Towards Automatic Image Analysis for Computerised Mammography

Christina Olsén
colsen@cs.umu.se

UMINF-08.04
Detta verk skyddas enligt lagen om upphovsrätt (URL 1960:729).

© Christina Olsén, Frame.
© SPIE, Paper II.
© Springer Verlag, Papers I, III, IV.
© Elsevier, Paper V.

Printed by Print och Media, Umeå University, 2008: 2004607.
To my beloved

Olivia and Jesper
Abstract

Mammographic screening is an effective way to detect breast cancer. Early detection of breast cancer depends to a high degree on the adequacy of the mammogram from which the diagnosis is made. Today, most of the analysis of the mammogram is performed by radiologists. Computer-aided diagnosis (CAD) systems have been proposed as an aid to increase the efficiency and effectiveness of the screening procedure by automatically indicating abnormalities in the mammograms. However, in order for a CAD system to be stable and efficient, the input images need to be adequate. Criteria for adequacy include: high resolution, low image noise and high image contrast. Additionally, the breast needs to be adequately positioned and compressed to properly visualise the entire breast and especially the glandular tissue.

This thesis addresses questions regarding the automatic determination of mammogram adequacy with the focus on breast positioning and segmentation evaluation. The goal and, thus, the major technical challenge is to develop algorithms that support fully automatic quality checks. The relevant quality criteria are discussed in Chapter 2. The aim of this discussion is to compile a comprehensive list of necessary criteria that a system for automatic assessment of mammographic adequacy must satisfy. Chapter 3 gives an overview of research performed in computer-aided analysis of mammograms. It also provides basic knowledge about image analysis involved in the research area of computerised mammography in general, and in the papers of this thesis, in particular. In contrast, Chapter 4 describes basic knowledge about segmentation evaluation, which is a highly important component in image analysis. Papers I–IV propose algorithms for measuring the quality of a mammogram according to certain criteria and addresses problems related to them. A method for automatic analysis of the shape of the pectoralis muscle is presented in Paper I. Paper II proposes a fully automatic method for extracting the breast border. A geometric assumption used by radiologists is that the nipple is located at the point on the breast border being furthest away from the pectoralis muscle. This assumption is investigated in Paper III, and a method for automatically restricting the search area is proposed. There has been an increasing need to develop an automated segmentation algorithm for extracting the glandular tissue, where the majority of breast cancer occur. In Paper IV, a novel approach for solving this problem in a robust and accurate way is proposed. Paper V discusses the challenges involved in evaluating the quality of segmentation algorithms based on ground truths provided by an expert panel. A method to relate ground truths provided by several experts to each other in order to establish levels of agreement is proposed. Furthermore, this work is used to develop an algorithm that combines an ensemble of markings into one surrogate ground truth.
Acknowledgements

I would like to express my sincere gratitude to the following people:

Senior Lecturer Fredrik Georgsson, my associate supervisor and co-author of papers III and V, for giving me the opportunity to write this thesis, for his continuous encouragement and for spending time discussing all kinds of issues within the scope of this thesis.

Associate Professor Frank Drewes for his support and encouragement, for thoroughly reading this thesis, for valuable and instructive comments and for being my supervisor.

Niclas Börlin for reading and commenting this thesis. I would also like to thank him for answering all kinds of \( \LaTeX \) questions, for discussing all kinds of questions concerning research thoughts and for helping me out in finalising this thesis in the dying moments before the printing deadline.

The staff at the Department of Computing Science for their help and support during lecturing and concerning research questions.

The technical and administrative staff at the Department of Computing Science for always helping me out with all my questions with such enthusiasm and high proficiency.

I would like to thank Aamir Mukhdoomi and Mattias Pierre, the co-authors of Paper IV and Paper V, respectively.

On a more personal level I would like to thank:

My parents Ewy Olsén, Torsten Hjärtstam, Ingvar Olsén and Ulla-Britt Olsén for their belief in me, help and support.

My brothers and sisters, Åsa, Erik, Robert, Anders, and Emilie.

My grandma Edith Olsén for her indefatigable encouragement, cheering, joy and love.

My uncle Arne Olsén for planting a scientific seed that eventually became this thesis.

Helena, Lena P., Erik, Anders, Ulrika, Yvonne, Johan, Anne-Lie, Stefan H. Mikael and Lena KW. and all the rest of you at the department for making the time in the “fika-room” and time off work pleasant as well.

Synnöve Solbakk Karlsson for being a good friend and for sharing my love for cats.

Kristina Rylander for being one of my best friends.

Niclas Börlin for all the experiences we have shared (including astronauts, fallen trees and locked buildings), for all the steps we have taken together and individually with each others' support towards finding the answers to the difficult questions concerning life, and for being my bonus-brother that I always can count on. You’re the best!

Olivia Rönqvist for being a pleasant input in my life.
Paulina Wikstén, my bonus-daughter, for welcoming me into her life.
Jonas Wikstén for the love and support you have given me during the end of this journey. Also and most importantly for making me and my life complete. I love you!

Finally, and most importantly, I would like to thank my children, Olivia and Jesper for their love and support, and for their patience. You are my inspiration for everything in life. This thesis would not have been possible without the two of you by my side. I love you!

Umeå, May 2008
Christina Olsén
Preface

This thesis is divided into the following parts:

**Frame:** Chapter I–V.


**Appendix A:** Quality Assessment Questionnaire — Part I, questions.

**Appendix B:** Quality Assessment Questionnaire — Part II, mammogram.
In addition to the papers included in this thesis the following papers have been produced in relation to the Ph.D. studies:


# Contents

1 Introduction .............................................. 1  
   1.1 Relations to Other Research Areas ...................... 2  
   1.2 Objectives ........................................ 2  
   1.3 Outline ........................................... 3  

2 Medical Background ...................................... 5  
   2.1 The Anatomy of the Female Breast ..................... 5  
   2.2 Breast Cancer ...................................... 7  
   2.3 Mammography ....................................... 9  
   2.4 Mammographic Imaging Techniques ..................... 11  
   2.5 Quality Criteria Concerning Breast Positioning ...... 21  
   2.6 Mammographic Interpretations ......................... 28  

3 Computer-Aided Analysis in Mammography ................. 29  
   3.1 Computerised Analysis of Mammograms ................. 29  
   3.2 Computer-Aided Detection ........................... 30  
   3.3 Computer-Aided Diagnosis ........................... 30  
   3.4 Image Analysis ..................................... 31  
   3.5 Segmentation of the Breast Region ................... 48  
   3.6 Lesion Extraction ................................... 49  
   3.7 Clinical Aspects of Computerised Aids ................. 56  
   3.8 Mammographic Image Databases ....................... 58  
   3.9 Future Prospects ................................... 59  

4 Segmentation Evaluation ................................. 61  
   4.1 Background ......................................... 61  
   4.2 Analytical Methods .................................. 62  
   4.3 Empirical Methods .................................. 62  
   4.4 Segmentation Evaluation in Image Analysis for Computerised  
       Mammography ....................................... 67  
   4.5 Discussion ......................................... 68
5 Summary of Contributions

5.1 Detecting the Pectoralis Muscle (Paper I) ................. 72
5.2 Extracting the Breast Border (Paper II) .................... 73
5.3 Detecting the Nipple (Paper III) ........................... 73
5.4 Extracting the Glandular Tissue (Paper IV) ............... 74
5.5 Identifying Ground Truth Based on Multiple Markings by Do-
main Experts (Paper V) ........................................ 75

6 Future Work

6.1 The Decision Making Process of Experts .................... 79
6.2 Assessing Ground Truth Useful in Segmentation Evaluation for
Real-Life Problems ............................................. 80

References .......................... 81

Paper I .................................. 93
Paper II .................................. 105
Paper III .................................. 127
Paper IV .................................. 141
Paper V .................................. 157
Appendix A ................................ 185
Appendix B ................................ 193
Chapter 1

Introduction

Breast cancer is the most common form of cancer in Sweden (Nyström, 2000). Despite the increased incidence of breast cancer, the mortality from breast cancer has decreased in all ages. A possible reason is the introduction of mammographic screening – x-ray imaging of the breast in large groups of asymptomatic patients. Studies indicate that screening plays an important role in detecting cancers at an early stage (Nyström, 2000; Georgsson, 2001). Mammography screening is defined as the continuous evaluation of an asymptomatic group of women, where the signs of cancer are either very subtle or not present at all. The primary reason for performing mammography is to detect breast cancer in such an early stage that it will be clinically impalpable, i.e. difficult to detect via a physical examination. Mammography is currently the best method for the early detection of breast cancer (Nishikawa, 2007). In approximately two-thirds of the false-negative diagnosed mammograms, the radiologist failed to detect a cancer that was evident retrospectively. The lack of detections may be due to the subtle nature of the radiographic findings, poor image quality, eye fatigue, or oversight by the radiologist. Therefore, it is common practice that two radiologists should analyse each mammogram to increase the sensitivity (Thurfjell et al., 1994). To minimise this to the extent possible, a computer-aided diagnosis (CAD) system can serve as a secondary opinion. This is expected to improve consistency by providing a standardised approach to mammogram interpretation, and increasing detection sensitivity.

However, an early detection of breast cancer depends on the adequacy of the mammogram from which the diagnosis is made. In turn, the adequacy of the mammogram is dependent on the image acquisition procedure. Thus, a diagnostic quality assurance system would be an important step when constructing a computer-aided diagnosis system. Using a quality assurance program at the medical facilities would maximise the likelihood of consistently producing high quality images, so that the mammograms provide adequate diagnostic information with low radiation dose. The latter is important since radiation is by itself carcinogenic. However, in a clinical quality evaluation carried out within the American College of Radiology Mammography Accreditation Program, 1034 out
of 2341 mammograms (44%) failed in quality, which shows that medical facilities must strive to improve the quality of mammography. The quality evaluation revealed that improper breast positioning was the most frequent problem, followed by problems related to breast compression, x-ray exposure, and image sharpness (Bassett et al., 2000).

It can even be argued that a diagnostic quality assessment system is more important than a CAD system in order to improve the efficiency of mammographic screening. Such system would identify mammograms with inferior diagnostic quality during image acquisition and thus prevent unnecessary recalls due to technical problems since low-quality images could be retaken immediately. This is important since there is mental pressure on the patient to be recalled for another examination, among other things because of the concern that there might be cancer present (Gilhuijs et al., 2002). A diagnostic quality assessment system may also prevent failure in the algorithms of a CAD system, which are caused by inadequate images. Finally, if the recall rate due to inadequacy in image quality can be reduced, the screening will be more cost-effective and more comfortable for the patient.

This research area contains a number of interesting problems within computing science in general and image analysis in particular, e.g. texture classification, expert systems, object classification, and segmentation. An important research question addressed within this thesis is performance evaluation. Since segmentation evaluation is essential for providing a scientific basis for image analysis in general and for medical image analysis in particular, this question is as important as the development of the segmentation algorithms themselves.

