the other hand, only components representing information relevant to the presence of breast cancer are of interest. A set of component features together with the breast template are used to reduce the number of components.

In our application, a template is extracted for each mammogram using the method introduced in Chapter 5, which is based on ROW filtering and MRF smoothing. The templates are applied to the components and any components that are outside the breast are filtered out.

Four features, including *area, ratio, s1, s2*, are used to further reduce the unwanted components. The reason for selecting these four features for the filtering is that, in our application, computation of these four features is relatively convenient. Two other features *int* and *relint* are naturally derived from the AP segmentation and thus are also convenient to include. However, these two features are more database dependant, and hence are not used. The filter is set by simply thresholding the four features. In the experiment, a training dataset (see Section 8.2.1) is used to select the thresholds. The maximum and minimum value of each feature within all the malignant masses of the training dataset are used as the thresholds.

6.5   Mass-like Score

A mass-like score is assigned to each component to indicate how likely a component is a mass. The mass-like score is used during the graph matching and the components with higher mass-like score are to be matched first.

Before computing the mass-like score, all features are normalized to zero mean and unit variance according to the following formula,

$$f' = \frac{f - E(f)}{\sigma(f)}.$$

To compute the mass-like score, features are linearly combined.

$$\text{mass-like score} = \sum_{i=1} a_i f_i,$$

where f_i is the ith feature. Parameters a_i are optimized based on Fisher linear discriminant analysis (LDA) on the training dataset.
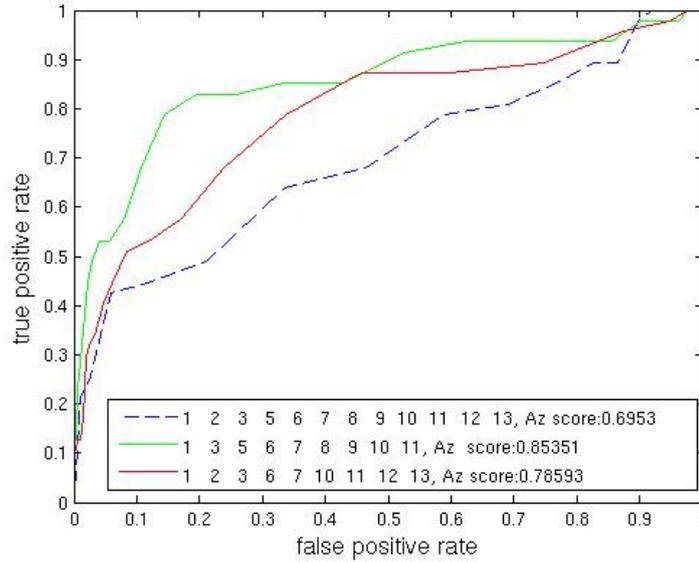
Figure 6.4: ROC curve for 3 different combinations of features based on the entire data set (see Section 8.2.1). The legend in the image shows the detailed combinations and their A_z score. Numbers in the legend represent different features. The correspondence of the numbers and the features are as follows: *s1-1, ratio-2, area-3, relint-4, int-5, dradi-6, pri-7, s2-9, radi-10, circularity-11, c2-12, c3-13.*

Based on reports in the literature and some initial experiments with our own data, some different combinations of features were tested for their ability to distinguish between malignant and benign masses. The receiver operating characteristic (ROC) analysis is used to evaluate the performance of these different subset of features in classification of malignant masses. For this purpose, the 190 mammograms (the same set of mammograms as described in Section 8.2.1) are divided into two groups, those having malignant masses forming one group and the others forming the other group. The area under the ROC curve, the A_z score, is used for the feature subset selection. Figure 6.4 shows the ROC performance of three different combinations of features together with the A_z scores. The highest A_z score is 0.85 reached by the combination of feature *s1, area, int, dradi, pri, relint, s2, radi* and *circularity*. Figure 6.6 shows the distribution of this combination. However, in our application, we used *solidity, ratio, area, dradi, pri, radi, circularity, c2* and *c3* for the computation of mass-like score. The *int* and *relint* are not selected, due to their database dependency. Most of the malignant masses in our experimental dataset are bright. Thus the inclusion of *int* and *relint* increases the performance in classification for the used dataset. Distribution of the used combination of features is shown in Figure 6.5.

Based on the mass-like score, the components are further filtered to reduce the number. In this thesis, for time processing reasons, only the components having
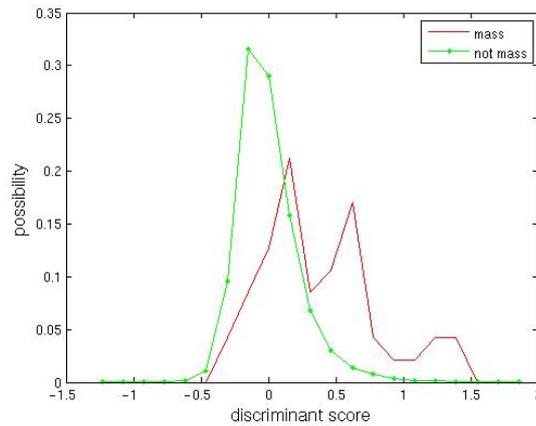
Figure 6.5: Distribution of the combination of the features used in this study. The features are *solidity, ratio, area, dradi, pri, radi, circularity, c2* and *c3*. This set of features is used for the mass-like score in this thesis.

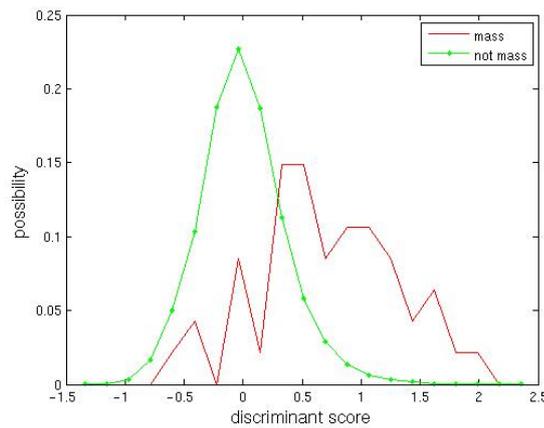**Figure 6.6:** Distribution of the best combination of the features. The features are *s1, area, int, dradi, pri, relint, s2, radi* and *circularity*.

mass-like scores in top 50 are kept for further investigation.

Chapter 7

Graph Matching

In this chapter we describe our method of matching corresponding mass-like objects in temporal mammograms based on spatial relations between the mass-like objects and a graph matching technique. A fuzzy spatial relation representation is used to represent the spatial relations between the mass-like objects. Section 7.1 describes the fuzzy spatial relation representation. A detailed description of the graph matching algorithm is introduced in section 7.2.

7.1   Fuzzy Spatial Relation Representation

A spatial representation between two objects given by the angle and distance between the centroid of the objects was initially used in our experiments. The angle is made by the line connecting the centroid of two objects and the vertical line, and the distance is measured between the centroid of two objects. Our experiments showed that this spatial representation only performs well when two objects are small and far apart. Fig. 7.1 illustrates the drawback of the angle and distance spatial representation.

In many situations, especially in representing spatial relations between complex objects, "all-or-nothing" relations lead to unsatisfactory results. Freeman (1975) proposed to use primitive spatial relations to describe the spatial relations between objects. He also proposed to use fuzzy set theory in representing spatial relations. Many approaches have been made in representing spatial relations based on the fuzzy set theory (e.g. Miyajima & Ralescu 1994, Matsakis & Wendling 1999, Bloch 1999). Rather than giving an explicit number (e.g. angle, distance) to describe the spatial relation, fuzzy spatial relation representation defines the spatial relation between objects as a fuzzy set. In this thesis, a fuzzy spatial relation representation introduced by Miyajima & Ralescu (1994) is implemented. This representation takes into account the geometric properties of the objects by using the histogram of all possible angles between all pairs of points in the considered objects (Fig. 7.2). In this method, four spatial relations, "right

(a)   (b)
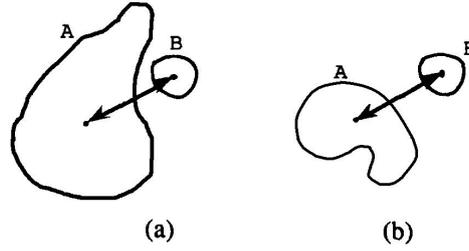
Figure 7.1: Ambiguity spatial representation by using centroid of objects. Using centroids of the objects to represent the spatial relation, objects in (a) and (b) have exactly the same spatial relation. However, shape, size and orientation of object A in (a) and (b) are different. The centroid spatial relation representation can not accurately express the spatial relations between the objects in this example.
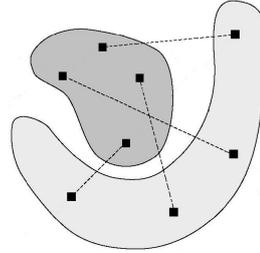
Figure 7.2: Fuzzy spatial relation representation. Angles between all pairs of points are calculated in the implemented fuzzy spatial relation representation.

of", "left of", "below", "above", are considered. The spatial relations between two points p and q are determined by the angle θ made by the line passing through p and q and the x-axis. The spatial relations between points p and q are fuzzy sets whose membership functions are given by the trigonometric functions.

$$\mu_{\text{right}}(\theta) = \begin{cases} \cos^2 \theta & \text{if } \frac{-\pi}{2} \leq \theta \leq \frac{\pi}{2}, \\ 0 & \text{otherwise}, \end{cases}$$

$$\mu_{\text{below}}(\theta) = \begin{cases} \sin^2 \theta & \text{if } 0 \leq \theta \leq \pi, \\ 0 & \text{otherwise}, \end{cases}$$

$$\mu_{\text{above}}(\theta) = \begin{cases} \sin^2 \theta & \text{if } -\pi \leq \theta \leq 0, \\ 0 & \text{otherwise}, \end{cases}$$

$$\mu_{\text{left}}(\theta) = \begin{cases} \cos^2 \theta & \text{if } -\pi \leq \theta \leq \frac{-\pi}{2}, \frac{\pi}{2} \leq \theta \leq \pi, \\ 0 & \text{otherwise}. \end{cases}$$

For two objects $A = \{a_1, a_2, \ldots, a_n\}, B = \{b_1, b_2, \ldots, b_m\}$, to calculate the spatial relations between them, all $m \times n$ pairs of points $(a_i, b_j), i = 1, \ldots, n, j = 1, \ldots, m$, are considered. Let Θ denote the collection of angles $\theta_{ij} = \angle(a_i, b_j), a_i \in A, b_j \in B$. As different pairs of points may have same angle, Θ is a multiset. For each $\theta \in \Theta$, let f_θ be the number of pairs (a_i, b_j) for which $\angle(a_i, b_j) = \theta$, the histogram associated with Θ is defined as

$$H_\Theta(A, B) = \{(\theta, f_\theta)\}.$$

The frequency f_θ is normalized by the largest frequency, $f_\theta = f_\theta / \max(f_\theta)$.

To evaluate the degree to which the spatial relation between H_Θ and each of four spatial relations holds, the histogram H_Θ is treated as a fuzzy set with membership function μ_H defined as

$$\mu_{H_\Theta}(\theta) = f_\theta, \text{ where } (\theta, f_\theta) \in H_\Theta.$$

The problem of to what extent is H_Θ each of four spatial relations is then treated as a problem of compatibility of fuzzy sets (Miyajima & Ralescu 1994). Let G be one of the four spatial relations and μ_G be the corresponding membership function, the compatibility of the fuzzy set H_Θ to a fuzzy set G is a fuzzy set $CP(H_\Theta; G)$ whose membership function is defined as

$$\mu_{CP(H_\Theta;G)}(v) = \begin{cases} \sup_{\theta, v = \mu_G(\theta)} \mu_{H_\Theta}(\theta) & \text{if } \mu_G^{-1}(v) \neq \emptyset, \\ 0 & \text{if } \mu_G^{-1}(v) = \emptyset. \end{cases}$$

To obtain the final degree to which a spatial relation holds, the compatibility fuzzy set $CP(H_\Theta; G)$ is defuzzyficated using the center of gravity of the fuzzy set

$$\frac{\sum_v v \cdot \mu_{CP(H_\Theta;G)}(v)}{\sum_v \mu_{CP(H_\Theta;G)}(v)}.$$

7.2   Graph Matching

Graph matching for temporal mammogram registration is inexact graph matching. This is because on one hand, the same object of a breast may appear very different in temporal mammograms. This may be caused by natural changes in breast tissue, different dosages, deformation at image acquisition in combination with the fact that mammograms are projection of 3D objects onto 2D images. On the other hand, mammograms are very complex images. Accurate isolation of the ROIs is a difficult task. Thus it is highly unlikely that the segmentation results of temporal mammograms are strictly identical or a graph isomorphism can be found between the graphs of temporal mammograms.