1.1 Relations to Other Research Areas

Computer-aided analysis in mammography mainly concerns the development of algorithms, but it includes elements from several other research areas as well. Anatomy, breast cancer, and clinical aspects are related to medical areas such as epidemiology, oncology, and radiology. Mammography techniques are closely related to engineering sciences and physics. Both mathematics and mathematical statistics are involved in pattern classification and pattern recognition, but even more in image processing and image analysis. Cognitive science is involved in the decision making process because of the need to model the cognitive ability of the expert. Most applications in computer-aided mammography require a way of handling uncertainty in order to make adequate decisions.

1.2 Objectives

In order for the radiologist to make an accurate diagnosis, the ability to recognize structures in the mammogram needs to be considered. The visibility of these structures depends on variables such as resolution, image noise, and contrast. These variables must be optimised to keep the radiation dose as low as possi-
ble. Another important requirement is to determine if all regions of interest are properly imaged. As an ongoing quality control activity, mammograms should be evaluated for quality by both the personnel performing the examination and the radiologist interpreting the examination (Kimme-Smith et al., 1997).

Automatic quality assurance of mammograms will assist the personnel working with mammography. Thus, the main goal of the research presented in this thesis is to develop methods for the automatic determination of mammogram adequacy. This includes a discussion of the research area “computer-aided analysis in mammography”, for instance accounting for general mammography, computerised analysis of mammograms and image analysis techniques. In particular, both old and new problems within the area will be identified.

A computerised quality assessment system needs to extract, analyse and evaluate mammograms according to the quality criteria an adequate mammogram must fulfil. Therefore, the quality criteria for diagnostic quality in mammography will be addressed. These include: breast positioning, breast compression, contrast, exposure, noise, sharpness, and artefacts. Furthermore, this thesis contains a review of recent literature within the mammographic image analysis. The focus will be on the field of computerised mammographic image analysis and quality assurance.

1.3 Outline

What follows is a short description of the remaining chapters of this thesis.

Chapter 2, provides an orientation of the technical and mammographic concepts. It also describes and presents all the quality criteria investigated and compiled from the literature and interviews.

Chapters 3, presents concepts and work related to computer-aided mammography systems. The sections concerning image analysis are supposed to be useful for the understanding of the difficulties of computer-aided diagnosis (CAD). The aims of these sections are to describe some of the image analysis techniques used in the papers and in computerised mammography in general. Another aim of Chapter 3 is to introduce the reader to CAD as a research area. Some of the material found in publications on CAD systems is of direct interest for a quality assessment system.

Chapter 4, address the challenges involved when conducting segmentation evaluation. Segmentation evaluation can be grouped into analytical and empirical methods. However, the focus of the chapter is on empirical methods where an ideal truth is impossible to acquire.

Chapter 5, describes the aims and contributions of the five papers included in this thesis.

Chapter 6, describes some interesting and important questions that remain to be addressed within these research areas.
Chapter 2

Medical Background

Since computerised mammography is the topic of this thesis it is essential to gain a better understanding of the problems involved in taking and assessing mammographic images. Therefore, an introduction of the anatomy of the female breast is given as well as a section describing breast cancer.

2.1 The Anatomy of the Female Breast

The female breast is highly specialised for the production of milk to nourish infants. Yet, its importance exceeds its biologic function. The breast is also important for the self-esteem of the woman. Therefore, cancer in the breasts has a serious effect on the health of the woman, and if the breast must be surgically removed, it can be psychologically detrimental for her.

The size and weight of a breast changes during the life of a woman. They are mainly determined by the amount of adipose (fatty) tissue present in the breast. As the female approaches puberty, her breasts grow in response to the female hormone oestrogen. The second stage in breast development occurs when ovulation begins, caused by progesterone production. The development is mostly completed within one and a half years after the menarche (first menstruation), when the development of alveoli (small sacs within the breast) continues in response to hormonal stimulations. The breasts gradually atrophy after menopause.

The cone-shaped structure topping the breast is known as the mamilla, or nipple. In front, breast tissue may extend from the collarbone to the middle of the breastbone. On the side, breast tissue may reach as far as the latissimus dorsi (the muscle extending from the lower back to the humerus bone of the upper arm) and it may continue into the axilla.

Figure 2.1 illustrates important anatomical details of the breast, together with the corresponding terminology. The sweat glands, that are adapted to secrete milk, contain a number of secretory lobes, where each lobe is arranged around one of the secondary tubules and is connected with a main duct that
Figure 2.1: The anatomy of the female breast (Gaudin and Jones, 1989).
leads to the nipple. Every lobe is made up of smaller lobules which, in turn, are composed of the small sacs mentioned above, called alveoli. A lobule consists of a single layer of secretory cells arranged around a duct. About 15 to 20 excretory ducts come together near the areola, where they form enlarged duct terminals. These ducts serve as milk receptacles along the path to the nipple surface (Vorherr, 1974).

A piece of connective tissue known as fascia encloses each breast. Fasciae also extend down into the breasts as “dividing walls”. The uppermost fascia separates the breast tissue from the skin. The deep fascia separates the breast tissue from the muscle of the chest wall. Cooper’s ligaments, the fibre-like bands that are attached to the pectoralis muscle, help to support the breasts (Vorherr, 1974; Cardenosa, 1996). The medical term for function-specific cells of a gland or an organ, contained in and supported by the connective framework, is parenchyma, e.g. the functioning tissue of the breast.

The breasts cover important muscles of the chest wall, such as the pectoralis, which is attached to the collarbone, breastbone and the cartilage of most of the ribs. The pectoralis also connects with the humerus bone of the upper arm. Another chest muscle is the pectoralis minor, which is a triangular muscle that lies under the pectoralis and is attached to the third, fourth, and fifth ribs. This connects with the intercostal muscles, the muscles between the ribs, and the shoulder joint.

In the breast, most lymphatic vessels drain into a network of lymph nodes that are located around the edges of the breast or in nearby tissues of the axilla and collarbone. Such lymph nodes are embedded within fat pads, an arrangement that complicates lymph node removal during breast cancer surgery. The axillary lymph nodes are especially important in breast cancer, because they are often the sites of metastasis (Vorherr, 1974).

Terms that determine certain locations within a specific part of the human body are of widespread use in the medical literature. The terms of location which are of interest in connection with the female breast are shown in Figure 2.2.

### 2.2 Breast Cancer

Breast cancer is by far the most common cancer disease among western women. About every tenth woman develops breast cancer during her lifetime (Nyström, 2000). For women under the age of 50 it is one of the most common causes of death. The disease consumes a considerable amount of resources within the health and medical care system. Apart from the severe threat against health and life, the disease also affects the identity of the woman.

Breast cancer is relatively uncommon in individuals younger than 35 years of age. Although the incidence increases with age, the disease is of particular concern for women at the age of 40–74. Breast cancer among men is rare. Survival from breast cancer has improved considerably during recent decades. Screening programmes with mammography were gradually introduced during
the 1990s into the four counties of northern Sweden. The aim was to reduce breast cancer mortality. It was concluded that 6 years after their introduction the population-based service screening programmes substantially reduced effect on breast cancer mortality in women aged 50–69 years at invitation to screening (Lenner and Jonsson, 1997).

In D’Orsi and Newell (2007b) it is shown that the survival of women with invasive breast cancer varied according to detection mode. Detection mode was categorised as cancer detected in women not yet invited to screening (uninvited), screening-detected, interval-detected and cases detected among women who were invited to, but did not attend, screening (non-participants). A significant survival difference in favour of women with cancer detected in the intervals between screenings compared with uninvited cases was found. Furthermore, it was shown that breast cancer cases detected in the screening intervals (interval cancers) constitute 20–40% of the incident cases in a population-based screening programme.

Breast cancer is defined as a malign tumour that starts out in the glandular tissue within the breast. It can appear as a non-invasive cancer, called cancer in situ. In these cases, the cancer cells are confined within the basement membrane and have not invaded surrounding tissues. Cancer in situ can be of ductal type, which means that the cancer starts growing in the gland ducts, or it can be of lobular type, which starts out in the gland lobule. The most common type of breast cancer is ductal cancer (Cardenosa, 1996). Lobular cancer is more often found in both breasts than other types of breast cancer.

When the cancer invades surrounding tissue it is called invasive breast cancer. The invasive breast cancer has the ability to metastasise, which means that the disease spreads from one part of the body to other parts through the bloodstream or the lymph system. At the time of discovery of invasive cancer,
the patient prognosis is related to several aspects (Giger et al., 2000):

1. the size of the lesion;
2. a variety of histological, biochemical, and genetic indicators;
3. its spread, if any, to regional lymph nodes; and
4. its spread to distant sites in the body.

A general practitioner may notice changes in the breasts. If that is the case, the general practitioner may suggest that the woman should have a mammogram taken. A mammogram is a special x-ray image of the breast that may depict tumours that are too small to be palpable. If a mass is found in the breast, the radiologist pulls out a small piece of the mass, a core biopsy, by inserting a needle into the breast (Heywang-Köbrunner et al., 1997). A histological study is undertaken to analyse if the cells are malignant.

### 2.2.1 Radiographic Manifestations of Breast Cancer

The radiographic discovery of breast cancer is possible because of the local manifestations of an ever-increasing number of neoplastic cells. Usually, what is found belongs to one of the two large categories microcalcifications and masses (Giger et al., 2000). Microcalcifications related to malignancy typically arise in ducts that are occluded by neoplasms and are usually as small in size as a tenth of a millimetre. However, many are only visible on magnification mammograms or during microscopic analysis of surgical specimens. Characteristic of malignancy are morphologies calcifications, while other forms can be seen in both malignant or benign disease. The earliest detectable breast cancers usually manifest radiographically with microcalcifications (Highnam and Brady, 1999; Giger et al., 2000; Karahaliou et al., 2007). If the microcalcifications are arranged as clusters they are strong indicators of malignancy, and they appear in 30-50% of the mammographically diagnosed cases (Karahaliou et al., 2007).

Unlike microcalcifications, neoplasms of the breast can also be presented as mass-like findings on a mammogram, that can easily be obscured by the fibroglandular elements of the breast. Most breast cancers that are presented as masses are invasive cancers. Mass lesions projecting in dense areas of the breast may, however, be visible on the basis of asymmetry in bilateral comparison of mammograms. Other radiographic manifestations of breast cancer are: skin thickening, nipple or skin retraction and vascular asymmetry. These are typically associated with a more advanced disease. Such findings are of less value in the screening setting, but are useful to confirm radiographic impressions of malignancy (Giger et al., 2000).

### 2.3 Mammography

There are basically two different kinds of mammographic examination, Clinical Mammography and Mammographic Screening.
2.3.1 Clinical Mammography

The clinical examination often starts with a contact with a physician as a concern of symptoms. A complete breast examination includes the physical examination of the breast in addition to mammography. The physical examination begins with a visual inspection looking for asymmetry in size, changes in contour and shape, and changes of the skin and nipple. After the visual inspection, a palpation is performed, which provides information on the structure of the glandular tissue and possible differences between breasts, such as location and consistency of lumps to the surrounding breast tissue, the skin, and the pectoralis muscle. During palpation, pain and mobility in the nipple and surrounding regions can be assessed. These findings should be documented because information from the physical examination is useful in interpreting the mammogram. Whenever a clinical finding exists, further examination needs to be done by using mammography.

2.3.2 Mammographic Screening

Mammographic screening is defined as examinations performed regularly on a broad population group that does not show any sign of the disease. The goal is to detect hidden breast cancer at an early stage, the primary task being searching, as opposed to diagnosing. Potential abnormalities visible on screening mammograms are more extensively evaluated during diagnostic examinations in which a variety of additional mammographic views may be taken. Importantly, mammography does not replace careful physical examination (Heywang-Köbrunner et al., 1997). Studies during the last twenty years have given evidence that mammographic screening can significantly reduce mortality from breast cancer (Nyström, 2000; Socialstyrelsen, 1998; Heywang-Köbrunner et al., 1997).

However, screening involves very large groups to be examined. The highest interest lies within the group that shows an increased probability of developing the disease. Since women younger than 40 and men are not a potential risk group, they are usually not involved in screening programmes. In order to motivate such a screening program the risk of getting the disease must be severe.

The technical image quality of a mammogram does not differ between a screening mammogram and a mammogram taken during a clinical examination. However, in the latter case, it is the woman who contacts the physician. Therefore, clinical mammograms are more likely to show more developed changes, such as lumps or skin changes, than mammograms taken during screening. Consequently, it seems likely that the clinical mammogram will be easier to judge for the radiologist than the screening mammogram. Thus, efforts must be put on interpreting the mammogram when constructing computerised equipment for screening.

Furthermore, there are clear indications that radiologists do not detect all breast cancers that are visible in retrospect on screening mammograms (Giger et al., 2000). Often no available screening modality is uniquely ideal. During the last two decades, tremendous advances have been made in the performance and
interpretation of breast magnetic resonance imaging (MRI) examinations (Orel, 2008). MRI of the breast is rapidly gaining popularity in clinical practice in both the diagnostic setting and, more recently, in the screening setting. For breast MRI, there is an increasing body of observational data showing that screening can identify cancer in patients of specific risk groups, i.e. high-risk related to family history (Saslow et al., 2007). There are clinical indications that have emerged where MRI, as an adjunct to mammography, seems to be the imaging study of choice (Orel, 2008; Saslow et al., 2007). Examinations carried out by using magnetic resonance imaging are further addressed in Section 3.7.

2.4 Mammographic Imaging Techniques

Mammography demands higher stringency on equipment and image quality compared with radiographic studies of other parts of the body. In addition, there are special demands on the positioning of the breast, which make mammography one of the most difficult radiographic examinations (Heywang-Köbrunner et al., 1997).

Some examples of the specific requirements regarding image quality and positioning are the following:

- Fine microcalcifications and fibrotic strands must be imaged sharply, with high contrast and a low level of image noise.

- Even if the highest possible contrast is achieved, mammography must permit adequate assessment of areas of greatly varying density. These can be fatty areas behind the nipple or close to the skin in small breasts, areas of dense fibrocystic tissue in large breasts, and the tissue overlying the pectoralis muscle. This imposes particularly stringent demands on the imaging system.

- Since the mammary tissue is sensitive to radiation, the examination should involve the minimum dose of radiation sufficient for producing an image of adequate quality.

- Imaging the entire body of the breast tissue is very important for an accurate assessment in screening examination and clinical diagnostic problem solving. In order to achieve accurate assessment, optimum standard positioning is important. Even knowledge regarding additional views of the breast, whenever needed, is necessary.

It is of utmost importance that radiologists and radiographers are thoroughly familiar with mammographic techniques and persistently supervise quality. There are studies showing that only an optimal imaging technique ensures early detection of breast cancer. Stringent requirements on both the equipment and expertise are a prerequisite for good results (Heywang-Köbrunner et al., 1997; Perry et al., 2001). The National Board of Health and Welfare in Sweden has published their demands on both technique and knowledge concerning
quality assurance (Socialstyrelsen, 1996). In the USA, similar demands are controlled by an external review which must be performed at least every third year by specially trained and qualified interpreting physicians under the support of the Food and Drug Administration (Kimme-Smith et al., 1997).

In the sections below, the components of the mammographic equipment are described. Some are more important for the understanding of the mammographic technique and are thus described in more detail. In Figure 2.3, an overall description of the mammographic equipment is shown. Further details concerning technical aspects of mammography can be found in (Heywang-Köbrunner et al., 1997; Kimme-Smith et al., 1997; Olsén, 2003; Perry et al., 2001).

2.4.1 Physics of X-ray Imaging

A special x-ray tube that produces low-energy radiation is required in mammography in order to produce high tissue contrast. The low-energy radiation is accomplished by combining special targets and filters.

Mammography tubes must have an extremely small focal spot in order to
obtain the required sharpness (Heywang-Köbrunner et al., 1997). The local projection of the width of the focal spot will vary depending on its distance from the chest wall and the angularity of the tube. Focal spots in this size range put practical limitations on tube current, and thus x-ray output, with exposure times that are often large fractions of seconds. Motion unsharpness is therefore of serious concern for the quality of the mammogram. This can be avoided by a proper compression of the breast, but also by minimising the exposure time. Proper geometric configuration is important in order to achieve the necessary sharpness, thus avoiding geometric blurring. Geometric blur is caused by technical aspects, such as an increase in focal spot size, a longer object-to-film distance, or a shorter source-to-image distance. Thus, the focal spot sizes of the mammography equipment have been reduced for e.g. magnification mammography (Kimme-Smith et al., 1997). Importantly, adequate compression is necessary to reduce the distance to the film. On the film plane shown in Figure 2.4, the light grey area surrounding the projection of the object is called the penumbra. The penumbra is an example of geometric blur.

The radiation produced in an x-ray tube consists of a spectrum of electromagnetic energies. Since the spectrum of imaging radiation greatly influences contrast and radiation dose, there are some aspects to consider. With the low-energy radiation slight differences in density can be visualised with high contrast. Increased energy of the radiation decreases the soft-tissue contrast. By balancing the energy radiation carefully, it will be high enough to penetrate thick breasts and breasts with abundant fibrotic or glandular tissue. Insufficient energy of radiation will not penetrate the breast and cannot be compensated even with long exposure time. Since such radiation will unnecessarily increase the radiation dose it is not suitable for imaging. Because of this complexity of the energy selection, it is advisable to adapt the radiation spectrum as closely as possible to the thickness and density of the breast. The radiation spectrum
is determined by the target/filter combination of the x-ray tube and the peak kilo-voltage (kVp) setting on the x-ray unit (Heywang-Köbrunner et al., 1997). Scattered radiation is an undesirable effect in breast imaging. It causes opacification and decreased image contrast and thus limits the information available on the film, which hampers the ability to diagnose properly (Eklund et al., 1994). Scattered radiation arises as the beam penetrates the breast tissue. By adequate breast compression and the use of the grid technique, the effect of scattered radiation can be reduced effectively (Perry et al., 2001). The grid consists of strips of lead, which are focused on the focal spot. The lead strips absorb obliquity oriented radiation, whereas radiation parallel to the lead strips passes through. The grid is then moving perpendicular to the path of the beam during the exposure. The movement of the grid prevents stripes from appearing as shadows on the mammogram. The grid absorbs scattered radiation but also some of the useful radiation. Consequently, it requires a longer exposure time and thus increased radiation dose. This is a problem because of the fact, mentioned above, that the radiation dose should be kept as low as possible in order to prevent the patient from being exposed to unnecessary radiation. Adequate breast compression reduces scattered radiation, since compression shortens the distance through the tissue passed by the beam.

The automatic exposure control system uses a photocell placed under the cassette to determine the radiation dose. When the required dosage for a mean optical film density has reached the photocell, the beam is switched off. To get correctly exposed images, it is important to place the photocell under a representative part of the glandular tissue. If the breast does not completely cover the photocell, this results in an underexposed image, because the photocell turns off the beam too soon (Heywang-Köbrunner et al., 1997).

Underexposure of dense tissue increases the risk that breast abnormalities may be masked, which may lead to missing lesions. Overexposed images of fatty areas and skin can often be evaluated adequately with high-intensity light (Eklund et al., 1994; Kimme-Smith et al., 1997).

### 2.4.2 Image Sharpness

Kimme-Smith et al. (1997) describe sharpness as the ability of the mammography system to define an edge. One type of unsharpness is screen unsharpness resulting from light dispersion. A single x-ray photon absorbed in the screen is converted to a large number of visible light photons, and the spread of these photons from the point of the x-ray interaction with the screen to where they are absorbed by the film creates blurring of fine details. Faster screens (i.e. more sensitive), which are usually thicker, yield greater photon spread and increased screen unsharpness.

In mammographic screen-film systems, low-dose screen-film systems are used almost exclusively as the image receptor system. Since film resolution is always higher than that of the screen (Kimme-Smith et al., 1997), the resolution of the screen-film system is primarily determined by the resolution of the intensifying screen. If a double-emulsion film is used, the image is captured on each side of
the film, separated by the width of the film base. This results in slightly offset details i.e. blurring of fine details. Consequently, double-emulsion films have not gained wide acceptance despite the lower radiation dose and longer tube life that they provide (Kimme-Smith et al., 1997).

It is also important to guarantee a close contact between film and screen to prevent further spread of light from the screen before it reaches the film. Poor film-screen contact can result from poorly designed or damaged cassettes, improper placement of the film in the cassette, dirt lying between the film and the screen, or air trapped between the film and the screen at the time the film is loaded (Kimme-Smith et al., 1997).

Spatial resolution is usually measured with patterns of closely spaced lines. It is defined in terms of line pairs per millimetre (lp/mm). According to Heywang-Köbrunner et al. (1997) most screen-film systems currently used give a high-contrast spatial resolution of 14-18 lp/mm, which can be verified using a lead line grid. The size of the smallest detectable object can be estimated from the number of lp/mm as corresponding to the width of a line.

### 2.4.3 Image Contrast

Because of the fact that the various tissue types, such as fat, glandular and fibrous tissue, but also calcifications within the breast, have small differences in intensity, it is important to look for methods that magnify these subtle variations. Due to small density variations within the breast, contrast requirements used in general radiography, such as x-rays of the skeleton, are not sufficient, even though many of the standards used in general radiography are useful to increase contrast. However, the need for maximised contrast is far more critical for mammography than in general radiography (Eklund et al., 1994).