In many image analysis applications, matching of the objects in analyzed images is based on a set of features measured on the segmented objects. In

temporal mammogram analysis, however, the success of feature based matching is limited by the variations of the segmented objects which may be caused by the changes of the temporal mammograms or by the segmentation.

This section describes our graph matching algorithm that is based on the spatial relations between the mass-like objects.

7.2.1   Complete Graph Representation

In many applications, graphs are built based on the neighborhood relations of the segmented components, that is, two vertices in the graph are connected if the components they represent are neighbors. In this application, a graph is naturally generated with the AP segmentation algorithm to represent the segmented components. As the AP algorithm is based on the neighborhood relations of the vertices, so is the generated graph. The AP produced graph was initially used in our graph matching. However, the matching based on the generated graph results in false matches (Fig. 7.3). In this thesis, after segmentation and filtering, a *weighted complete graph* (see Section 3.1) is used to represent the mammogram with each vertex in the graph representing a mass-like object of the mammogram.
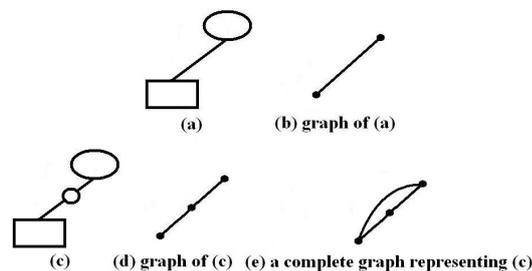
Figure 7.3: Problem with graph constructed based on neighborhood relation of objects. The big circle and big rectangle of (a) should match the big circle and big rectangle of (c). Because of the complexity of mammogram segmentation, a small extra circle in (c) adds an extra node in its corresponding graph (d). Graph matching based on (b) and (d) can only result in the big circle of (a) matching to the big circle of (c), the rectangle of (a) matching to the small circle of (c) or the big circle of (a) matching to the small circle of (c), the rectangle of (a) matching to the rectangle of (c), either of which yields a false match. The desired match can be reached when (d) is replaced by (e).

7.2.2   Match Cost Function

In inexact graph matching, a match cost function is usually involved to measure the similarity of subgraphs and thus to allow comparison between the different

solutions. The match cost function is usually based on the weights or attributes associated with vertices or edges. In this thesis, to find the best common subgraph between two weighted complete graphs that has the most similar spatial relations between the vertices of the graphs, the difference of four spatial relations between vertices is used to measure the spatial similarity. The similarity of spatial relations between an individual pair of matched vertices is defined as follows. Let $H_c = \{V_c, E_c\}$, $V_c = \{v_c^1, v_c^2, \ldots, v_c^n\}$, denote the subgraph of the weighted complete graph G_c of a current mammogram M_c and $H_p = \{V_p, E_p\}$, $V_p = \{v_p^1, v_p^2, \ldots, v_p^n\}$, denote the subgraph of the weighted complete graph G_p of a previous mammogram M_p. Suppose $v_c^k \in V_c$ matches $v_p^k \in V_p$. The similarity between v_c^k and v_p^k in terms of spatial relation μ_t is defined based on the difference of corresponding spatial relations between v_c^k to $V_c \backslash \{v_c^k\}$ and v_p^k to $V_p \backslash \{v_p^k\}$

$$C(v_c^k, v_p^k) = \sqrt{\sum_{i=1}^{n} (\mu_t(v_c^k, v_c^i) - \mu_t(v_p^k, v_p^i))^2}.$$

Here t is one of the four spatial relations and $\mu_t(v_m, v_n)$ is the μ_t spatial relation between vertices v_m and v_n. $V_c \backslash \{v_c^k\}$ is the set V_c but with element v_c^k removed. $\mu_t(v_c^k, v_c^k)$ is set to 0.

Based on the spatial similarity measure for the single match, the global spatial similarity measure of subgraphs H_c and H_p in terms of spatial relation μ_t is calculated as follows

$$\phi(\mu_t) = \sum_{i=1}^{n} C(v_c^i, v_p^i).$$

The final match cost function $\Phi(H_c, H_p)$ is the sum of four spatial relations

$$\Phi(H_c, H_p) = \phi(\mu_{\text{left}}) + \phi(\mu_{\text{right}}) + \phi(\mu_{\text{above}}) + \phi(\mu_{\text{below}}). \tag{7.1}$$

However, as only spatial relations between mass-like objects, which are local information, are used, it may happen that one of the matched subgraph globally offsets the other one, causing match error. Figure 7.4 gives an example of this problem. To avoid the global offset, the breast boundary is treated as a mass-like object and the spatial relation between other mass-like objects and the breast boundary is included in the match cost function. Let B_c, B_p be the breast boundaries of current and previous mammogram M_c, M_p. The similarity between v_c^k and v_p^k in terms of spatial relation μ_t is redefined as

$$C(v_c^k, v_p^k) = \sqrt{\sum_{i=1}^{n} (\mu_t(v_c^k, v_c^i) - \mu_t(v_p^k, v_p^i))^2 + (\mu_t(v_c^k, B_c) - \mu_t(v_p^k, B_p))^2}. \tag{7.2}$$

Breast boundaries are used in many multiple mammogram registration methods to provide partial global registration of multiple mammograms (see Chapter 2,
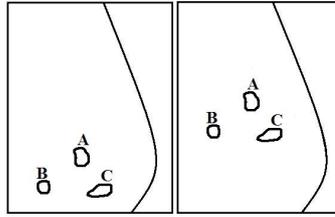
Figure 7.4: Example of global offset of mass-like objects. The spatial relations between the mass-like objects of the two mammograms are exactly the same, but mass-like objects in one mammogram are significantly shifted with respect to the mass-like objects of the second mammogram.

Section 2.1). In most of these methods, the breast boundaries of multiple mammograms are aligned and mammograms are transformed based on the aligned boundaries. For these methods to work, the breast boundaries need to be both accurately extracted and aligned, which is not a trivial task. The fuzzy spatial representation adopted in this thesis enables the spatial relation between any mass-like objects and the boundaries to be represented. By using the spatial relations between mass-like objects and the boundary in measuring the similarity of the subgraphs, we avoid the strict condition of accurate alignment of breast boundaries and still promise a global reference.

7.2.3   Identification of the Best Solution

Many inexact graph matching applications restrict themselves to finding a vertex in the bigger graph for each vertex of the smaller graph to match. For these applications, the length of the solution is fixed and the best solution is reached when the match cost function attains its minimum or maximum. Here a solution means a common subgraph of two graphs. The length of the solution is the number of the vertices of the common subgraph (the size of the common subgraph). In some real graph matching applications, however, vertices in the smaller graph may not have correspondences in the bigger graph. This can be due to the segmentation, if for example, the corresponding objects are segmented to be combined with other objects, or due to the fact that a new tumor has developed and so does not have a correspondence, or if filters are applied and one object of a corresponding pair is removed. For this type of inexact graph matching, the length of the solution is not fixed, and in many cases, is unknown. Finding the solution for this type of inexact graph matching is thus more complex. Some methods address this problem by using so-called *null vertex* or *dummy vertex*, which represents a null value. These methods still find at least one vertex in the bigger graph for each vertex of the smaller graph to match, but allowing vertices in the smaller graph to match the dummy vertex.

Due to the changes that can happen in the temporal mammograms, as mentioned before, temporal mammograms, even taken in a short period of interval, can be quite different. The difference between the corresponding segmented mass-like objects in a temporal mammogram pair is exaggerated by the difficulties in segmenting complex mammograms. Thus vertices in one of the graphs do not necessary have corresponding vertices in the other graph. To pick up the best sequence of matches during graph matching, not only the same length sequences of matches need to be compared, but also the different length sequences of matches. Kumar, Sallam & Goldgof (2001) reported the same problem. To solve the problem, Kumar et al. (2001) combined the matching results obtained by using both the Greedy Algorithm and Hungarian Method and simply ignored the conflicted matches. Unlike the case in this thesis, however, Kumar et al. (2001) used a weighted bipartite graph and matched two sets of vertices of the bipartite graph by finding a match vertex in one set for each of the vertices of the other set. Also, the graph matching in this thesis is based on the backtracking schema. We developed a new strategy to deal with the problem. We first find the best sequence of matches for each possible sequence of matches length. Thus for two graphs G_1 and G_2, assume the longest possible sequence of matches has length k, we find k sequences of matches $S_1, S_2, \cdots, S_k$, with $S_i = \{(x_{i,1}, y_{i,1}), (x_{i,2}, y_{i,2}), \cdots, (x_{i,i}, y_{i,i})\}$ the best sequence of matches having length i. We then combine all these sequences of matches together to form one final sequence $S = S_1 \cup S_2 \cup \cdots \cup S_k$. The final solution is derived by further filtering this final sequence.

7.2.4 Graph Matching Algorithm

In this section, we develop a graph matching algorithm which is based on the backtracking schema. This algorithm is based on the work proposed by Krissinel & Henrick (2004).

Graph matching is a computationally expensive process, with the increase of the size of graphs, the processing time increases exponentially. To keep the running time in a reasonable range, for each mammogram, we only use the components whose mass-like score is in the top 50. Experiments on the training data set shows that most malignant masses are within top 10, a worst case is with the malignant mass on a position 28. Thus 50 is a safe number to reduce processing components whilst retaining a very high percentage of masses.

Function *Backtrack* (Function 1) lists the main steps that are used to recursively enumerate all possible matches to find the best solutions. The function accepts two parameters, M and *Depth*. *Depth* records the number of pairs that have been matched in current solution. The matched vertices are kept in two vectors X and Y. For graphs $G_1 = \{V_1, E_1\}$ and $G_2 = \{V_2, E_2\}$ with $|V_1| = n_1$ and $|V_2| = n_2$, M is defined as an $n_1 \times n_2$ logic matrix. $m_{ij} = 1$ indicates that the ith vertex of G_1 can be matched to the jth vertex of G_2.

Function 1 $Backtrack(M, Depth)$

1: $p = PickupVertex$
2: **if** not p is null **then**
3:     $Depth = Depth + 1$
4:     **for all** vertex q in G_2 **do**
5:       **if** $M(p, q) \neq 0$ **then**
6:         Store p, q to X, Y
7:         $M1 = Refine(M)$
8:         $EvaluateMatch(X, Y, Depth, M)$
9:         $Backtrack(M1, Depth)$
10:       **end if**
11:     **end for**
12:     $Depth = Depth - 1$
13:     $M1 = M$
14:     $M1(p, :) = 0$
15:     $Backtrack(M1, Depth)$
16: **end if**

In each entry to $Backtrack$, the function $PickupVertex$ is called to pick up a vertex p from G_1 to match. If the current solution is still extendable, then p will be not null, thus a new pair is to be formed and $Depth$ increases by 1. Vertex p of graph G_1 is then tested as a match to all unmatched vertices q in G_2 in line 4. If in step 5, p and q are matchable, the new match p and q is added to the current solution by storing them in X and Y. Next, a $Refine$ function is called to refine matrix M based on the new formed match. $Backtrack$ is recursively called after the refinement to search for the next match for the current solution. After all the vertices of G_2 has been tested as a match to p, in step 14 of $BackTrack$, p is temporarily removed from G_1 by setting the pth row of M to 0. This is because the best solution may not contain the particular vertex p of G_1. After removing p, $BackTrack$ is called again to search for the best solution not containing particular vertex p.

For a simple version, $PickupVertex$ only needs to pick up an unmatched vertex of G_1. However, complex selections can be combined into the picking up procedure to facilitate any particular purpose. In this study, we combine the mass-like score to pick up the next vertex to match. Within all the remaining vertices of G_1, we always pick up the one having biggest mass-like score. The bigger the mass-like score is, the more likely the corresponding mass-like object is a mass. By picking up the vertex with largest mass-like score first, the matching process is focusing more on the suspicious mass-like objects.

Whenever a new pair has been matched, matrix M is refined to remove the search branch that will not lead to the best solution. The main idea of this refinement is based on the spatial relation compatibility. Assume $X = \{x_1, x_2, \ldots, x_k\}, Y = \{y_1, y_2, \ldots, y_k\}$ is the current solution. Then for any un-

mapped vertices x_{k+1} of G_1, a unmapped vertex y_{k+1} of G_2 is matchable to x_{k+1} if for any spatial relation μ_t, $\mu_t(y_i, y_{k+1}) \propto \mu_t(x_i, x_{k+1}), i = 1, \ldots, k, t =$ left of, right of, above, below. Here $\propto$ denotes compatibility. For a perfect case, the spatial relations related to the matched vertices should be exactly the same, thus the spatial compatibility between spatial relations $\mu_t(y_i, y_{k+1})$ and $\mu_t(x_i, x_{k+1})$ for this ideal case means

$$|\mu_t(y_i, y_{k+1}) - \mu_t(x_i, x_{k+1})| = 0, i = 1, \ldots, k, t = \text{left of, right of, above, below.}$$

However, nonrigid changes between the temporal mammograms may cause slight differences of the spatial relations between the corresponding matched vertices. The differences are usually very small. In this graph matching algorithm, we relax the definition of the spatial relation compatibility to allow for small nonrigid changes that may happen to the temporal mammograms. The spatial compatibility between spatial relations $\mu_t(y_i, y_{k+1})$ and $\mu_t(x_i, x_{k+1})$ is defined as

$$|\mu_t(y_i, y_{k+1}) - \mu_t(x_i, x_{k+1})| < \gamma, i = 1, \ldots, k, t = \text{left of, right of, above, below.}$$

As nonrigid changes may happen in all four directions, the four spatial relations are treated equally and the same γ is used for all four spatial relations. In this thesis, γ was determined empirically and $\gamma = 0.1$ was used for all four spatial relations "left of", "right of", "above", "below".

The function *EvaluateMatch* is called after *refine* whenever a new pair is formed. *EvaluateMatch* calculates the match cost of current solution according to Eq. 7.1. The best solutions of different length are saved in a global variable. The match cost of current solution is compared to the same length best solution that has been found so far. The length of the current solution is recorded by *Depth*, so the match cost of the current solution will be compared to the stored best solution that also has length *Depth*. If the current solution is better than the stored best solution, the current solution will replace the stored best solution.

7.2.5 Final Result Filtering

After graph matching, the best solutions of different length are combined together. For two graphs G_1 and G_2, assuming that longest possible sequence of matches has length k, we have k sequences of matches $S_1, S_2, \cdots, S_k$, with $S_i = \{(x_{i,1}, y_{i,1}), (x_{i,2}, y_{i,2}), \cdots, (x_{i,i}, y_{i,i})\}$ the best sequence of matches having length i. All these sequences of matches are then combined together to form one final sequence $S = S_1 \cup S_2 \cup \cdots \cup S_k$. It is obvious that this final sequence of matches S may contain duplicate matches and false matches. The final sequence of matches is then cleaned to remove duplicate matches. To reduce false matches, we next apply a filter on the final sequence S. The filter is based on the similarity of fuzzy spatial relations of the matched mass-like objects to the boundaries of the breast and the relative gray level differences.

The similarity of fuzzy spatial relations of the matched mass-like objects to the boundary is computed as follows. For matched pair (x_i, y_i), the similarity d_i of fuzzy spatial relation of x_i, y_i to the corresponding boundaries of current and previous mammogram B_c, B_p is defined as

$$d_i = \sum_{t=\text{"left of", "right of", "above", "below"}} |\mu_t(x_i, B_c) - \mu_t(y_i, B_p)|.$$

If $d_i > E(D) + \sigma(D)$, matched pair (x_i, y_i) is removed. Here $D = \{d_1, d_2, \cdots, d_n\}$ with n being the number of total pairs in the final sequence and $E(D), \sigma(D)$ are the mean and standard deviation of D.

The fuzzy spatial relation promises compatible spatial relations between the mass-like objects but does not promise compatible gray level of the matched mass-like objects. Thus it is possible that a bright mass-like object in one mammogram matches to a dark mass-like object in the other mammogram. The relative gray level difference is used to remove these false matches. As the intensities of mass-like objects of different mammograms are not directly comparable, the gray level of one mammogram is transformed to match the gray level histogram of the paired mammogram. The first step of this transformation is to calculate the cumulative histograms of the intensity inside the breast area. For intensity level t, the cumulative histogram $f(t)$ of the mammogram is defined as

$$f(t) = \frac{1}{Q} \sum_{i=0}^{t} H_{histo}(i),$$

where H_{histo} is the intensity histogram calculated inside the breast area, and Q is the total number of pixels inside the breast area. Let f_1 be the cumulative histogram of the first mammogram and f_2 be the cumulative histogram of the paired mammogram. Each gray level t in f_1 is then transformed to

$$\tilde{t} = f_2^{-1}(f_1(t)).$$

After the histogram match, the relative gray level difference of matched pair (x, y) was measured as

$$\text{Diff} = \frac{|E(x) - E(y)|}{E(x) + E(y)},$$

where $E(x)$ is the mean gray level of mass-like object x.

A threshold ξ is then set, and those pairs with $\text{Diff} > \xi$ are filtered out. In this thesis, based on a training dataset, ξ is set to 2.45.

Chapter 8

Experiments and Results

In this chapter we investigate the performance of the proposed registration method. The method is applied both to constructed mammogram pairs and to real temporal mammogram pairs. The constructed mammogram pairs are used to investigate the performance of the matching process independent of the segmentation process. Section 8.1 introduces the experiments with constructed mammogram pairs. To evaluate the performance of the method on real temporal mammograms, 95 pairs of temporal mammograms are selected from a local screening archives. Section 8.2 describes the experiments on the temporal mammograms.

8.1 Experiments with Constructed Mammogram Pairs

In this thesis, the results of the AP segmentation algorithm (Chapter 3) are used as the input to the proposed registration method. Good matching depends on the quality of the segmentation step as well as the quality of the matching algorithm itself. In order to evaluate the performance of the matching step independently of the segmentation, this section describes the performance of the matching process on constructed mammogram pairs. By this, we mean that the matching process is applied to a real mammogram and a derived version of the mammogram obtained by applying known transformations to the original mammogram. The constructed mammogram is designed to imitate the deformation that can be expected over time. It provides an ideal "segmentation" which allows evaluation of the performance of matching process eliminating a poor segmentation effect.

Artificial data are used in many mammogram registration applications to evaluate the registration method. Evaluation of complex registration with artificial data is convenient as "true answers" are already known. In many of these applications, a new mammogram is created by transforming an existing real mammogram. The registration method is then applied to the real mammogram and the

transformed mammogram. A slightly different approach is taken in this section. The proposed method is also applied to the real mammograms and their derived images. The real mammograms are segmented using the AP algorithm resulting label images. To create an "ideal" segmentation and thus to avoid segmentation effect, instead of using AP to segment the derived images, the "segmentation" to the derived images is realized by transforming the label images of the original mammograms in a same way that is used to obtain the derived images from the original mammograms. This way, the generated label images and their original label images originate from the same segmentations.

8.1.1   Database and Methods

For this experiment, 18 MLO view mammograms from Mini-MIAS database were used. As there were no particular requirements to the images, we simply used the first 20 images (mdb001 to mdb020). Images mdb007 and mdb008 were excluded because relevant data of these two images, such as pectoral muscle boundary and breast boundary, were not available. All images were preprocessed as described in Section 3.4.1 so that the resulting images were of size 256×256 which corresponds to spatial resolution of 800 μm per pixel. All 18 images were preprocessed to obtain the pectoral muscle boundaries and breast boundaries as described in Chapter 4 and Chapter 5.

All 18 images were segmented by the AP segmentation algorithm. A label image was generated from the segmentation for each mammogram with all pixels belonging to a same segmented component being labelled by the same number. The label image was relabelled, shifted and warped to form new label images. By relabelling, shifting and warping the label image, the resulted new label images contained exactly the same segmentations. The shifting and warping are designed to imitate the deformation that can be expected between temporal pairs of mammograms. With shifting and warping, the images are slightly deformed. Relabelling does not change the images at all, it simply relabels each segmented component.

The registration method was applied to the image pairs consisting of the original images and the derived images. All images were filtered before the matching process by both template filter and feature filters. The pre extracted pectoral muscle boundary and breast boundary were used to create a template for each image. To set the feature filters, a different set of 20 mammograms was selected from Mini-MIAS database. The images were selected based on the criteria that a mass exists in the image (either benign or malignant). All 20 images were segmented using AP algorithm. We then recorded the component that mostly covers the mass for each image. We call these components the mass objects. Features were calculated for all mass objects and were used to set the feature filter and mass-like score. The features used for the feature filter include *area, s1, s2,*

relint, ratio. In this experiment, the features were not normalized as introduced in Section 6.5. The settings of the filters are as follows.

- area < 10

- s1 < 0.35

- s2 < 0.18

- relint < 1.2

- ratio > 8.0 and relint < 1.38

- area > 1000 and s1 < 0.5

The mass-like score was calculated based on four features, *s1, dradi, pri, relint.* Based on the calculated features of mass objects and Fisher LDA, the mass-like score was calculated as follows

$$
\begin{aligned}
\text{mass-like score} = {} & 0.8649 \times s1 + 0.0073 \times dradi \\
& + 0.1930 \times pri + 0.4633 \times relint
\end{aligned}
\tag{8.1}
$$

Based on the mass-like score, for the convenience of processing, in the experiments, only components with their mass-like score in the top 20 were kept for the matching process. In many images, however, the total number of remaining components was less than 20. In these cases, all remaining components were kept for the matching.

Three separate experiments were conducted. In the first experiment, described in section 8.1.2, no transformation was applied to create the second image. The second image was created by relabelling the segments in the original image. In the second experiment, described in section 8.1.3, the second image was created by applying a shift to the original mammogram and segmentation to the second image was obtained by relabelling the original label image and then applying the same shift. In the third experiment, described in section 8.1.4, the second image was created by warping the original mammogram and was segmented by relabelling and warping the original label image. Figure 8.1 outlines these three experiments. For convenience, in the following sections, the second images will be called the relabelled images, the relabelled and shifted images and the relabelled and warped images respectively.

Several parameters were used to describe the graph matching results. Total matches found is the number of matches that were identified by the graph matching process. As the relation between the original images and their constructed images is preknown, the correspondence between the mass-like objects of original and constructed images can be pre-identified. Total true matches represents
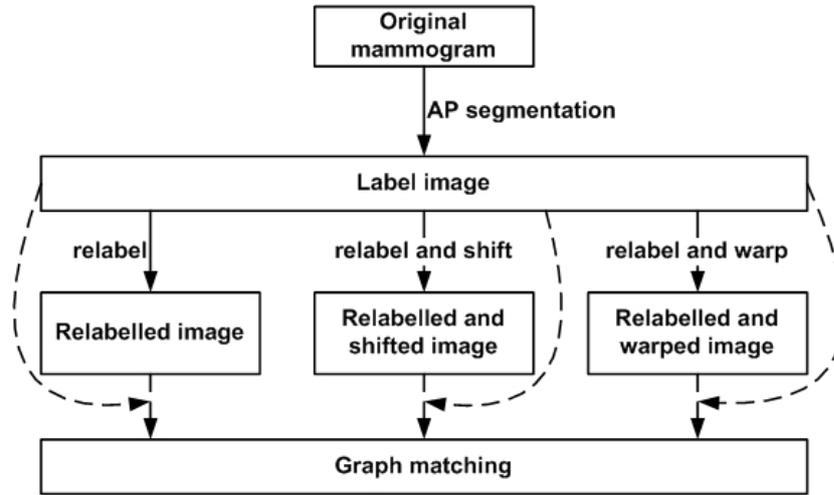
Figure 8.1. Outline of three experiments.

the maximal number of matches that can be identified. Note that in the shifted and warped image test, as the images are deformed, the number of the mass-like objects that will attend the graph matching process may not be the same for the original images and the deformed images. For example, for a mammogram M_1, after the filtering, there may be only 15 mass-like objects left for the graph matching, and for its deformed version M_2, there may be 13 mass-like objects left. Even the leftover mass-like objects in M_2 may not correspond mass-like objects in M_1 since, for example, a mass-like object may lie in the strip of the image that is removed during the shifting process. Hence the total true matches for this pair may be less than 13. Correct matches found is the number of matches that were identified by the graph matching process and that are true matches. The correct match rate (cmr) is defined as cmr = correct matches found / total matches found. The match efficiency (me) is calculated as

$$me = \frac{\text{correct matches found}}{\text{total true matches}} - \frac{\text{total matches found - correct matches found}}{\text{total true matches}}.$$

8.1.2   Relabelled Images

In this experiment, the second image is the same as the original mammogram. To segment the second image, for each label image of the original mammogram, a second label image was obtained by randomly re-assigning the labels in the original label image. Thus the segmentation is identical for the two images except that labels of the segments do not match. This simple experiment checks the performance of the matching process without the influence of the segmentation process. Table 8.1 shows the results of matching on relabelled images. The table shows that the registration method obtains perfect results under the most ideal circumstances with 100% of correct match rate and match efficiency in all 18 images. An example is shown in Figure 8.2.
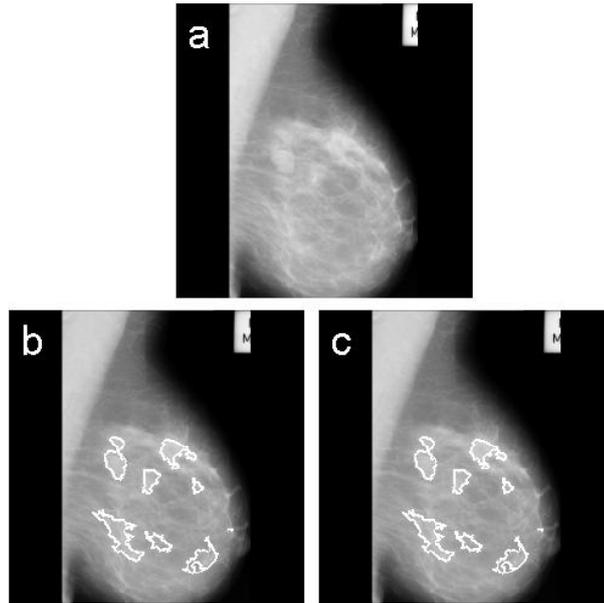
Figure 8.2: Example of matching results for a relabelled image. (a) is the original mdb019. Eight matched mass-like objects are marked out with white boundary in (b) and (c). Mass-like objects in (b) are from the original labelled image and mass-like objects in (c) are from the relabelled image.