Formally, contrast can be defined as the relative difference in density between an object and its surroundings (Heywang-Köbrunner et al., 1997):

\[
Contrast = \frac{Density_{obj} - Density_{surroundings}}{Density_{surroundings}}.
\]  

(2.1)

There are various factors that affect contrast, such as breast thickness and density of the tissue, but also radiation qualities. The radiation qualities involve the target/filter combination, kilovolt-peak settings, breast compression, scatter-reducing techniques (e.g. grids), and the choice of an adequate screen-film system.

For a mammogram to be useful, the optical density of regions composed of glandular tissue versus those composed of fat must be different. Glandular tissue should be light grey to white and regions of fat dark grey to black on the image, see Figure 2.5(b) for example. However, with too high contrast it may be impossible to see both thick and thin parts of the breast on the same image, see Figure 2.5(a).
Towards Automatic Image Analysis for Computerised Mammography

Figure 2.5: a) An underexposed mammogram. In this image, tissue details within the dense fibroglandular tissue (asterisk) in the lateral aspects of the breast is not clearly depicted due to underexposure of the dense tissue. b) A properly exposed mammogram. In this mammogram, more tissue details in the dense area and greater differences in optical density between radiodense fibroglandular and radiolucent fatty tissue can be seen, due to improved contrast. ©Lippincott Williams & Wilkins.

2.4.4 Artefacts

Mammography artefacts are any density variations on the image that do not reflect true attenuation differences in the breast, thus complicating the interpretation of the image. Artefacts can conceal remarkable abnormalities or create fictitious lesions that will challenge even the best radiologist (Eklund et al., 1994). An image should not show any scratches, shades or other marks indicating artefacts, and therefore this should be tested daily using a phantom made of synthetic material. There are six major types of artefacts (BenComo, 2000): 1) artifacts caused by the x-ray equipment; 2) patient artefacts; 3) cassette/film/screen artefacts; 4) artefacts caused by improper film-handling (in cases of analog systems); 5) film processor artefacts; 6) random artefacts.

The presence of multiple artefacts on images suggests problems in quality control at a medical facility. Artefacts related to the dedicated mammography equipment include grid lines and improperly aligned equipment components that are superimposed on the image. In (Kimme-Smith et al., 1997), examples such as improper size, alignment, or design of the compression device are given.
Figure 2.6: a) In this mammogram, the hair of the patient (arrow) is superimposed on posterior aspects of the image. b) On the posterior inferior part of this mediolateral oblique view, we can see that a finger (arrow) from the hand displacing the contralateral breast is superimposed on the image. ©Lippincott Williams & Wilkins.

Body parts other than the breast can also be superimposed on the image. These include the hair of the patient (Figure 2.6(a)) and the hand used to displace the contralateral breast from the area of the image (Figure 2.6(b)).

2.4.5 Image Acquisition

In Sweden, mammographic screening is still almost exclusively performed with radiographic screen-film systems using two standard projections. The cranio-caudal (CC) view is performed with the x-ray source above the breast and the receptor below. The mediolateral oblique (MLO) view is performed with the x-ray tube medial to the breast and the receptor position laterally, the tube and receptor angled approximately with 30–70° inclination depending on the body posture. When properly obtained, the MLO view is the most important view since it best visualises the tissue adjacent to the chest wall and the axillary tail, where the majority of cancers occur (Kimme-Smith et al., 1997). Whether CC and MLO mammograms allow optimal mobilisation of the glandular tissue away
from the chest wall is highly dependent on the skill of the radiographer taking the image. Greater mobility of the inferior and lateral margins of the breast aid in the positioning process (Heywang-Köbrunner et al., 1997; Eklund et al., 1994).

Concerning mammographic positioning, it is essential that the breast is properly compressed by a flat radiolucent paddle that is affixed to the tube-image receptor axis. Correct positioning of the breast is also significant in order to ensure that the entire glandular body is imaged (Perry et al., 2001; Heywang-Köbrunner et al., 1997), see Section 2.5 for further details.

**Breast Compression**

To summarise, a proper breast compression is important for the following reasons (Heywang-Köbrunner et al., 1997; Perry et al., 2001):

- It decreases the distance the beam passes, thus improving the contrast of the images by reducing scattered radiation.

- The thickness of the breast is reduced, which decreases the radiation dose and geometric unsharpness.

- The various healthy structures in the breast are separated, whereas true masses persist, thus reducing overlapping of tissue shadows and resulting in a better visualisation of the breast tissue.

- Blurriness caused by motion of the breast during exposure is reduced.

- The thickness of the breast is made more uniform so that the optical densities of the resulting film are more likely to correspond to subtle attenuation differences rather than to differences in tissue thickness.

Figure 2.7(a) visualises inadequate compression manifested by overlapping breast structures. Furthermore, Figure 2.8 visualises a skin fold that obstructs diagnosis, also caused by superimposed tissue.

**2.4.6 Digitisation**

The availability of high-speed computers and high-resolution film digitisers gives rise to changes in the application of computer vision to mammography. Research groups within academia and industry are developing computerised algorithms for the detection or characterisation of masses and clustered microcalcifications in digital mammograms.

The parameters of the acquisition of a digital image are important in the development of any computerised image analysis method. Mammograms can be acquired in digital form either with a film digitiser or with a direct digital acquisition device. The quality of the initial mammogram is important, since poor film image quality affects both radiologist and computer performance in
Figure 2.7: a) Drooping breast. This is an inadequate mammogram because the nipple (N) is located very low in the image, the subareolar fibroglandular tissue (asterisk) is not well separated, and there is a prominent skin fold (arrow) at the posterior inferior aspect of the breast. b) Same breast properly positioned. In this mammogram, the nipple is higher in the image, the subareolar tissues are well separated, and the inframammary fold (arrow) is open. More of the pectoralis muscle is shown as well. ©Lippincott Williams & Wilkins.

the analysis of the images. Currently, in Sweden, only screen-film combinations are used for image recording in screening mammography. Consequently, researchers of computer-aided diagnosis (CAD) use digitised mammograms almost exclusively, see Section 3.8. However, developments in full-field digital mammography (FFDM) are in clinical use and will yield the opportunity to perform CAD on FFDM.

Digital image acquisition (film digitisation or FFDM) can vary in terms of spatial resolution and number of quantisation levels. Since film-based computerised image analysis methods may be dependent on the digitiser, digitisation as well as image analysis approaches should be treated as important subcomponents of a computer-aided diagnosis system. For example, as further described in Section 3.8, it is important to use a database with digital images, from similar acquisition sources, while comparing computerised image analysis algorithms.

There are several types of detectors used for digital image acquisition. However, basically, only two of the available detectors are useful for digital mammography based on (D’Orsi and Newell, 2007a). The first is the well known digital detector, the Charge Coupled Devices (CCD), which has been used in
20 Towards Automatic Image Analysis for Computerised Mammography

Figure 2.8: In this mammogram, the breast is properly positioned (no droopiness of the breast), the pectoralis muscle is well visualised to the level of the nipple, and exposure is correct. Unfortunately, a skin fold posterior to the nipple may obstruct diagnosis (conceal calcifications or breast architecture). For this reason, the breast has to be re-examined by taking a new image. ©Lippincott Williams & Wilkins.

early experimental versions of FFDM and for digital stereotactic breast biopsy. According to the authors, they deliver very good spatial resolution (6 to 12 lp/mm) with excellent contrast properties and dynamic range. However, the limitation with these detectors is that they can only cover an area of about $6 \times 6 \text{ cm}^2$. A second type of detector is the flat panel. Depending on their construction, flat panels may be used either for direct or for indirect image capture. In the indirect capture design, a scintillator (usually cesium iodide) produces a light image that is directed to small photo diodes, producing the digital or electronic image. Direct capture design uses amorphous selenium, which eliminates the intermediate scintillator layer enabling these detectors to attain excellent spatial resolution (D’Orsi and Newell, 2007a).

2.4.7 Digital Mammography

Digital mammography is a technique in which the screen-film image receptor is replaced by an x-ray detector which detects x-ray photons. The photon energy is converted into an electronic signal, which is then digitised and stored as a computer image. The operations of image acquisition, display and storage are decoupled and can thereby be optimised independently without the
compromises inherent in screen-film imaging, where increased contrast necessitates a reduction in exposure latitude (Yaffe, 1996). The detector can have linear response over a wide dynamic range, and it can make extremely precise measurements of the transmitted x-ray intensity. The image information, once displayed, can be transformed and presented such that relevant anatomy for breast cancer detection or diagnosis is visualised in an optimal manner.

Over the last 20 years, tremendous advances in the technology and technique of digital mammography have taken place. Already in 1996, Yaffe (1996) pointed out that there had been a remarkable improvement in spatial resolution, contrast and the amount of breast tissue imaged. At the same time, techniques for positioning the breast and exposing the image had been improved and training in the interpretation of digital mammograms had become better.

Recently, digital systems for clinical mammography have become available on a commercial basis. Full field digital mammography (FFDM) offers potential improvements over the limitations of SFM (D’Orsi and Newell, 2007a; Bick and Diekmann, 2007; Turco et al., 2007; Skaane et al., 2007; Pisano et al., 2008). While a film acts both as the detector and as the display medium for the breast, the digital technique can separate these two functions with the possibility of maximising the performance of each independently.

The remarkable advances in the technology for digital mammography and the systems are, now in 2008, sufficient for mammographic examination. It appears to perform significantly better for women who are pre- or peri-menopausal with dense breasts (Pisano et al., 2008; Skaane et al., 2007). However, it is important to keep in mind several aspects while turning from screen film mammography to digital mammography. The expensiveness of the units and the initial difficulty of comparing digital to prior analog examinations are some aspects (D’Orsi and Newell, 2007a). Furthermore, when considering softcopy read, LCDs seem to offer the best compromise with regard to detection and size but one must ensure that there are at least 1010 bits of grey displayed (D’Orsi and Newell, 2007a). Thus, storage is also a consideration and should be considered before implementation of digital mammography (D’Orsi and Newell, 2007a). Assume that a facility performs 5000 screening examinations per year. This may require up to 1 terabyte of storage space. Digital mammography is still in its infancy and many of the issues raised by studying clinical trials will be addressed and rectified in the near future (D’Orsi and Newell, 2007a; Pisano et al., 2008; Skaane et al., 2007).

2.5 Quality Criteria Concerning Breast Positioning

Mammography positioning of the breast is an art; if the breast is incorrectly positioned, the radiologist cannot evaluate the mammogram properly. The mediolateral oblique view and the craniocaudal view (Figure 2.9), used in combination, have become international standard (Heywang-Köbrunner et al., 1997;
Figure 2.9: a) The mediolateral oblique (MLO) view. b) The craniocaudal (CC) view, both taken from (Georgsson, 2001).