8.1.3 Relabelled and Shifted Images

In this experiment, to create the segmentation of the shifted images, the original label images are first relabelled, and then shifted. The same shifting method as described in Chapter 3, Section 3.4.1 is used to create both the second image and its label image. Breast boundaries of second images are obtained by shifting the boundaries of the original images in the same way. Table 8.2 lists the results of the matching on the original images and the relabelled and shifted images. The proposed method reached almost perfect results on all 18 images, with 100% of correct match rate on all 18 images and 100% match efficiency on most of images. Figure 8.3 shows an example of matching results.

mammogram	total matches found	total true matches	correct matches found	cmr	me
mdb001	20	20	20	100%	100%
mdb002	13	13	13	100%	100%
mdb003	9	9	9	100%	100%
mdb004	13	13	13	100%	100%
mdb005	20	20	20	100%	100%
mdb006	14	14	14	100%	100%
mdb009	11	11	11	100%	100%
mdb010	14	14	14	100%	100%
mdb011	17	17	17	100%	100%
mdb012	17	17	17	100%	100%
mdb013	17	17	17	100%	100%
mdb014	17	17	17	100%	100%
mdb015	19	19	19	100%	100%
mdb016	17	17	17	100%	100%
mdb017	10	10	10	100%	100%
mdb018	5	5	5	100%	100%
mdb019	20	20	20	100%	100%
mdb020	18	18	18	100%	100%

Table 8.1. Matching results for relabelled images.

mammogram	total matches found	total true matches	correct matches found	cmr	me
mdb001	20	20	20	100%	100%
mdb002	12	12	12	100%	100%
mdb003	9	9	9	100%	100%
mdb004	10	11	10	100%	90.1%
mdb005	20	20	20	100%	100%
mdb006	12	12	12	100%	100%
mdb009	9	10	9	100%	90%
mdb010	13	13	13	100%	100%
mdb011	16	17	16	100%	94.1%
mdb012	17	17	17	100%	100%
mdb013	16	17	16	100%	94.1%
mdb014	16	16	16	100%	100%
mdb015	16	16	16	100%	100%
mdb016	16	16	16	100%	100%
mdb017	9	9	9	100%	100%
mdb018	4	4	4	100%	100%
mdb019	18	18	18	100%	100%
mdb020	17	17	17	100%	100%

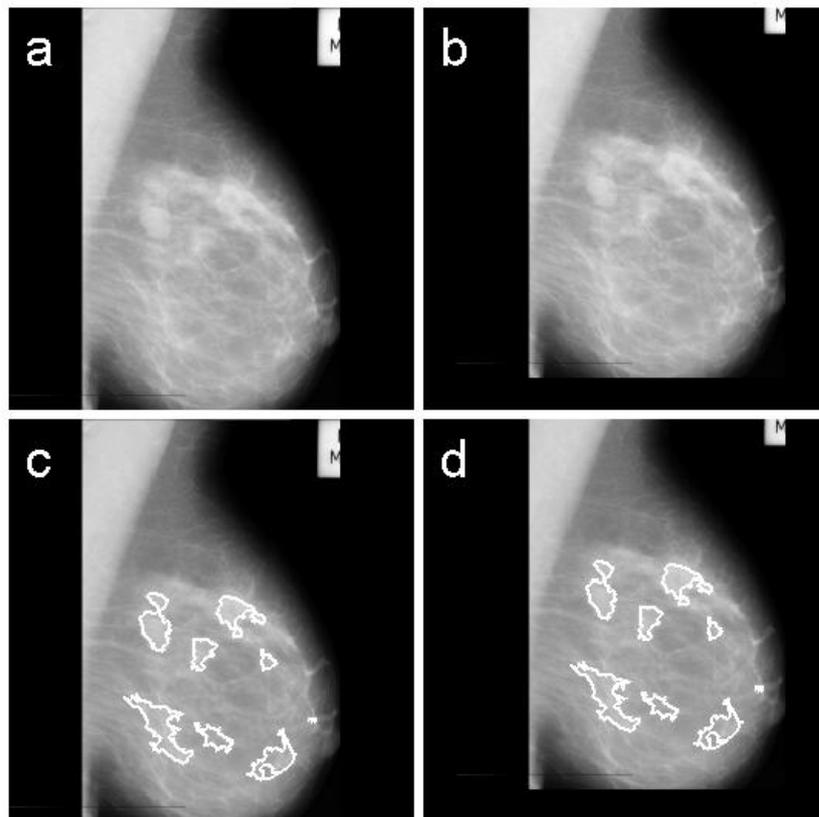**Table 8.2. Matching results for relabelled and shifted images.**

Figure 8.3: Example of matching results for a relabelled and shifted image. (a) is the original mdb019. (b) is the shifted version of mdb019. Eight matched mass-like objects are marked out with white boundary in (c) and (d). Mass-like objects in (c) are from the original labelled image and mass-like objects in (d) are from the relabelled and shifted image.

8.1.4 Relabelled and Warped Images

For this test, the proposed matching method is applied to the original images
and their warped images. Segmentation of the warped images is obtained by
first relabelling, and then warping the label images of the original mammograms.
The same warping method as used in Chapter 3, Section 3.4.1 is used for this
experiment. In this experiment, the breast boundaries of the warped images are
obtained by using the AP segmentation based initial breast boundary extraction
method (see Section 5.1) on the warped images. Results on all 18 images are
presented in Table 8.3. The matching failed dramatically in one case (mdb017).
The correct match rate is high in most of the images with 100% in 14 images.
However, match efficiency varies between images. The failed case (mdb017) is
shown in Figure 8.4 and Figure 8.5 shows a more typical example.

mammogram	total matches found	total true matches	correct matches found	cmr	me
mdb001	13	13	13	100%	100%
mdb002	6	11	4	66.7%	18.1%
mdb003	7	7	7	100%	100%
mdb004	11	13	11	100%	84.6%
mdb005	16	17	16	100%	94.1%
mdb006	14	14	14	100%	100%
mdb009	4	4	4	100%	100%
mdb010	8	13	7	87.5%	46.1%
mdb011	8	9	7	87.5%	66.7%
mdb012	9	16	9	100%	56.2%
mdb013	12	12	12	100%	100%
mdb014	13	16	13	100%	81.3%
mdb015	8	11	8	100%	72.7%
mdb016	14	17	14	100%	82.4%
mdb017	3	3	0	0%	-100%
mdb018	3	5	3	100%	60%
mdb019	16	16	16	100%	100%
mdb020	16	18	16	100%	88.9%

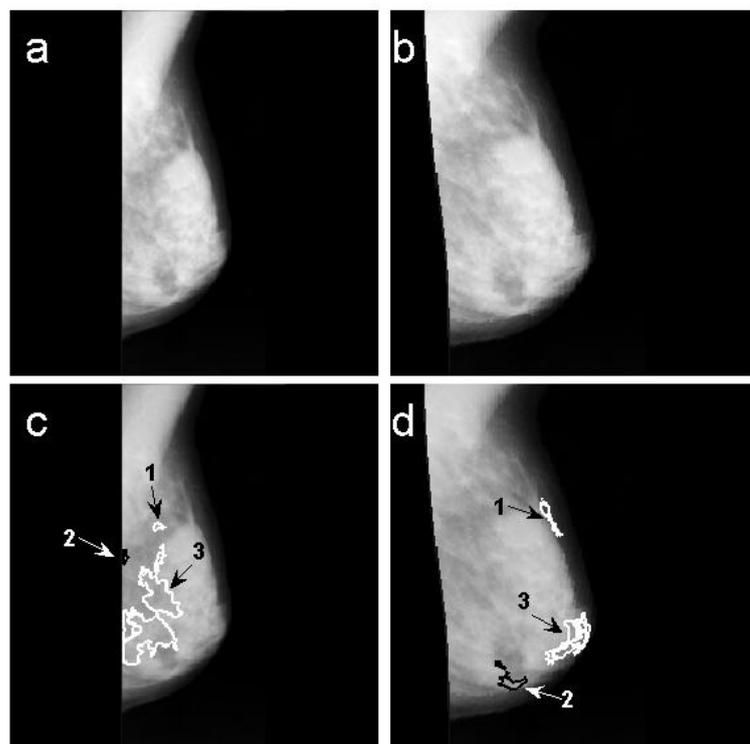**Table 8.3. Matching results for relabelled and warped images.**

Figure 8.4: Example of matching results of mdb017 and its relabelled and warped image. (a) is the original mdb017. (b) is the warped image of mdb017. Three matched mass-like objects are marked out with boundaries (both in black and white) in (c) and (d). Mass-like objects pointed out by the same number in (c) and (d) are the matched mass-like objects. Mass-like objects in (c) are from the original image and mass-like objects in (d) are from the relabelled and warped image.
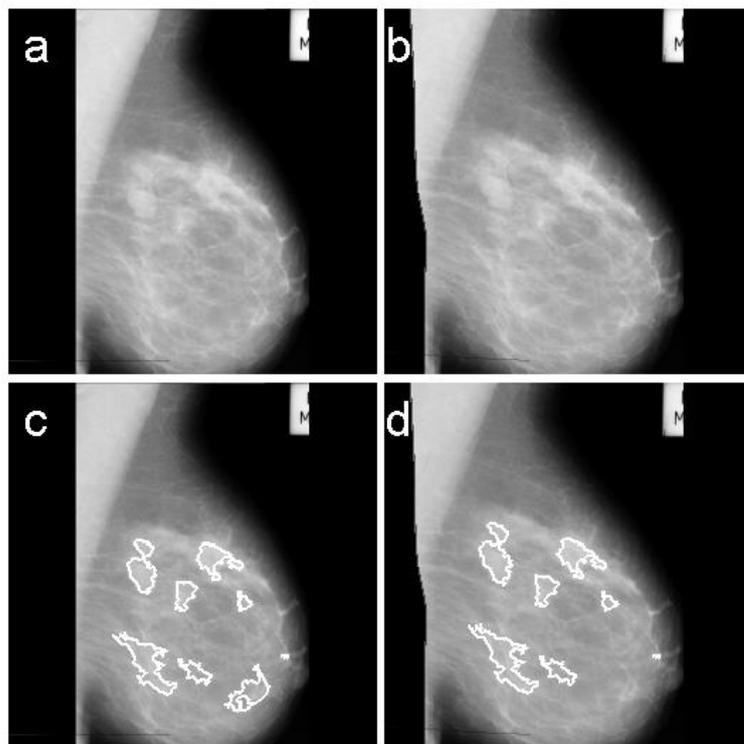
Figure 8.5: Example of matching results for a relabelled and warped image. (a) is the original mdb019. (b) is the warped image of mdb019. Eight matched mass-like objects are marked out with white boundary in (c) and (d). Mass-like objects in (c) are from the original image and mass-like objects in (d) are from the relabelled and warped image.

8.2 Experiments with Real Temporal Mammograms

Although artificial data is convenient in evaluation of complex registration problems, Marias et al. (2005) pointed out that the artificial data can not mimic the actual changes that happened on the breast. Further more, the registered pairs are not generally expected to be identical as breast structure may change significantly over time.