Perry et al., 2001). The aim with these standard views is to permit reliable information if there are malignant processes. However, additional views should be used whenever a standard view is not adequate or does not sufficiently visualise the areas of interest. The major goal in positioning a patient for a screening examination is to show all breast tissue on this combined two-view examination. At some occasions, so-called single view mammography, which basically consists of the MLO view, is sufficient. However, most radiologists do not regard the single view enough for a complete diagnosis (Heywang-Köbrunner et al., 1997).

2.5.1 The Mediolateral Oblique View

Since the breast lies primarily on the pectoralis major muscle, a generous amount of the muscle should be included in the image. When properly obtained, the mediolateral oblique (MLO) view offers the best opportunity to show all breast tissue in one image (Figure 2.10(a)), since it best visualises the tissue adjacent to the chest wall and the axillary tail. This is of particular importance, because the majority of cancers occur in the axillary tail (Kimme-Smith et al., 1997).

When conducting the examination, more muscle can be included if the x-ray tube is rotated so that the cassette holder (also called bucky or positioning table) is parallel to the plane of the pectoralis muscle (Figure 2.11). This can be accomplished with a 30–70° inclination depending on the body posture. The mobile lateral border of the breast is moved as far toward the fixed medial border as possible before placing the breast on the cassette holder. The x-ray beam will go from medial and superior to lateral and inferior, hitting the image detector system perpendicularly. By positioning the cassette and compression paddles in line with the pectoralis muscle in this way, it allows optimal mobilisation of the glandular tissue away from the chest wall.

The visibility and the form of the pectoralis muscle in a mammogram is one
Figure 2.10: a) Right and left MLO view. b) Right and left CC view. These views reveal a cancer (arrows) deep in the medial aspect of the breast. This small nonpalpable cancer was discovered only because a generous amount of pectoralis muscle was included in the MLO view (a). As we can see, it would also have been missed in the CC view (b) if this view had not been properly acquired. ©Lippincott Williams & Wilkins.

Figure 2.11: Bilateral mediolateral views. In these mammograms, a proper amount of pectoralis muscles has been included in the images, and they are triangular in shape with slightly anterior convexity. There is, however, a skin fold (arrow) in the axilla. According to Kimme-Smith et al. (1997), this a common finding that does not interfere with image interpretation. ©Lippincott Williams & Wilkins.
Figure 2.12: Quality criteria for the mediolateral oblique view. a) Inadequate mammogram. b) Adequate mammogram. The quality criteria for a good MLO view are (Carlson, 2002; Eklund et al., 1994; Heywang-Köbrunner et al., 1997; Perry et al., 2001): The pectoralis muscle must be shown at least down to nipple level (1), The pectoralis muscle should course superiorly along the lateral border of the image to an angle of about 20° (2), The nipple must be in profile (3), The inframammary fold must be included, which is important because the inclusion indicates sufficient visualisation of the medial glandular tissue (4), The glandular tissue should be well spread (5), Retroglandular fat should be visible between the posterior glandular margins of the back of the image (6), The entire breast must be shown. The upper glandular tissue projecting over the pectoralis muscle should not be “cut off” at the back of the image (7) and The pectoralis muscle should have a convex anterior margin (8). The outlines of the two mammograms taken, one on each breast, should be symmetrical.
Figure 2.13: These two mammograms are examples of improper positioning. In the left mammogram, not enough axilla is displayed. The pectoralis muscle is shown only down to the top of the glandular tissue. As a consequence, a small mass with calcification in posterior aspects of the breast is missed, whereas it can be seen in the properly positioned mammogram on the right (arrow). ©Lippincott Williams & Wilkins.

of the most important criteria considering the assessment of the image quality. The reason is that the entire tissue will not be viewed properly in the MLO view if the pectoralis muscle is tighten. Since mammography is a painful procedure for the woman the muscle may be tense. This may result in an incorrect MLO view and the radiologist will not be able to judge the image properly. The muscle should be as relaxed as possible, evidenced by a generous amount of muscle on the image. A correct MLO view will visualise the pectoralis muscle as a triangular shape with a convex anterior margin (Figure 2.11).

The inclusion of the retroglanular fat posterior to all glandular tissue is another important criterion for the quality of the mammogram. If the glandular tissue extends all the way to the posterior edge of the film, it must be assumed that some glandular tissue is excluded from the image.

Proper positioning methods are required during the initiation and application of breast compression for the MLO. The breast will fall or “droop” if it is not held up by the radiographer until there is adequate pressure by the compression device to maintain an upright position of the breast. Droop of the breast is witnessed by a low position of the nipple on the image, a superimposition of subareolar tissues, and a prominent skin fold near the inframammary fold (Figure 2.7(a)). The quality criteria for a good MLO view are summarised in Figure 2.12. Figure 2.13 visualises a possible consequence of improper positioning.
Figure 2.14: Bilateral CC views. In these mammograms, the lateral part of the breast is at the top of the images. We also see the pectoral muscle (arrowhead) along the posterior medial parts of the image. The accessory muscle, called the sternalis muscle (arrow), is included, and all medial fibroglandular tissue (asterisk) has been imaged. ©Lippincott Williams & Wilkins.

2.5.2 The Craniocaudal View

The craniocaudal (CC) view is a routine supplement to the MLO view serving two important purposes. First, the CC can more effectively visualise the medial glandular tissue. Since this is the portion of the breast that sometimes is not visualised in the MLO view at all, an important objective when positioning the breast for the CC is to include all medial tissue (Figure 2.14). Second, the CC can often depict structures more clearly because more compression is possible, and motion blur is uncommon because the breast is resting directly on the cassette holder. In the CC view, the beam travels from superior to inferior.

In mammography, the radiologist uses the posterior line as a measurement to decide if the entire breast is imaged in the craniocaudal view (CC). The posterior line for the CC view is defined as the distance from the nipple-skin junction to the back of the image. The posterior line for the MLO view is the distance along the nipple axis perpendicular toward the pectoralis muscle. According to Eklund et al. (1994), when a proper mobilisation and positioning of the breast are obtained, the posterior nipple line on the MLO view should not exceed that on the CC view by more than 1 cm (Eklund et al., 1994). This quality criterion requires a comparison between the two standard views, MLO and the CC (Figure 2.15). It is often necessary to combine the information from the mediolateral oblique and craniocaudal mammograms, not only in quality assessment analysis but also in lesions detection. The quality criteria for a good CC view are summarised in Figure 2.16.
Figure 2.15: The posterior nipple line outlined in both the MLO (a) and the CC (b) views. For a properly positioned CC view, the pectoralis muscle (white arrow) should be shown at the posterior aspect of the breast (if possible) to confirm the inclusion of deep posterior breast tissue. If the pectoralis muscle is absent, the posterior nipple line is used to document an adequate imaging of posterior tissue. Therefore, one of the quality criteria is that the difference between the posterior nipple line in the MLO view and the posterior nipple line in the CC view should be at least 1 cm. ©Lippincott Williams & Wilkins.

Figure 2.16: Quality criteria for a good CC view. The quality criteria for a good CC view are (Carlson, 2002; Eklund et al., 1994; Heywang-Köbrunner et al., 1997; Perry et al., 2001): The medial fold of the breast is shown (1), as much as possible of the lateral breast tissue is shown (2), if possible, the pectoralis muscle shadow is just barely shown on the posterior edge of the breast (3), the difference between the posterior nipple line in the MLO view and the posterior nipple line in the CC view should be at least 1 cm (4), the nipple should be in profile (5) and the entire body of the glandular tissue (6) with the retroglandular fat should be included (7). The outlines of the two mammograms taken, one on each breast, should be symmetrical.
2.6 Mammographic Interpretations

As seen in Section 2.1, connecting the superficial and deep pectoral fasciae, Cooper’s ligaments provide a skeleton along which breast tissue is loosely supported, see Figure 2.1. Within a honeycomb-like pattern, variable amounts of fatty and glandular tissue are distributed along Cooper’s ligaments, giving the breast tissue a characteristically scalloped appearance on mammograms (Cardenosa, 1996). Therefore, if there is no history of trauma or surgery, straight lines and convex contours are usually not a characteristic of healthy breast tissue. Hence, such structures can indicate an underlying breast cancer (Cardenosa, 1996). In most women, the medial half of the breast is fatty on cranio-caudal views and any perceived parenchymal asymmetry or tissue protruding into the fat demands attention. The retro-glandular fat, and in particular the interface between glandular and fatty tissues, in both the CC and MLO views, requires careful assessment, as well. One suspicious irregularity is if the anterior boundary between the glandular tissue and subcutaneous fat on the cranio-caudal and mediolateral oblique views has a typically scalloped appearance. If the harmonious appearance of the scalloped pattern is disrupted by tissue bulging into the subcutaneous fat, spot compression and tangential views should be obtained (Cardenosa, 1996). Correlative physical examination in these patients can be helpful in confirming the presence of palpable abnormalities in these superficial areas of contour irregularity.

Regions such as the area closely behind the nipple, the medial half of the breast on cranio-caudal views, and the retro-glandular area on both the mediolateral and cranio-caudal views (especially the glandular/fat boundary) deserve particularly close attention. If a potential problem is perceived in these, or any other areas of a screening study, additional evaluation, including spot compression, micro-focal spot (0.1 mm) magnification, rolled or tangential views may be indicated to examine and evaluate a possible lesion appropriately. Ultrasoundography can be used to characterise masses as either cystic or solid, to evaluate palpable abnormalities or to guide interventional procedures including cyst aspirations, fine needle aspirations, or core biopsies (Cardenosa, 1996). According to Cardenosa (1996), the second most common location for breast cancer is the subareolar area. This area can be difficult to evaluate because of the confluence of ductal structures and, in some patients, superimposition of the nipple. Furthermore, in a significant number of women, the subareolar area remains relatively undercompressed in routine MLO and CC views due to the limitation of compression by the thicker, chest wall portion of the breast. If there is inadequate separation of subareolar structures, the problem can be solved with full paddle anterior compression views or spot compression of the subareolar area. These modalities can be used to differentiate superimposed normal anatomical structures from significant abnormalities.
Chapter 3

Computer-Aided Analysis in Mammography

Since quality assessment of mammograms tries to estimate whether the quality of a given mammogram is good enough to allow for a reliable analysis, it is essential to gain a better understanding of the problems involved in computer-aided analysis of mammograms. Identifying and evaluating links between this area and related ones are another important prerequisite for the research to be successful. Therefore, a survey of literature within mammographic image analysis was performed and constitutes parts of this chapter. In order to give a better understanding, some of the basic image analysis techniques are elucidated. Furthermore, image segmentation and feature extraction techniques, which are essential in computer-aided mammography, will be briefly explained.