In this section, the real temporal mammograms are used to evaluate the performance of the proposed registration method. The difficulty with real temporal data is that the true correspondences between the mass-like objects of temporal mammograms are never known with absolute certainty. As breast structure can be significantly different between consecutive examination, and due to the compression and twist that happens during the image acquisition, correspondence between the mass-like objects of temporal mammograms may not exist at all. For this reason, as also pointed out by Marias et al. (2005), the registration method needs to be evaluated both with quantitative measures and by the judgment of experts. In this section, the matching results are evaluated in two ways. First, results are visually assessed and classified into "good", "average", "poor" and "unknown". Second, the match results are applied directly to the mass detection to reduce the false detection rate of an existing mass detection scheme.

8.2.1 Dataset

The dataset used for the experiments is composed of 95 temporal mammogram pairs. Each pair consists of a current mammogram that was called back at screening and a previous same view mammogram taken for the same breast. The mammograms were randomly selected from a local screening archives from the same time period under the restriction that the woman was recalled based on the current mammogram and that a previous mammogram (less than 3 years old), that had not resulted in recall, was available.

Images were digitized using a Vidar Diagnostic Pro Advantage digitizer (48 μm spatial resolution and 12 bit depth). All images are originally 5296×3478 pixels in size. For processing, images were downsampled by a factor of $8 \times 8 \rightarrow 1$.

The mammogram pairs were divided into a training dataset and a testing dataset. Table 8.4 summarizes the information of each dataset. The training dataset contains 51 temporal mammogram pairs, 25 of which are CC views and 26 are MLO views. This set contains 21 malignant masses (confirmed by histopathology). The testing set contains 44 temporal mammogram pairs, with 19 CC view mammogram pairs, 25 MLO view mammogram pairs and 23 malignant masses. This set became available later than the training dataset and thus is used as testing dataset.

	training dataset	testing dataset
no. of image pairs	51	44
no. of malignant images	21	23
MLO views	26	25
CC views	25	19

Table 8.4: Composition of the datasets from a local database used for the experiments.

All malignant images in both datasets were annotated by an experienced radiologist. For each malignant mass, a box was drawn by the radiologist to include the whole mass. The coordinates of four vertices of the box were stored in a file.

8.2.2   Feature Filter and Mass-like Score

Parameters involved in the feature filters and mass-like score (see Section 6.4 and Section 6.5) were selected based on the training dataset. For this purpose, for each malignant mass in the training dataset, based on the radiologist marked box, we manually pick up the segmented component that best covers the mass (Fig. 8.6). Features were then calculated for these manually picked components. Parameters used in the feature filters were set as the minima or maxima of the corresponding features measured on the manually recorded masses. In this way, the filters did not remove the malignant masses.

Based on the training dataset, the filters were set as follows.

- $-0.58 < \text{area} < 1.6$

- $-2.38 < \text{s1} < 1.78$

- $-1.1 < \text{ratio} < 1.28$

- $-1.63 < \text{s2} < 2.24$

To set the coefficients for the mass-like score calculation, the segmented components of each mammogram were filtered by both the template filter and the feature filters. Features listed in Section 6.4 were calculated on all remaining components. Next, the remaining components of all mammograms in the training dataset were put together and divided into two groups, those that were manually picked up as malignant masses forming one group, all other components that are not associated with malignant masses forming the other group. Figure 6.5 shows the distribution of these two groups. Coefficients used to compute the mass-like
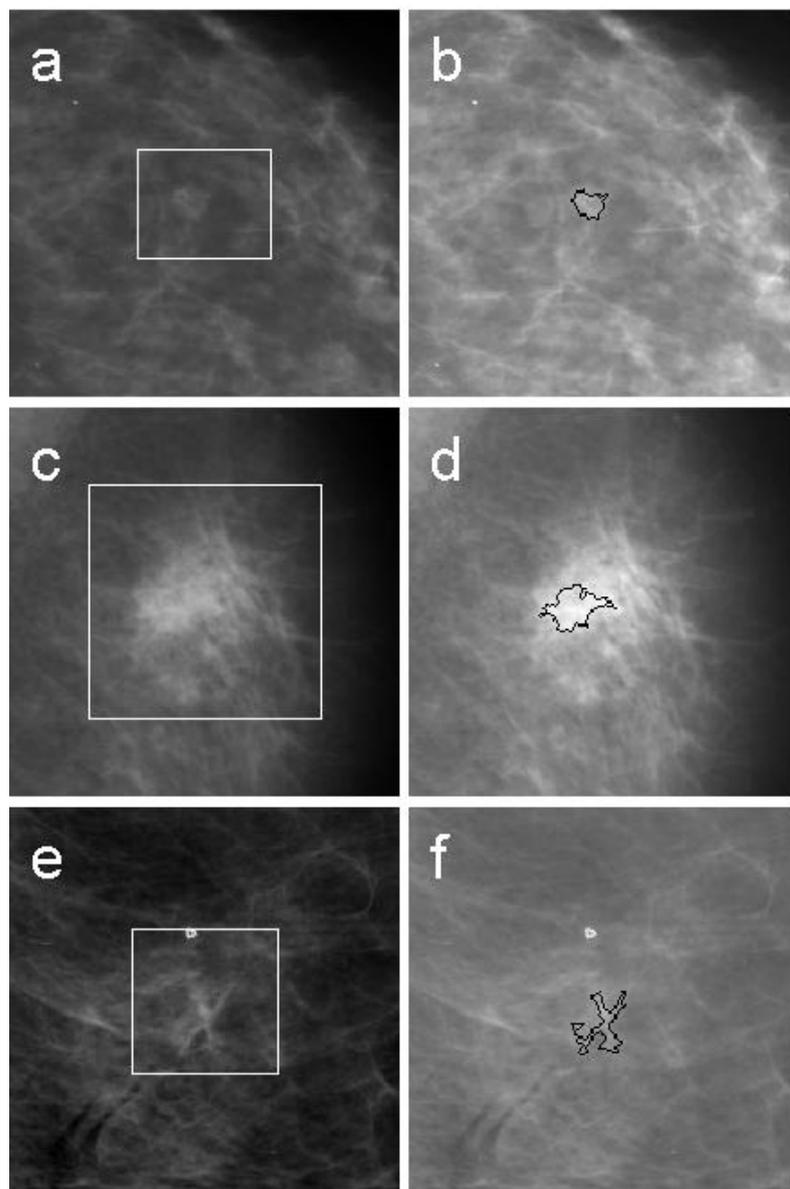
Figure 8.6: Examples of malignant mass components. The images in the left column show three malignant masses enclosed in boxes marked by the radiologist. The components that best cover the masses are shown in the corresponding right column with the boundaries of the components marked out in black line.

score were optimized using Fisher LDA on the two groups. In this thesis, the coefficients were

$$
\begin{aligned}
\text{mass-like score} = {} & -0.10820143991282 \times solidity + (-0.00251710862324) \times ratio \\
& + (-0.12595050961751) \times areas + (-0.53221670446709) \times dradi \\
& + 0.12273889473518 \times pri + 0.79985455221125 \times radi \\
& + (-0.10625337678614) \times circularity + 0.13674010732452 \times c2 \\
& + 0.06594075503656 \times c3
\end{aligned}
\tag{8.2}
$$

8.2.3 Evaluation Based on Visual Perception

The process of visual perception evaluation is as follows. First, the examiner identified several correspondences between the mammograms. Distinct objects (landmarks), such as line structures or clear masses, were picked up to establish the correspondence. Figure 8.7 shows two examples with corresponding objects marked. The matched mass-like objects were then compared, based on those identified correspondence. A match was "good" if according to the reference landmarks, they were exactly the same mass-like object. If the mass-like objects were not exactly the same, but were very close by, they were identified as "average". "poor" matches refer to those that were clearly different mass-like objects, and also not close by mass-like objects. In some cases, even with clearly identified corresponding landmarks, it was still difficult to identify a match as corresponding well or not. These matches were marked as "unknown". Figures 8.8 - 8.11 give examples of each type.

An average of 13.2 matches were made over all the mammogram pairs from both data sets. 63.5% of matches were identified as "good", and 23.6% as "average". The percentages of "poor" and "unknown" were 10.9% and 2% respectively. An example of matching appears in Figure 8.12.
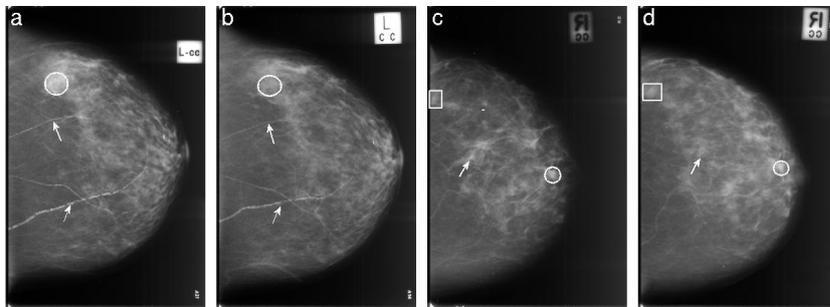

Figure 8.7: Examples of manual matching. In mammogram pair (a) and (b), three correspondences, two elongated structures (arrows) and one bright region (circle) are marked out. In (c) and (d), 3 corresponding bright regions are identified in the pair.
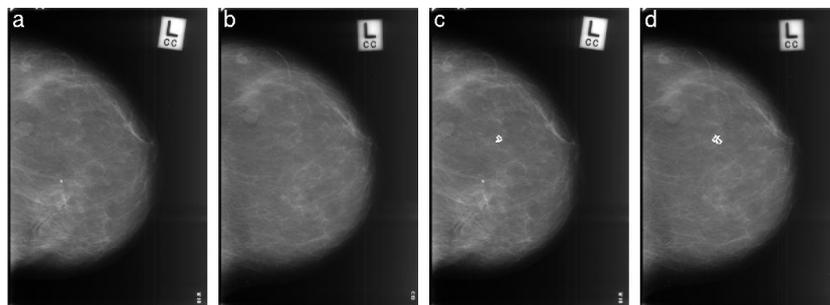
Figure 8.8: Example of a good match. (a)(b) are paired mammograms and (c)(d) are the same as (a)(b) with one matched mass-like object marked out by a white border. This match was identified as a "good" match.
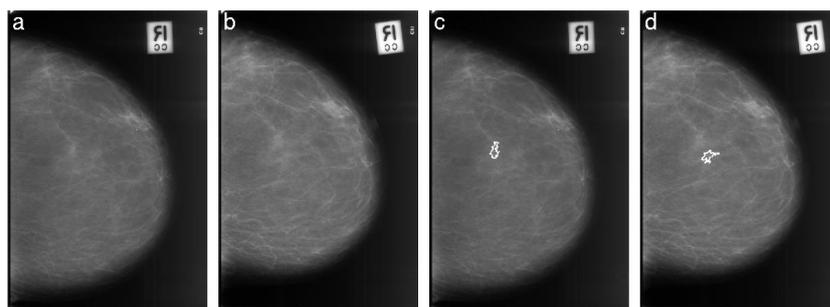

Figure 8.9: Example of an average match. (a)(b) are paired mammograms and (c)(d) are the same as (a)(b) with one matched mass-like object marked out by a white border. This match was identified as an "average" match.
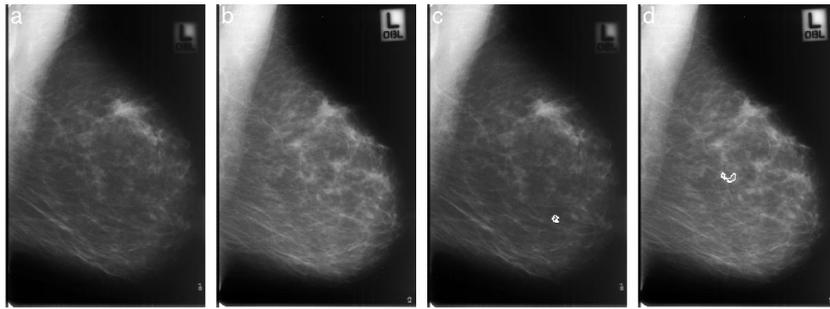
Figure 8.10: Example of a poor match. (a)(b) are paired mammograms and (c)(d) are the same as (a)(b) with one matched mass-like object marked out by a white border. The match in this example is identified as "poor".
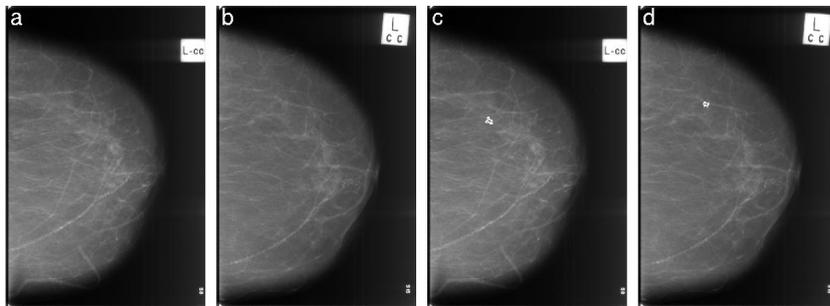

Figure 8.11: Example of an unknown match. (a)(b) are paired mammograms and (c)(d) are the same as (a)(b) with one matched mass-like objects marked out by a white border. Although the matched in (c) and (d) are in relatively similar position, the nearby structures, especially the elongated structures close to the marked mass-like objects are different. This match was identified as "unknown".