3.1 Computerised Analysis of Mammograms

The development of analysis algorithms in mammography requires knowledge regarding automatic extraction of features from a digital image, about medical images (i.e. mammograms), and about various computer processing techniques. The required a priori knowledge includes properties of the digital image acquisition system and morphological quality of the abnormality (e.g. mass lesion or cluster of microcalcifications) along with its associated anatomic background. An adequate database is needed in order to cover the entire range of abnormals and normals.

The aim of developing computer-aided diagnosis (CAD) is to overcome some of the limitations of mammography. The term CAD system covers two different types of systems (Nishikawa, 2007): computer-aided detection (CAde), with the purpose to aid radiologists find breast cancer on screening mammograms, and computer-aided diagnosis (CADx) with the purpose to support radiologists in the decision whether a known lesion on diagnostic mammograms is benign or malignant.
3.2 Computer-Aided Detection

Most computerised schemes in mammography are developed for the detection of mass lesions or clustered microcalcifications. The computer-aided detection (CADe) of lesions requires having the computer locate suspicious regions, but leaves the subsequent classification of the lesion entirely to the radiologist.

The CADe of mass lesions, which includes spiculated lesions, asymmetries, developing densities, and circumscribed lesions as well as parenchymal distortions in digital mammograms is a more difficult task than that of detecting microcalcifications, because of the similarity of many mass lesions and the surrounding normal parenchymal tissue.

The goal of CADe is not the detection of cancer. The goal is to help radiologists avoid overlooking a cancer that is visible in a mammogram. According to Nishikawa (2007) CADe schemes have high sensitivity, but poor specificity compared to radiologists. CADe has been shown in both observer studies and clinical evaluations to help radiologists find more cancers. Recent clinical studies indicate that CADe increases the number of cancers detected by approximately 10%, which is comparable to double reading by two radiologists (Nishikawa, 2007).

3.3 Computer-Aided Diagnosis

Since computer-aided diagnosis (CADx) systems are developed to aid in distinguishing between malignant and benign lesions, the systems need to improve both the sensitivity and specificity of mammography in order to be successful for clinical use. CADx methods may use features extracted either by the computer or the radiologist. Radiologists employ many radiographic image features that they seem to extract and interpret simultaneously and instantaneously. The development of methods using computer analysis techniques requires the determination of which individual features are clinically significant and then extract each of these features.

One of the aims of computerised classification is to reduce the number of women undergoing biopsy for benign disease. Since the price of a missed cancer is much higher than that of a misclassification of benign findings, such a reduction will be clinically acceptable only if it does not result in malignant cases going unbiopsied (Giger et al., 2000). Therefore, the systems developed should improve specificity but not at the loss of sensitivity.

Various investigators have used computers to aid in the decision-making process regarding probability of malignancy and patient management, using features which were extracted from the images by radiologists. Such methods are dependent on the subjective identification and interpretation of the personnel within the mammographic facility. The interpretation of a mammogram is inherently variable because the mammograms are read by human beings. There is both inter- and intra-variability among radiologists (Nishikawa, 2007). Furthermore, there are studies showing considerable differences in performance
between radiologists in Europe and in North America. In the USA, for instance, the positive predictive value (PPV) for diagnostic breast imaging is generally less than 50% (Nishikawa, 2007). The PPV measures the percentage of all breast biopsies that are positive for cancer. These results indicate that there is a need to develop methods based on direct computer extraction of features to increase the objectivity and reproducibility of the result.

One of the earliest aims of computer analysis systems was to perform as well as an expert radiologist. Since then, researchers have succeeded in extracting a variety of clinically relevant features and merging them into an estimate of the likelihood of malignancy, with observer studies showing improvement when the aid is used (Jiang and Metz, 2006).

### 3.4 Image Analysis

In computerised mammography, there are several image analysis tasks involved. This section will give a review of the techniques used in this area and the papers of this thesis. First the involved terms and techniques will be described and later on the image analysis problem solving task for computerised mammography will be discussed.

#### 3.4.1 Terms and Techniques

The fundamental steps in digital image processing can be divided into low level, intermediate level, and high level image processing. Each level is characterised by its objectives and the techniques that are used (Figure 3.1). For example, discretization and pre-processing are objectives of low level image processing. Discretization is the image acquisition process, which yields the digital images to work with. Preprocessing adapts the image to a specific application. It involves image enhancement and image restoration. The basic ideas behind both image enhancement and image restoration are to bring out details that are obscured, or simply to highlight certain features of interest in an image. Image enhancement aims at subjective improvements, i.e. to make the image “look better”, whereas image restoration deals with improving the image in a more objective manner by using mathematical or probabilistic models (Gonzalez and Woods, 2008).

The intermediate level includes segmentation, representation and description. Segmentation partitions an image into its constituent parts or objects, which makes it possible to talk about objects in the image. It involves techniques such as morphological processing, which is a tool for extracting image components that are useful in the representation and description of an object in the image. Representation and description make it possible to talk about the properties of the extracted objects. Finally, the high level image processing involves interpretation. Interpretation makes it possible to talk about the contents of the image by classifying objects.
Digitising Images

Computers cannot handle analog images, and thus such images have to be converted into digital form. Digitising images involves certain steps. The first step is scanning, in which the spatial locations have to be selectively addressed with a grid within the domain of an image. The resulting small subregions are called picture elements or pixels. The next step is called sampling. It measures the intensity of an image at the location of each picture element. After the scanning and sampling steps, the measured intensities have to be quantified to distinct degrees of intensity. This step is called quantisation.

Histogram

One simple and useful tool in digital image processing is the intensity histogram. The intensity histogram of a digital image is a discrete function summarising the grey level in the range \([0, L - 1]\) of an image, where \(L\) is the number of intensity levels. The histogram is defined as the function \(h(r_k) = n_k\), where \(r_k\) denotes the \(k\)th intensity and \(n_k\) is the number of pixels in the image that have that specific intensity \(r_k\) (Gonzalez and Woods, 2008). If the area under the histogram is normalised, that is, each of its values is divided by the total number of pixels in the image, denoted by \(n\), the histogram can be viewed as a probability function \(P(r_k) = n_k/n\) for \(k = 0, 1, \ldots, L - 1\). In other words, \(P(r_k)\) is the probability that a randomly picked pixel has the intensity \(r_k\).

Histograms are the basis for numerous spatial domain processing techniques,
and histogram manipulation can be used effectively for image enhancement. In addition to providing useful image statistics, the information inherent in histograms is quite useful in other image processing applications, such as for instance segmentation.

The histogram gives a visual indication as to whether or not an image is properly scaled within the available range of grey levels. A digital image should make use of all of the available intensities. It is almost always an information loss to discretize a continuous image into a digital image. However, if 256 (8-bit image) intensities are available and the image is digitised to fewer than 256 intensities, more information than necessary will be lost. That lost information cannot be restored without redigitising the continuous image. If the image has a brighter range than the digitiser is set to handle, then the histogram can give information during the digitising process, so that problems can be avoided.

**Operations Used in Image Processing**

**Point Operation**  Point operations modify the intensity histogram of an image by a pixel-by-pixel copying operation, except that the intensities are modified according to the specific intensity transformation function $\gamma$. A point operation that takes an input image $f(x, y)$ into an output $g(x, y)$ may be expressed as

$$ g(x, y) = \gamma[f(x, y)]. $$ \hfill (3.1)

**Usefulness of Point Operations**  It can be useful to represent the images in a consistent format prior to visual comparison or segmentation. *Histogram equalisation* enhances the contrast of an image by transforming the grey value in an intensity image (Castleman, 1996). The intensity values of the input image are transformed by a point operation such that the histogram of the output image is approximately flat. This is useful for, e.g. segmenting images or comparing images acquired under different imaging conditions. If an image has low contrast with most values in the middle of the intensity range, histogram equalisation can be used to evenly distribute the values throughout the range of all available intensity levels.

**Algebraic Operations**  Algebraic operations are useful operations that are performed on two input images and give one output image. Since an image is often represented as a matrix, the sum, difference, product or quotient of the two input images is often defined pixel-wise (Gonzalez and Woods, 2008):

$$ h(x, y) = f(x, y) \odot g(x, y), $$ \hfill (3.2)

where $f(x, y)$ and $g(x, y)$ are the input images, $h(x, y)$ is the output image and $\odot$ is one of $+, -, \cdot, /$, denoting addition, subtraction, multiplication, and division, respectively. The output $h$ is not necessarily of the same type (for instance 8–bit
image) as \( f \) and \( g \). For this reason \( h \) may need to be scaled and re-quantified to keep the grey levels within range.

**Usefulness of Algebraic Operations** Image addition is useful for adding together multiple images of the same scene by using proper normalisation. This can be used to reduce effects of additive random noise with zero mean (Castleman, 1996).

Image subtraction can be used to remove an undesired additive pattern from an image, that is, a pattern such as slowly varying background shading or periodic noise. Subtraction can also be useful in detecting changes between two images of a scene. A closely related application is the computation of an approximation of the gradient, which is a useful function for locating edges.

Multiplication and division can be used to correct for the effects of a digitiser in which the sensitivity of the light sensor varies from point to point (Castleman, 1996). Multiplication by a mask image is useful if it is desired to block out certain areas of an image, leaving only the area of interest.

**Geometric Operation** Geometric operations change the spatial relationships among the picture elements in an image. Such operations may be thought of as printing the image on a rubber sheet, stretching the rubber sheet and tracking it down at various points. In general, it is possible to transform one point to any other arbitrary point but that would scramble the image content. Thus, the geometric operations have to be constrained (Castleman, 1996).

**Gradient Magnitude** The rate of change of the grey value yields an important method for differentiation in image processing (Gonzalez and Woods, 2008). It is particularly useful for edge detection, where a sudden variation in grey value indicates a border.

The gradient of an image is defined as follows: Given an image \( f(x, y) \) and a coordinate with unit vectors \( \mathbf{i} \) in the \( x \)-direction and \( \mathbf{j} \) in the \( y \)-direction, the gradient is the vector function

\[
\Delta f(x, y) = i \frac{\partial f(x, y)}{\partial x} + j \frac{\partial f(x, y)}{\partial y}.
\]

(3.3)

It is known from vector analysis that the gradient vector points in the direction of the maximum rate of change of \( f \) at \((x, y)\) and the magnitude (length) of the vector \( \Delta f(x, y) \) is equal to the value of the slope. The gradient magnitude is given by:

\[
|\Delta f(x, y)| = \sqrt{\left( \frac{\partial f}{\partial x} \right)^2 + \left( \frac{\partial f}{\partial y} \right)^2}.
\]

(3.4)

This represents the steepness of the slope at every given point. Its direction is calculated as follows:
\[ \theta = \tan^{-1} \left( \frac{\frac{\partial f}{\partial y}}{\frac{\partial f}{\partial x}} \right) \] (3.5)

Convolution Filtering of images is useful to accomplish particular goals. Linear filtering of images is implemented through two-dimensional convolution. A convolution of the digital image, i.e. with integer pixel coordinates becomes a double summation:

\[ h(i, j) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n)g(i - m, j - n) \] (3.6)

for \( i = 0, 1, 2, \ldots, M - 1 \) and \( j = 0, 1, 2, \ldots, N - 1 \). The image is denoted \( f \) and \( g \) is the filter kernel. The resulting image after \( f \) is convolved with the filter kernel \( g \) is \( h \).

Fourier Transform

A powerful tool within image processing is the Fourier transform, which can be used to quantify the effects of digitising systems, sampling spots, electronic amplifiers, convolution filters, noise, and display spots (Castleman, 1996). The image processing analyst uses the Fourier transform to move back and forth between the spatial domain and the frequency domain while analysing problems. The forward transform and the inverse transform constitute a transform pair (the Fourier transform is completely reversible). Each Fourier transform coefficient depends on all spatial sample values (pixels).

Working with the Fourier transform on a computer usually involves a variant of the transform known as the discrete Fourier transform (DFT). The DFT is defined for a discrete function \( f(m, n) \) that is nonzero only over the finite region given by \( 0 \leq m \leq M - 1 \) and \( 0 \leq n \leq N - 1 \). The two-dimensional DFT and inverse DFT are given as follows (Gonzalez and Woods, 2008):

\[
\begin{align*}
F(p, q) & = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n)e^{-j(2\pi/M)m}e^{-j(2\pi/N)n} \quad \{ \begin{array}{c} p = 0, 1, \ldots, M - 1 \\ q = 0, 1, \ldots, N - 1 \end{array} \\
& \sum_{p=0}^{M-1} \sum_{q=0}^{N-1} F(p, q)e^{j(2\pi/M)p}e^{j(2\pi/N)q} \quad \{ \begin{array}{c} m = 0, 1, \ldots, M - 1 \\ n = 0, 1, \ldots, N - 1 \end{array} 
\end{align*}
\] (3.7)

\[
\begin{align*}
f(m, n) & = \frac{1}{MN} \sum_{p=0}^{M-1} \sum_{q=0}^{N-1} F(p, q)e^{j(2\pi/M)p}e^{j(2\pi/N)q} \quad \{ \begin{array}{c} m = 0, 1, \ldots, M - 1 \\ n = 0, 1, \ldots, N - 1 \end{array} 
\end{align*}
\] (3.8)

The values \( F(q, p) \) are called the DFT coefficients of \( f(m, n) \). The discrete Fourier transform produces the same number of Fourier transform coefficients as the number of pixels in the image. It should be noted that each Fourier coefficient is a complex number.
The Convolution Theorem  It can be shown that if $F$ and $G$ are the Fourier transform of $f$ and $g$, respectively, the following holds:

$$f * g \Leftrightarrow FG$$  \hspace{1cm} (3.9)

$$F * G \Leftrightarrow fg$$  \hspace{1cm} (3.10)

This means that the convolution in the spatial domain reduces to multiplication in the frequency domain and the convolution in the frequency domain reduces to multiplication in the spatial domain. This is an important property since it, for instance, allows for developing filters in the frequency domain while applying them (or at least an approximations of them) in the spatial domain.

Filter

A filter may be implemented by convolving the image by an appropriate convolution kernel ($g$ in Equation 3.6). There are several useful filters within image processing that are commonly used. For instance, a filter can be constructed to transform one image into another with certain properties changed. This transformation is done by simply computing the Fourier transform of the image to be enhanced ($F$ in Equation 3.10), multiplying the result by a filter transfer function ($G$ in Equation 3.10), and take the inverse Fourier transform to produce the enhanced image. Some kinds of filters presented within the literature concerning image processing are, lowpass, highpass and bandpass (Castleman, 1996).

Lowpass filters reduce the energy for the high frequencies. In general, an image has a majority of its energy in the low- and mid-frequency range of its frequency spectrum. Noise and other sharp transitions contribute to the high-frequency content of its Fourier transform. The information of interest is often buried by noise. A lowpass filter can then be used to reduce the amplitude of high-frequency components and thus the visible effect of noise.

Highpass filters reduce the energy for low frequencies. Since abrupt changes in intensities are associated with high frequency components, image sharpening can be achieved in the frequency domain by a highpass filter process. If there are small details with high frequency these can be detected much easier after the image is convolved with a highpass filter. The highpass filter process reduces the low frequencies without disturbing the high frequency information.

Bandpass filters reduce the energy for specific frequencies. In some cases, it is desirable to eliminate only some frequency ranges of the spectrum. A bandpass filter can then be used to let only remaining frequencies pass. Bandpass filters combine the noise reduction of lowpass filters with the detail enhancement of highpass filters.

There are other useful filters such as the Sobel filter. A Sobel filter is an edge-emphasising filter which is used very often when edge detection is desired. This is explained more thoroughly in chapter 3.4.2.
3.4.2 Pattern Recognition

In order to be able to analyse a mammogram with computer equipment, the different features within the mammogram have to be extracted, represented, and measured, and it must be found out what pattern class the objects belong to. These three steps are, within the computer vision field, called pattern recognition (Castleman, 1996), see Figure 3.2. Consider the problem of analysing the pectoralis muscle within a mammogram. The digital image contains several objects. The recognition involves three phases, namely (1) image segmentation, (2) feature extraction, and (3) classification. The first two are also called object isolation and object measurement, respectively.

Image Segmentation

Castleman (1996) defines image segmentation as a process that partitions a digital image into disjoint (non-overlapping) regions. The level of this subdivision into disjoint regions depends on the problem being solved. When the object of interest has been isolated, the segmentation should stop. Image segmentation is a difficult task within image processing, and the analysis of the image is highly affected by the success of this process. A region can be seen as a connected set of pixels. The definition of connectedness relies on the notion of adjacency used. Gonzalez and Woods (2008) discuss: 4-adjacency and 8-adjacency, which are based on the neighbourhood of pixels (given that the sampling has been carried out with a square lattice):

- A pixel $p$ at coordinate $(x, y)$ has four horizontal and vertical neighbours, namely $(x + 1, y)$, $(x - 1, y)$, $(x, y + 1)$, and $(x, y - 1)$.

- The diagonal neighbours of $p$ are $(x + 1, y + 1)$, $(x + 1, y - 1)$, $(x - 1, y + 1)$, and $(x - 1, y - 1)$. Together with the horizontal and vertical neighbours these are the 8-neighbours of $p$.

A connected set is a set such that, between each pair of pixels, there is a path that does not leave the set. A path from $p$ to $q$ is defined as a finite
sequence of pixels $p_0, p_1, \ldots, p_n$, where $p_i$ and $p_{i+1}$ are adjacent and $p_0 = p$ and $p_n = q$ (Gonzalez and Woods, 2008). If only laterally adjacent pixels are connected, the objects are four-connected. If the diagonally adjacent pixels also are considered, the objects are eight-connected.

General image segmentation algorithms for intensity images are based on one or two basic properties of intensity values: discontinuity and similarity. The first approach is to partition an image based on abrupt changes in intensity, where the principles areas of interest are the detection of the boundaries between objects within an image. The second is a region approach, which assigns each pixel to a particular object (Gonzalez and Woods, 2008).

**Image Segmentation by Thresholding**  Thresholding is a particular region-approach technique for images depicting solid objects resting upon a contrast background. The thresholding technique for image segmentation works in the following way. All pixels at or above the threshold intensity are assigned to the object, assuming that the object is brighter than the background, whereas all pixels having a intensity below the threshold fall outside the object. Thresholding works well if the object of interest has a uniform intensity and rests upon a background of different, but uniform, intensity. If the object differs not only in background intensity but also in its internal texture, the image can be pre-processed before being segmented by thresholding.

The simplest implementation of boundary location by thresholding is when the value of the threshold intensity is held constant throughout the image. In many cases, though, the background intensity is not constant and the contrast of objects varies within the image. This is due to several properties such as noise, uneven lightning, and uneven intensity of the object and the background, which, means that we cannot assume that there are only two different intensities. In those images a threshold works well in one area but works poorer in others. In such cases, it is convenient to use a threshold intensity that is a slowly varying function of the position in the image.

However, suppose for simplicity that there is one object and the background present in the image. Then we can say that the intensity of the object is given by the probability density function $p_1(z)$ and the intensity of the background is given by $p_2(z)$. The a-priori probability for a pixel to belong to the object is $P_1$ and the probability for a pixel to belong the background is $P_2$. Since a pixel either belongs to the background or the object it must hold that $P_1 + P_2 = 1$. We say that we have a mixture model, see Figure 3.3:

$$p(z) = P_1p_1(z) + P_2p_2(z).$$

The histogram of the image can be seen as an estimate of this model. Suppose that we establish a single threshold $T$ and say that pixels with intensity higher than $T$ belong to the object and pixels with intensity lower than $T$ belong to the background. Then the error of using this threshold is:

$$E(T) = P_2E_1(T) + P_1E_2(T)$$
where
\[ E_1(T) = \int_{-\infty}^{T} p_2(z)dz \]
and
\[ E_2(T) = \int_{T}^{\infty} p_1(z)dz \]

Obviously, we want to find the \( T \) that yields the least error. To find this \( T \) we solve \( \frac{dE}{dT} = 0 \) and obtains
\[ P_1p_1(T) = P_2p_2(T). \]

The observable error against the histogram \( h(z) \) is
\[ e = \frac{1}{n} \sum_{i=1}^{n} (p(z_i) - h(z_i))^2. \]

Without making any assumptions, it is very hard to minimise \( e \). This can be handled in two ways: On the one hand, we may assume that we know the form of the distributions \( p_1(z) \) and \( p_2(z) \) and solve the equation analytically. On the other hand, we may solve \( \frac{dE}{dT} = 0 \) for \( T \) by iterative numerical methods (Gonzalez and Woods, 2008).

An image containing an object on a contrasting background has a bimodal intensity histogram if the differences in intensities are large enough (Figure 3.3). The dip between the peaks corresponds to the relatively few points around the edge of the object. The dip in such a histogram is likely to be the proper threshold for that particular image. The histogram is commonly used to establish the threshold intensity for cases like this. For more challenging images, where the object may be irregular and create a noisy histogram, other methods should be used. One such method, similar to the one just described, is to estimate a
suboptimal threshold based on assuming some probability distribution of the pixels belonging to the background and the pixels belonging to the object, and find the intensity value that best separates these two. Based on this threshold, the region belonging to the background can be modelled and processed so that it becomes more uniform and will separate more from the object. In this way, a threshold closer to the optimum might be found. Another approach is to start with a suboptimal threshold and then trace an intensity profile from the object to the background until a more reasonable threshold for the specific task is found.