8.2.4 Evaluation Based on False Positive Detection Reduction

In mammography, a new born malignant mass means that no malignant mass was found in the prior examination but it exists in current mammogram. Detection of new born malignant masses is one of the most important task in mammography. In this experiment, the testing images were selected so that masses were not found in the previous mammogram but found in the current mammogram (Section 8.2.1). These images imitate the most common scenario in mammogram interpretation: radiologists compare the current mammograms with the previous ones, which do not have malignant masses detected, to see if there are new born malignant masses. With these images, we assume that the malignant masses in current mammograms do not have matches in the prior mammograms.
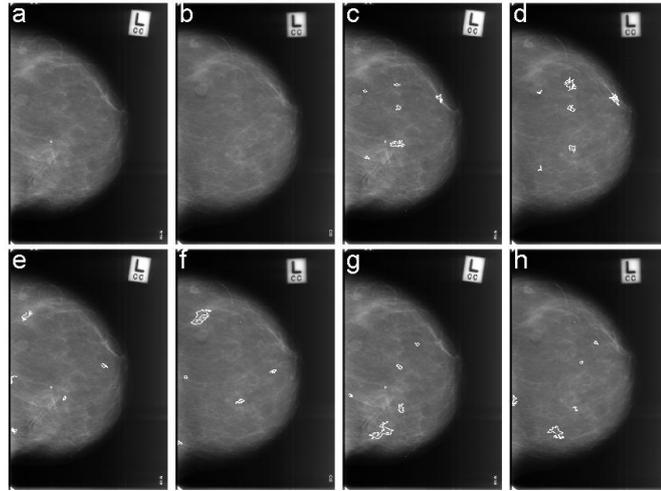
Figure 8.12: Example of matching results. (a) and (b) are the mammogram pair. Matched mass-like objects of mammogram (a) are displayed in (c)(e)(g) with mass-like objects marked out by a white boundary. (d)(f)(h) are the same as (b) with matched mass-like objects inclosed by a white line. Matched mass-like objects are in similar location in (c) and (d), (e) and (f), (g) and (h).
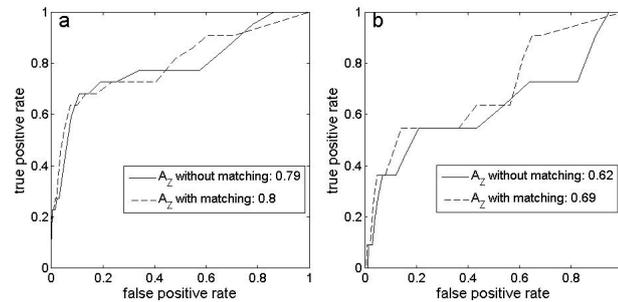

Figure 8.13: ROC curves with and without graph matching. (a) training data, (b) test data.

In this section, the results of proposed registration method was used to improve the performance of new born malignant mass detection by reducing the false positive detections. The process of using the AP algorithm to segment a mammogram, extracting features for segmented components, assigning a mass-like score to each component and using this score to classify each component as malignant or not, constitutes a mass detection scheme. We refer to this scheme as mass detection without matching. Under our assumption, mass-like objects that are associated with true malignant masses do not have corresponding components in the previous mammogram, thus a mass-like object in current mammogram that was matched to a mass-like object in the previous mammogram is not likely to be a true mass. The match results are used to reduce the false positive detections

by rejecting the candidate masses detected without matching if the associated mass-like object is matched to a mass-like object in the previous mammogram. The resulting detection scheme will be referred to as the detection with matching.

To test the contribution of the matching method, separate ROC curves were constructed for detection without matching and for detection with matching for both the training set and the testing set (Figure 8.13). For the training set, the A_z score (area under the ROC curve) with graph matching was 0.80 while A_z was 0.79 without graph matching. For the testing set, A_z was 0.69 with graph matching and 0.62 without.

8.3 Discussion

cmr measures how correct the registration method is in corresponding the mass-like objects of multiple mammograms of the same breast. *me* measures how efficient the method is in identifying the corresponding mass-like objects. False matches falsely bind up the mass-like objects belonging to different parts of breast and thus prejudice the registration and result in false information. On the other hand, very few matches can not register the multiple mammograms well. However, once there are enough matches to establish good correspondence, the total number of matches is not important. This is the reason why *cmr* is defined in terms of the total matches found instead of the total true matches. The *cmr* values for the experiments with relabelled and warped images (Section 8.1.4) indicate that mostly correct matches are found. This leaves the question as to whether sufficient matches are found to establish a useful correspondence. This cannot be judged by experiments with constructed mammograms as in sections 8.1 but can only be judged within a real application. The results of section 8.2 indicate that, at least in this application, the number of correspondences is sufficient.

Positioning of the breast is an important factor to many mammogram registration methods. In fact, correct positioning of the breast is a prerequisite for a nipple location based registration method (Engeland et al. Nov. 2003). Differently positioned mammogram pairs have different profile of the breasts and so may cause difficulties to mammogram registration methods based on the alignment of breast boundaries. As mainly based on the internal structure of the breast, the proposed method can cope with differently positioned mammogram pairs and mammograms with varying sizes. Figures 8.14 and 8.15 show examples of these situations.

For two cases in the training set, and one in the testing set, the segmentation step identified a bright region in the previous mammogram that was similar in appearance and location to the true mass in the current mammogram (Fig. 8.16, 8.17 and 8.18). Since a matching mass was found in the previous image, the true mass was rejected as a false positive. However, since a mass-like object was detected in the previous image, it is possible that the mass would have been
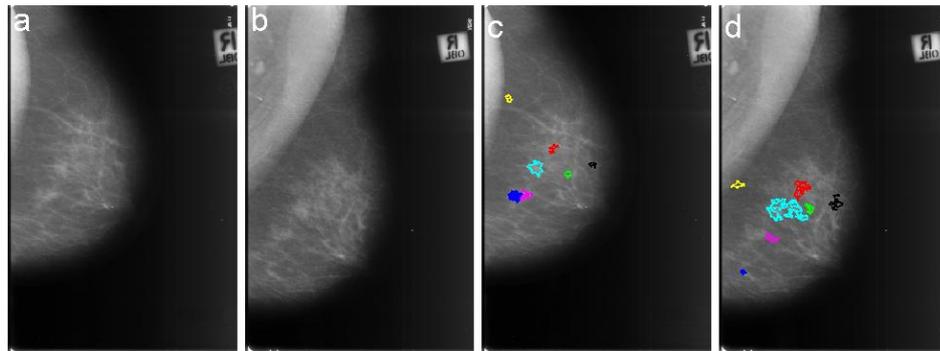
Figure 8.14: Example of differently positioned breasts. (a) and (b) are the paired mammogram. (c) and (d) are the same as (a) and (b) except that the matched mass-like objects are marked out by colored boundary with the matched mass-like objects in same color. The breasts in (a) and (b) are in totally different size, the size of pectoral muscles are different. Several good correspondence were found in this case.
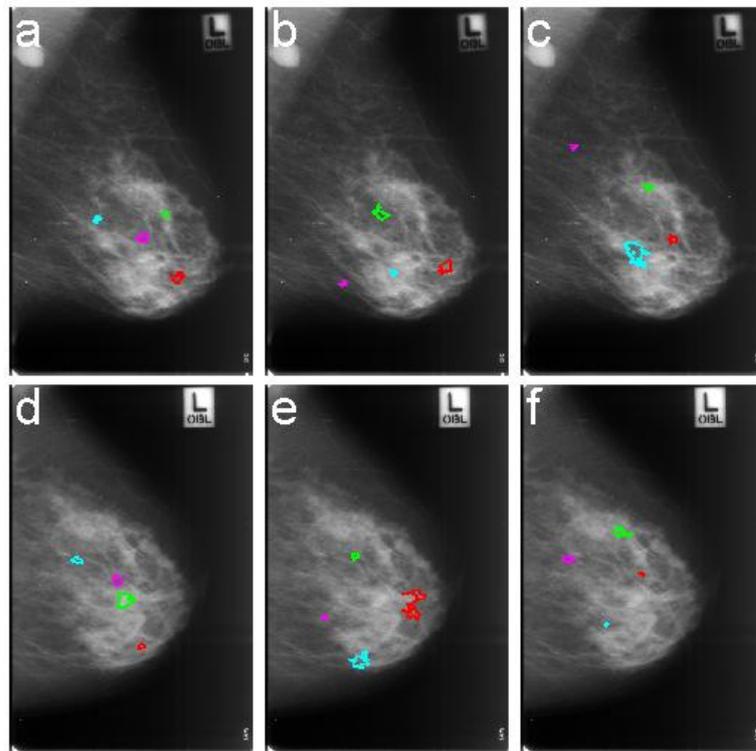

Figure 8.15: Another example of differently positioned breasts. (a)(b)(c) are the same mammogram and (d)(e)(f) are the same paired mammogram of (a)(b)(c). The same color line enclosed mass-like objects in (a) and (d), (b) and (e), and (c) and (f) are the matched mass-like objects.
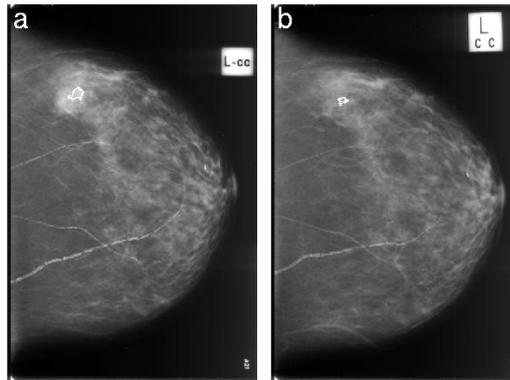
Figure 8.16: Example of a falsely rejected true mass in the training set. A component in the current mammogram (left) associated to a true mass was matched to a component in the previous mammogram (right) and so was rejected as a true mass.

detected at the previous round if graph matching had been used to compare the image from the screening round before the previous round. Such an image was not available for this study and so this could not be tested. Also, it is not possible to know if the bright region found in the previous image was indeed an early stage of the true cancer found in the current image or a bright region unrelated to cancer.

In this study, the performance was measured against the performance of the without-matching scheme. However, the without-matching scheme was not designed for mass detection on its own. A better test would be to compare results with state-of-the-art mass detection methods on a common set of images. Unfortunately, implementations of such methods were not available. Accordingly, a good estimate of the contribution of graph matching to the reduction of false positive detections is not yet known.
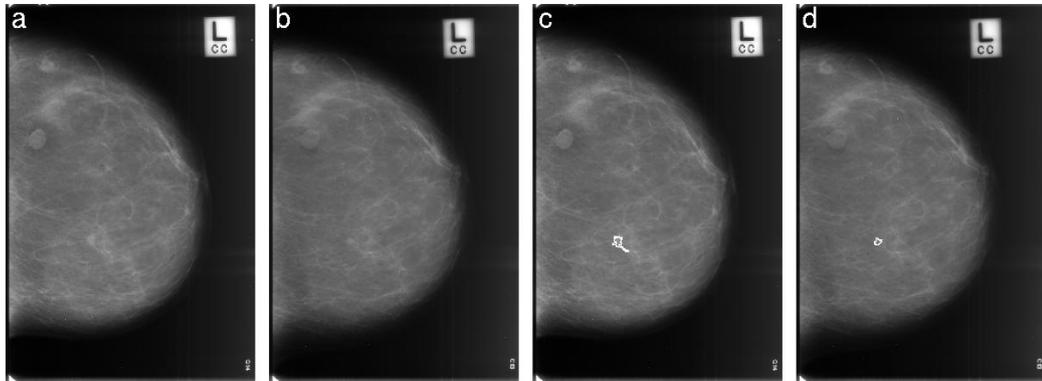
Figure 8.17: False rejection of a true mass in the training set. (a) current mammogram. (b) previous mammogram. (c) Same as (a) but with the boundary of a true mass marked according to the AP segmentation results. (d) Same as (b) but with the boundary marked according to the AP segmentation results. Since the mass was detected in the previous image, the true mass was rejected as a false positive in the current screening round.
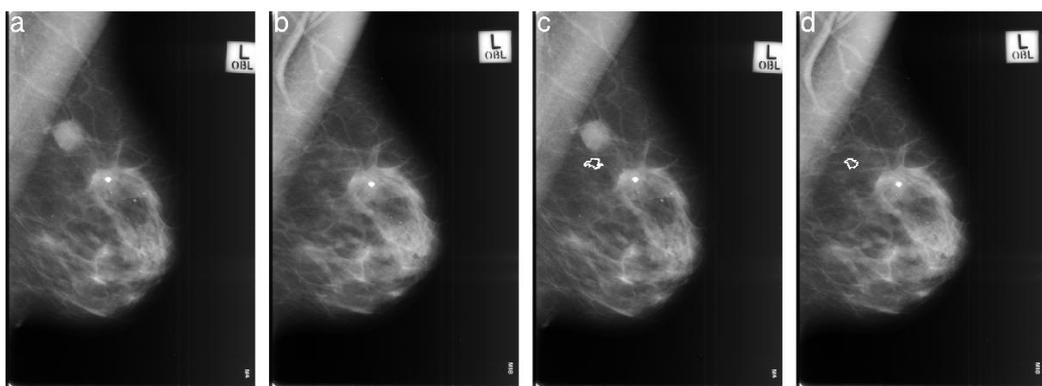

Figure 8.18: False rejection of a true mass in the testing set. A malignant mass in current mammogram (a)(c) was matched to a non-mass component of previous mammogram (b)(d).

Chapter 9

Final Remarks and Conclusion

This chapter consists of final remarks in Section 9.1 and the conclusion of the thesis in Section 9.2.

9.1 Final Remarks

In the proposed method, the breast boundary plays an important role, both during the graph matching process and in the final filter step. For this reason, a substantial effort was made to develop a very reliable method for finding the breast boundary (Section 5.2). The breast boundary as an entire entity provides a global spatial reference. Since the entire boundary is viewed as a single object, the extraction of this global reference is much more robust than individual landmark points. However, because the breast boundary plays such an important role, when the boundary is not determined correctly, graph matching results are likely to be very poor. Figure 9.1 illustrates such an example. This example is an extreme case and failed because the natural boundary of the breast lies, in part, outside the mammogram. For images in which the entire boundary appears on the mammogram the methods developed here work extremely well. For a full clinical application of graph matching to false positive mass reduction or to mass detection, a method must be found to treat cases in which the mass boundary extends outside the mammogram.

In this thesis, the effectiveness of the AP segmentation was only roughly estimated based on the number of successfully isolated true masses in the training set and the AP segmentation method does not find boundaries of masses that necessarily agree well with the true boundary (Section 3.4.3). The results of the AP segmentation was used as the input to the graph matching process without more careful effectiveness evaluation. This was done deliberately for three reasons. (i) If a method relies heavily on very accurate boundaries, this provides another opportunity for the method to fail at an early stage. (ii) Since changes in the boundary are expected for the same object in temporal images, the details
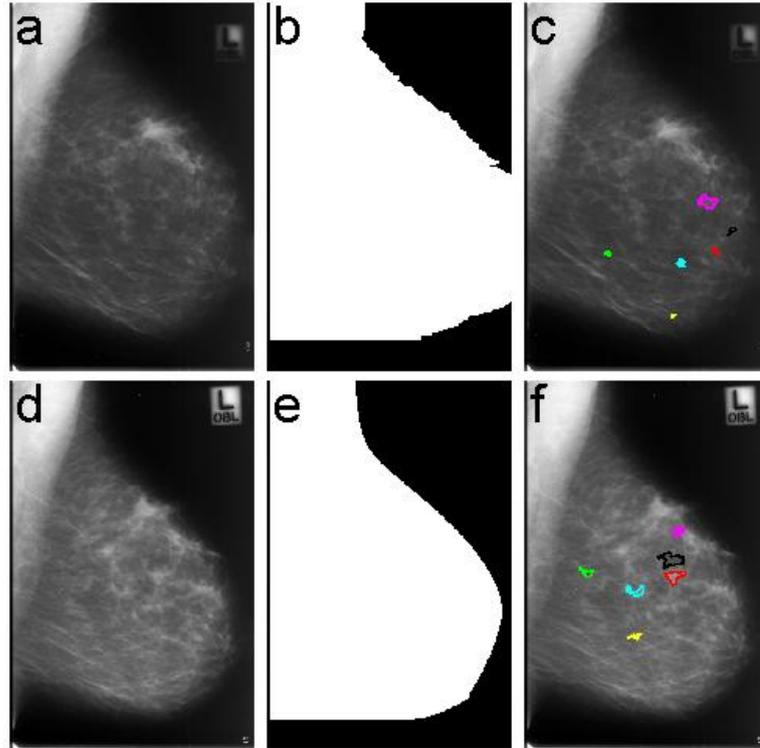
Figure 9.1: Example of poor boundary resulting in false matches. (a) and (d) are the paired mammograms. (b) and (e) are the templates of (a) and (d) with breast area in white and background in black. (c) and (f) display several selected matched mass-like objects with each matched mass-like objects in same color. Many false matches were formed in this case because of the unprecise boundaries.

of boundaries may confuse rather than enhance matching. (iii) State of the art mass detection algorithms, though reported in the literature to some extent, are not readily available for implementation. However, it may be possible to improve results by employing a very good mass segmentation scheme. Although the same mass may have different boundary details between images, these differences could be absorbed in the matching criteria. A careful study on a large data base could determine the nature of changes in boundary details expected of malignant masses compared to benign masses or non-mass components.

Although the experimental results show improvements on new born malignant mass detection by reducing the false positive rate (see Section 8.2.4) and capability in corresponding mass-like objects in temporal mammograms (see Section 8.1 and 8.2.3), the proposed method is not ready for clinical use. A clinical ready program should take care of all possible situations, such as breasts with boundaries that spill outside the mammogram (Fig. 9.1). Another reason for the proposed method not being clinical ready is, the without-matching scheme used in the experiments (see Section 8.3) on its own was not designed for mass

detection.

In this thesis, the proposed method was applied to both MLO view and CC view mammogram pairs without any modification. As no particular restrictions are applied, the method can be applied to other type multiple mammograms (MLO-CC, bilateral mammograms). However, in MLO-CC mammogram pairs, the mass-like objects may not have compatible spatial relations between them. As the proposed method is based on the spatial relations between the mass-like objects, the proposed registration method may produce poor results in MLO-CC view mammogram pairs.

9.2 Conclusion

In this thesis, we have presented a temporal mammogram registration method. In this method, spatial relationships between the mass-like objects were used to establish the correspondence between mass-like objects of temporal mammograms. The establishment of the correspondence is realized by using the graph matching to find a common subgraph between two graphs representing two mammograms of a mammogram pair.

We have tested the proposed method on both artificial data and real temporal mammograms. Artificial data were made by relabelling, shifting and warping the label images, which were resulted from the segmentation, for the purpose of investigating the performance of the method without the affection of poor segmentation. Experiments on artificial data provided initial evaluation of the proposed method. The results of the proposed registration method in relabelled and shifted image test were perfect. On the warped image test, a mean *cmr* of 91% was realized (when a failed cased was excluded, the mean *cmr* was 97%).

For the real temporal mammogram test, 95 mammogram pairs were selected out from a local archive. The selected mammogram pairs were divided into a training dataset and a testing set. The proposed method was trained using the training dataset and was then applied to the testing dataset. The results were evaluated based on both visual perception and false positive reduction. Based on visual perception, an average of 63.5% of matches were identified as "good", 23.6% of them were "average". The results were "poor" on 10.9% of the matches and "unknown" on 2% of the matches. For the quantitative assessment, the matching results were used to reduce the false detection rate. On both sets, A_z was improved with graph matching. For the training set, the A_z score improved from 0.79 to 0.80 with graph matching. For the testing set, A_z rise from 0.62 to 0.69 with graph matching.

The proposed registration method solves an important problem in the registration of mammograms. The method inherently allows for the natural changes in breast over time and the differences due to image acquisition. However, the

method as presented here is not ready for incorporation into existing hardware for computer-aided mammography (Section 9.1). However, as also described in section 9.1, the method does allow for natural extensions that could provide significant improvement in the use of multiple mammogram in the early detection of breast cancer.

Bibliography

Abdel-Mottaleb, M., Carman, C., Hill, C. & Vafai, S. (1996), Locating the boundary between the breast skin edge and the background in digitized mammograms, *in* 'Proceedings of the 3rd international workshop on digital mammography', Chicago IL, pp. 467–470.

Abdulkader, A. M. (1998), *Parallel Algorithms for Labelled Graph Matching*, PhD thesis, Colorado School of Mines.

Ackerman, L. V. & Gose, E. E. (1972), 'Breast lesion classification by computer and xeroradiography', *Cancer* **30**, 1025–1035.

AIHW (Australian Institute of Health and Welfare) and AACR (Australasian Association of Cancer Registries) (2007), *Cancer in Australia: an overview, 2006*, Cancer series no. 37. Cat. no. CAN 32., Canberra: AIHW.

American Cancer Society (2007), *Breast Cancer Facts & Figures 2007-2008*, Atlanta: American Cancer Society, Inc.

American Cancer Society (2008), *Cancer Facts & Figures 2008*, Atlanta: American Cancer Society.

Auscher, P. (1992), 'Wavelet bases for $L_2(R)$ with rational dilation factor', *Wavelets and their applications, Boston Jones and Barlett* pp. 439–451.

Australian Institute of Health and Welfare & National Breast Cancer Centre (2006), *Breast cancer in Australia: an overview, 2006*, Cancer series no. 34. cat. no. CAN 29., Canberra: AIHW.

Australian Institute of Health and Welfare 2008 (2007), *BreastScreen Australia Monitoring Report 2004-2005*, Cancer series no. 42. Cat. no. CAN 37., Canberra: AIHW.

Bajger, M., Ma, F. & Bottema, M. J. (2005), Minimum spanning trees and active contours for identification of the pectoral muscle in screening mammograms, *in* B. C. Lovell, A. J. Maeder, T. Caelli & S. Oursellin, eds, 'Digital Image Computing Techniques and Applications', IEEE Computer Society Conference Publishing Service, Cairns, Qld Australia.

Baussard, A., Nicolier, F. & Trucheter, F. (2004), 'Rational multiresolution analysis and fast wavelet transform: application to wavelet shrinkage denoising', *Signal Processing 84 ELSEVIER* pp. 1735–1747.

Bennett, R. L., Blanks, R. G. & Moss, S. M. (2006), 'Does the accuracy of single reading with CAD (computer-aided detection) compare with that of double reading?: A review of the literature', *Clinical Radiology* **61**, 1023–1028.

Besag, J. (1986), 'On the statistical analysis of dirty pictures', *J. Roy. Stat. Soc.* **48**, 259 – 302.

Bick, U., Giger, M., Schmidt, R., Nishikawa, R., Wolverton, D. & Doi, K. (1995), 'Automated segmentation of digitized mammograms', *Acad Radiol* **2**(1), 1–9.

Bister, M., Cornelis, J. & Rosenfeld, A. (1990), 'A critical view of pyramid segmentation algorithms', *Pattern Recognition Letters* **11**(9), 605–617.

Bloch, I. (1999), 'Fuzzy relative position between objects in image processing: a morphological approach', *IEEE Trans. on Pattern Analysis and Machine Intelligence* **21**(7), 657–664.

Bookstein, F. L. (1989), 'Principal warps: thin-plate splines and the decomposition of deformations', *IEEE Trans. on Pattern Analysis and Machine Intelligence* **11**(6), 567–585.

Brem, R. F., Baum, J., Lechner, M., Kaplan, S., Souders, S., Naul, L. G. & Hoffmeister, J. (2003), 'Improvement in sensitivity of screening mammography with computer-aided detection: a multiinstitutional trial', *AJR* **181**, 687–693.

Brem, R. F., Hoffmeister, J. W., Zisman, G., DeSimio, M. P. & Rogers, S. K. (2005), 'A computer-aided detection system for the evaluation of breast cancer by mammographic appearance and lesion size', *AJR* **184**, 893–896.

Brun, L. & Kropatsch, W. G. (2003), Construction of combinatorial pyramids, *in* E. Hancock & M. Vent., eds, 'Graph Based Representations in Pattern Recognition, Lecture Notes in Computer Science', Vol. 2726, Springer, Berlin, pp. 1–12.

Brzakovic, D., M., L. X. & Brzakovic, P. (1990), 'An approach to automated detection of tumors in mammograms', *IEEE Trans. Med. Imaging* **9**, 233–241.

Bunke, H. (2000), 'Recent developments in graph matching', *Proc. of ICPR'00* **2**, 117–124.

Caulkin, S., Astley, S., Asquith, J. & Boggis, C. (1998), Sites of occurence of malignancies in mammograms, *in* 'Proceedings of the 4th International Workshop on Digital Mammography', IWDM 1998, Nijmegen, The Netherlands, pp. 279–282.

Chan, H.-P., Wei, D., Helvie, M. A., Sahiner, B., Adler, D. D., Goodsitt, M. M. & Petrick, N. (1995), 'Computer-aided classification of mammographic masses and normal tissue: Linear discriminant analysis in texture feature space', *Phys. Med. Biol.* **40**, 857–876.

Chandrasekhar, R. & Attikiouzel, Y. (1997), 'A simple method for automatically locating the nipple on mammograms', *IEEE Trans. Med. Imaging* **16**(5), 483–494.

Conte, D., Foggia, P., Sansone, C. & Vento, M. (2004), 'Thirty years of graph matching in pattern recognition', *Internat. J. Pattern Recognit. Artif. Intell.* **18**(3), 265–298.

Dubes, R. C. & Jain, A. K. (1989), 'Random field models in image analysis', *J. Applied Statistics* **16**, 131 – 164.

Engeland, S. V., Snoeren, P., Hendriks, J. & Karssemeijer, N. (Nov. 2003), 'A comparison of methods for mammogram registration', *IEEE Trans. Med. Imaging* **22**(11), 1436–1444.

Ferlay, J., Autier, P., Boniol, M., Heanue, M., Colombet, M. & Boyle, P. (2007), 'Estimates of the cancer incidence and mortality in europe in 2006', *Ann Oncol.* **18**(3), 581–592.

Ferrari, R. J., Rangayyan, R. M., Desautels, J. E. L., Borges, R. A. & Frère, A. F. (2004*a*), 'Automatic identification of the pectoral muscle in mammograms', *IEEE Trans. Med. Imaging* **23(2)**, 232–245.

Ferrari, R. J., Rangayyan, R. M., Desautels, J. E. L., Borges, R. A. & Frère, A. F. (2004*b*), 'Identification of the breast boundary in mammograms using active contour models', *Med. Biol. Eng. Comput.* **42**, 201–208.

Freeman, J. (1975), 'The modelling of spatial relations', *Computer Graphics and Image Processing* **4**, 156–171.

Freer, T. W. & Ulissey, M. J. (2001), 'Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center', *Radiology* **220**, 781–786.

Garey, M. R. & Johnson, D. S. (1979), 'Computers and intractability: A guide to the theory of np-completeness', *W. H. Freeman, New-York* .

Geman, S. & Geman, D. (1984), 'Stochastic relaxation Gibbs distributions, and the Bayesian restoration of images', *IEEE Trans. on Pattern Analysis and Machine Intelligence* **6**(2), 721 – 741.

Gilbert, F. J., Astley, S. M., McGee, M. A., Gillan, M. G. C., Boggis, C. R. M., Griffiths, P. M. & Duffy, S. W. (2006), 'Single reading with computer-aided detection and double reading of screening mammograms in the united kingdom national breast screening program', *Radiology* **241**(1), 47–53.

Gonzalez, R. C. & Wintz, P. (1987), *Digital Image Processing*, Addison-Wesley, Reading, MA.

Gur, D., Sumkin, J. H., Rockette, H. E., Ganott, M., Hakim, C., Hardesty, L., Poller, W. R., Shah, R. & Wallace, L. (2004), 'Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system', *Journal of the National Cancer Institute* **96**(3), 185–190.

Hadjiiski, L., Sahiner, B., Chan, H. P., Petrick, N., Helvie, M. & Gurcan, M. (2001), 'Analysis of temporal changes of mammographic features: computer-aided classification of malignant and benign masses', *Med. Phys.* **28**(11), 2309–2317.

Haralick, R. M., Shanmugam, K. & Dinstein, I. (1973), 'Textural features for image classification', *IEEE Trans. Syst., Man, Cybern.* **SMC-3**(2), 610–621.

Hopcroft, J. & Wong, J. (1974), 'Linear time algorithm for isomorphism of planar graphs', *Sixth ACM Symposium on Theory of Computing* .

Jolion, J. M. (2003), 'Graph matching: what are we talking about ?', *Pattern Recognition Letters* **24**(8).

Jolion, J. M. & Montanvert, A. (1992), 'The adaptive pyramid: a framework for 2D image analysis', *Computer Vision, Graphics, and Image Processing* **55(3)**, 339–348.

Karssemeijer, N. (1998), 'Automated classification of parenchymal patterns in mammograms', *Phys. Med. Biol.* **43**(2), 365–378.

Kass, M., Witkin, A. & Terzopoulos, D. (1987), 'Snakes: active contour models', *International Journal of Computer Vision* **1**(4), 321–331.

Kok-Wiles, S. L., Brady, M. & Hignam, R. (1998), Comparing mammogram pairs for the detection of lesions, *in* 'Proceedings of the 4th International Workshop on Digital Mammography', IWDM 1998, Nijmegen, The Netherlands, pp. 103–110.

Krissinel, E. B. & Henrick, K. (2004), 'Common subgraph isomorphism detection by backtracking search', *Softw. Pract. Exper* **34**, 591–607 (DOI: 10.1002/spe.588).

Kropatsch, W. G. (1998), From equivalent weighting functions to equivalent contraction kernels, *in* E. Wenger & L. I. Dimitrov, eds, 'Digital Image Processing and Computer Graphics: Applications in Humanities and Natural Sciences', Vol. 3346, SPIE, pp. 310–320.

Kumar, S., Sallam, M. & Goldgof, D. (2001), 'Matching point features under small nonrigid motion', *Pattern Recognition* **34**(12), 2353–2365.

Lallich, S., Muhlenbach, F. & Jolion, J. M. (2003), 'A test to control a region growing process within a hierarchical graph', *Pattern Recognition* **36**, 2201–2211.

Lobregt, S. & Viergever, M. A. (1995), 'A discrete dynamic contour model', *IEEE Trans. Med. Imaging* **14**(1), 12–24.

Ma, F., Bajger, M., Slavotinek, J. P. & Bottema, M. J. (2006), Validation of graph theoretic segmentation of the pectoral muscle, *in* 'Digital Mammography, IWDM 2006, 8th International Workshop', Springer, Manchester, UK, pp. 642 – 649.

Ma, F., Bajger, M., Slavotinek, J. P. & Bottema, M. J. (2007), 'Two graph theory based methods for identifying the pectoral muscle in mammograms', *Pattern Recognition* **40**, 2592–2602.

Marfil, R., Molina-Tanco, L., Bandera, A., Rodríguez, J. A. & Sandoval, F. (2006), 'Pyramid segmentation algorithms revisited', *Pattern Recognition* **39**(8), 1430–1451.

Marias, K., Behrenbruch, C., Parbhoo, S., Seifalian, A. & Brady, M. (2005), 'A registration framework for the comparison of mammogram sequences', *IEEE Trans. Med. Imaging* **24**(6), 782–790.

Martí, R., Raba, D., Oliver, A. & Zwiggelaar, R. (2006), Mammographic registration: proposal and evaluation of a new approach, *in* 'Proceedings of the 8th International workshop on Digital Mammography', IWDM 2006, Manchester UK, pp. 213–220.

Matsakis, P. & Wendling, L. (1999), 'A new way to represent the relative position between areal objects', *IEEE Trans. on Pattern Analysis and Machine Intelligence* **21**(7), 634–643.

McGregor, J. J. (1982), 'Backtrack search algorithms and the maximal common subgraph problem', *Softw. Pract. Exper* **12**, 23–34.

Meer, P. (1989), 'Stochastic image pyramids', *Computer Vision, Graphics, and Image Processing* **45**, 269–294.

Miyajima, K. & Ralescu, A. (1994), 'Spatial organization in 2D segmented images: representation and recognition of primitive spatial relations', *Fuzzy Sets and Systems* **65**(2-3), 225–236.

Montanvert, A., Meer, P. & Rosenfeld, A. (1991), 'Hierarchical image analysis using irregular tessellations', *IEEE Trans. on Pattern Analysis and Machine Intelligence* **13(4)**, 307–316.

Mudigonda, N. R., Rangayyan, R. M. & Desautels, J. E. L. (2000), 'Gradient and texture analysis for the classification of mammographic masses', *IEEE Trans. Med. Imaging* **19**(10), 1032–1043.

Novak, R. (1988), 'Transformation of the female breast during compression at mammography with special reference to the importance for localization of a lesion', *Acta Radiologica Suplementum* **371**, 1–47.

Ojala, T., Näppi, J. & Nevalainen, O. (2001), 'Accurate segmentation of the breast region from digitized mammograms', *Computerized Medical Imaging and Graphics* **25**(1), 47–59.

Sahiner, B., Petrick, N., Chan, H.-P., Hadjiiski, L. M., Paramagul, C., Helvie, M. A. & Gurcan, M. N. (2001), 'Computer-aided characterization of mammographic masses: accuracy of mass segmentation and its efects on characterization', *IEEE Trans. Med. Imaging* **20**, 1275–1284.

Sallam, M. & Bowyer, K. W. (1999), 'Registration and difference analysis of corresponding mammogram images', *Medical image analysis* **3**(2), 103–118.

Sanjay-Gopal, S., Chan, H., Wilson, T., Helvie, M., Petrick, N. & Sahiner, B. (1999), 'A regional registration technique for automated interval change analysis of breast lesions on mammograms', *Med. Phys.* **26**(12), 2669–2679.

Shen, L., Rangayyan, R. M. & Desautels, J. E. L. (1993), 'Detection and classification of mammographic calcifications', *Internat. J. Pattern Recognit. Artif. Intell.* **7**(6), 1403–1416.

Shen, L., Rangayyan, R. M. & Desautels, J. E. L. (1994), 'Application of shape analysis to mammographic calcifications', *IEEE Trans. Med. Imaging* **13**(2), 263–274.

Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C., Ricketts, I., Stamatakis, E., Cerneaz, N., Kok, S., Taylor, P., Betal, D. & Savage, J. (1994), 'The mammographic images analysis society digital mammogram database', *Exerpta Medica, International Congress Series 1069* pp. 375–378.

Sun, Y., Suri, J. S., Desautels, J. E. L. & Rangayyan, R. M. (2006), 'A new approach for breast skin-line estimation in mammograms', *Pattern Anal Applic* **9**, 34–47.

Tanimoto, S. & Pavlidis, T. (1975), 'A hierarchical data structure for picture processing', *Computer Graphics and Image Processing* **4(2)**, 104–119.

Timp, S. & Karssemeijer, N. (2006), 'Interval change analysis to improve computer aided detection in mammography', *Medical Image Analysis* **10**(1), 82–95.

Timp, S., Varela, C. & Karssemeijer, N. (Jul. 2007), 'Temporal change analysis for characterization of mass lesions in mammography', *IEEE Trans. Med. Imaging* **26**(7), 945–953.

Ullmann, J. R. (1976), 'An algorithm for subgraph isomorphism', *Journal of the ACM* **1**(23), 31–42.

Vujovic, N. & Brzakovic, D. (Oct. 1997), 'Establishing the correspondence between control points in pairs of mammographic images', *IEEE Trans. Med. Imaging* **6**(10), 1388–1398.

Wei, D., Chan, H.-P., Helvie, M. A., Sahiner, B., Petrick, N., Adler, D. D. & Goodsitt, M. M. (1995), 'classification of mass and normal breast tissue on digital mammograms: Multiresolution texture analysis', *Med. Phys.* **22**, 1501–1513.

Williams, D. J. & Shah, M. (1992), 'A fast algorithm for active contours and curvature estimation', *CVGIP: Image Understanding* **55**(1), 14–26.

Wirth, M. A., Narhan, J. & Gray, D. (2002), 'Nonrigid mammogram registration using mutual information', *Proc. SPIE (Medical Imaging: Image Processing)* **4684**, 95–102.

Yu, L., Ma, F., Jayasuriya, A., Sigelle, M. & Perreau, S. (2007), A new contour detection approach in mammogram using rational wavelet filtering and MRF smoothing, *in* '9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications (DICTA 2007)', IEEE., Glenelg, South Australia, pp. 106–111.

Yu, L. & White, L. B. (2005), 'Broadband Doppler compensation for rational wavelet-based UWA communication systems', *Asia-Pacific Conferences on Communications (APCC05)* pp. 605 – 609.

Yu, L. & White, L. B. (2006), 'Complex rational orthogonal wavelet and its application in communications', *IEEE Sig. Proc. Letters* **13**(8), 477–480.

Yu, L. & White, L. B. (2007), 'Optimum receiver design for broadband Doppler compensation in multipath/Doppler channels with rational orthogonal wavelet signalling', *IEEE Trans. Sig. Proc.* **55**, 4091 – 4103.