There are several other image segmentation methods as well. One such is a boundary segmentation method. The thresholding region approach is based on partitioning the image into sets of interior and exterior points while boundary approaches attempt to find the edges directly by their high gradient magnitudes.

3.4.3 Edge Detection and Linking

An edge is the boundary between two regions with relatively distinct intensity properties (Gonzalez and Woods, 2008). After location of the pixels on the boundary of an object, the pixels exhibiting the required characteristics are labelled edge points. This can be displayed as an edge image showing how strongly each corresponding pixel meets the requirements of an edge pixel. In these edge images, the objects are outlined in edge points and are seldom the closed connected boundaries that an image segmentation process requires. Thus, the next step necessary before the objects are completely extracted is an edge linking process (Castleman, 1996).

Edge Detection

A pixel that lies on the boundary of an object within an image has a neighbourhood, which will be a zone of grey level transition. Edge detection operators examine the characteristics of that transition, the slope and the direction. These characteristics yield the magnitude and the direction, respectively, of the gradient vector. There are several ways to calculate this information. One of them is based upon convolution with a set of directional derivative masks (kernels). Some of the useful masks are described below.

The Sobel Operator

In Section 3.4.1, the concept of using the gradient vector for differentiation was explained. The Sobel operator performs a 2D spatial gradient measurement in an image and thus emphasises regions of high spatial gradient, i.e. regions that correspond to edges. Each point in the image is convolved with two $3 \times 3$ kernels. One kernel responds maximally to vertical edges and the other to horizontal edges. In this way, the absolute gradient magnitude at each point in an input intensity image is found. The result is an edge magnitude image (Castleman, 1996). The kernel used for detection of the vertical edges is
The transposed kernel used for horizontal edge detection is

$$S_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}. \quad (3.12)$$

The gradient magnitude is given by (Gonzalez and Woods, 2008)

$$|S| = \sqrt{S_x^2 + S_y^2}. \quad (3.13)$$

The Sobel operator has the advantage of providing both a differencing and a smoothing effect, because $S_x$ and $S_y$ actually correspond to bandpass filters based on the convolution of a highpass filter $h$ (derivative kernel) and a lowpass filter $l$ (smoothing kernel). If $h = [1, -1]$ and $l = [1, 1]$, the combination $h * l$ is a slightly noise reducing filter. The kernel $S_x$ will be the result of the convolution

$$[h * l] * [l * l]^T = S_x. \quad (3.14)$$

Since derivatives enhance noise, the smoothing effect is a particularly attractive feature of the Sobel operator. As shown above, for complete edge detection filtering, $S_x$ is simply rotated by 90°, yielding $S_y$, and the gradient information obtained from them is combined.

### 3.4.4 Edge Linking

The edge images described above will have gaps that need to be filled. Several techniques for edge linking are described in the image analysis literature (Castleman, 1996; Gonzalez and Woods, 2008). One of these techniques is the so-called Hough transform, which is extensively used in Paper I.

**Hough transform** Usually a straight line in the $xy$-plane is given by

$$y = kx + m, \quad (3.15)$$

where $k$ and $m$ are constants. However, it is impractical to use this formulation since it fails to express vertical lines due to the infinite value of $k$ required. This can be handled by expressing the straight line in polar coordinates:

$$\rho = x \cos(\theta) + y \sin(\theta), \quad (3.16)$$

where $(\rho, \theta)$ defines a vector from the origin to the line, which is perpendicular to the line. Consider a set of edge points $(x_i, y_i)$ that lie on a straight line having parameters $\rho_0$ and $\theta_0$. Then each point in $xy$-space corresponds to a sigmoid curve in $\rho\theta$-space. All these curves intersect at the point $(\rho_0, \theta_0)$, since this is
Figure 3.4: The Hough transform: a) polar coordinate expression of a straight line; b) set of known points in $xy$-space; c) resulting image in $\rho\theta$-space (Castleman, 1996).

...a line which they all have in common, see Figure 3.4. The Hough transform takes the original set of points and creates the intensity image consisting of all sigmoid curves in a discretized $\rho\theta$-space that correspond to lines through these points. The brighter a point $(\rho_0, \theta_0)$ in the resulting image is, the more does it indicate that the original image contains the straight line given by $\rho_0 = x \cos(\theta_0) + y \sin(\theta_0)$.

**3.4.5 Region Growing**

Region growing, as an image segmentation approach, is performed by merging regions in an image. The initial regions may be small neighbourhoods or single pixels. Starting from an initial region, neighbouring regions are merged if they meet certain criteria based on, for instance, region intensity. Region growing
algorithms are computationally more expensive than the simpler techniques, i.e. global thresholding, but since several merging criteria can be specified, the approach is able to utilise several image properties directly and simultaneously in determining the final boundary location (Castleman, 1996). One such method is flood-filling, which starts with the highest local maximum and proceeds down and including all touching pixels that extend down to some predefined limit, for instance, the 80% limit, while checking to see if the peak merges and flood the existing peaks into one peak (Russ, 1999). This particular technique is used in Paper II.

3.4.6 Morphological Image Processing

The result of an image segmentation operation is a binary image. If the segmentation result is not completely satisfactory, some form of post-processing performed on the binary image can often improve it.

A powerful set of image processing operations based on a set-theoretical approach is called morphological image processing (Castleman, 1996). In general, morphological image processing operates by passing a structuring element over the image in an activity similar to convolution. The structure element can be of any size and determines the effect of the operations on the image. The basic morphological operations are erosion and dilation. In the following, $B$ represents the binary image and $SE$ represents the structure element. $SE_{xy}$ represents the structure element after it has been translated so that the origin of the element is located at the point $(x, y)$.

To use mathematical morphology on grey-scale images, the crisp set theory must be replaced by fuzzy set theory. This means that each point $(x, y)$ has a membership function $f(x, y)$ describing the degree of membership that the point has to the foreground and background. Mathematical morphology on grey-scale images is used in papers I, II, IV, and V.

3.4.7 Definition of Set Operations

A binary image $B$ defined on $\mathbb{Z}^2$ may be identified with the set of all $(x, y) \in \mathbb{Z}^2$ such that $B(x, y) = 1$. Based on this identification, we defined a few basic operations on $B$.

The translation of $B$ by $x = (x_1, x_2)$ is defined as

$$ (B)_x = \{b + x \mid b \in B\}, \quad (3.17) $$

the reflection of $B$ is defined as

$$ \hat{B} = \{-b \mid b \in B\}, \quad (3.18) $$

and the complement of $B$ is defined as

$$ B^c = \{x \in \mathbb{Z}^2 \mid x \notin B\}. \quad (3.19) $$
3.4.8 Erosion

Erosion is the process of eliminating all the boundary points from an object. It is, for instance, useful for removing objects that are too small to be of interest. The erosion of \( B \subset \mathbb{Z}^2 \) by \( SE \subset \mathbb{Z}^2 \) is defined as (Castleman, 1996; Gonzalez and Woods, 2008):

\[
E = B \ominus SE = \{ (x, y) \mid (SE)_{xy} \subseteq B \}.
\]

(3.20)

In other words, the binary image \( E \) that results from eroding \( B \) by \( SE \) is the set of all points \((x, y)\) such that \(SE_{xy}\) is contained in \(B\).

3.4.9 Dilation

Dilation is the process of incorporating into the object all the boundary points that touch it. It is, for instance, useful for filling holes in segmented objects. The dilation of \( B \subset \mathbb{Z}^2 \) and \( SE \subset \mathbb{Z}^2 \) is defined as (Castleman, 1996; Gonzalez and Woods, 2008):

\[
D = B \oplus SE = \{ (x, y) \mid (\hat{SE})_{xy} \cap B \neq \emptyset \}.
\]

(3.21)

In other words, the binary image \( D \) that results from dilating \( B \) by \( \hat{SE} \) is the set of all points \((x, y)\) such that, if \( \hat{SE} \) is translated so that its origin is located at \((x, y)\), then its intersection with \(B\) is not empty.

3.4.10 Opening

Opening is the process of erosion followed by dilation. It has the effect of eliminating small and thin objects, breaking objects at places where they are very thin, and generally smoothing the boundaries of larger objects by removing pixels (Castleman, 1996; Gonzalez and Woods, 2008):

\[
B \circ SE = (B \ominus SE) \oplus SE.
\]

(3.22)

3.4.11 Closing

Closing is the process of dilation followed by erosion. It has the effect of filling small and thin holes in objects, connecting nearby objects, and generally smoothing the boundaries of objects by adding pixels (Castleman, 1996; Gonzalez and Woods, 2008):

\[
B \bullet SE = (B \oplus SE) \ominus SE.
\]

(3.23)

3.4.12 Feature Extraction

Another important task within pattern recognition is measuring the objects, so that they can be identified by their measurement. In Section 2.5, quality criteria for mammograms are discussed. Concentrating on the pectoralis muscle,
some measurement of the muscle is needed in order to check that the criterion is fulfilled. When constructing computer programs that will determine the quality of a mammogram, the object of interest must be isolated and extracted using some or several of the techniques addressed in this section. The object also has to be measured in order to make a proper assessment whether or not it fulfils the quality criteria for an adequate mammogram. In other words, we need to find the representation and description that best represents the object. By representation is meant that we make the object information more accessible for computer interpretation. By description is meant that we quantify our representation of the object.

3.4.13 Representation

In all representations it is desirable to choose methods where \( r(\zeta(f)) \) can be algorithmically transformed into \( r(f) \), in other words, the representation is approximately invariant of the effect of applying \( \zeta \) on \( f \). The function \( \zeta \) will denote a rotation, scale or translation.

One example of a method of representation is Polygon Approximation. Its goal is to represent an object by a polygon, that is a series of points connected by lines. The polygon should with as few points as possible capture the characteristics of the object.

The procedure of approximating a polygon based on an object starts by marking two points in the object that have the furthest distance from each other. Then these points are connected with lines to obtain a polygon. For each segment in the polygon, we now search for a point on the perimeter between the end points of the segments, that has the furthest distance to the polygonal line segment. If this distance is larger than a threshold, this point is added to the polygon (dividing the considered segment into two). This is performed iteratively until no more points have been added.

3.4.14 Description

The purpose with the description is to quantify a representation of an object. This implies that we, instead of talking about regions in the image, can talk about their properties, such as length, curvature, and so on.

Area and Perimeter are useful examples of descriptors. The area of an object is a convenient measure of its overall size. The area measurement depends only on the boundary and is not concerned with the intensity variations inside the object. The perimeter of an object is useful when the shape of the object is considered and when considering an object with a complex shape, because such objects have a longer perimeter. One of the simplest descriptions is the length \( P \) of the perimeter of an object. The obvious measure of perimeter length is the number of edge pixels. That is, pixels that belong to the object, but have a neighbour that belongs to the background. A more precise measure is obtained by considering each pixel centre as a corner of a polygon.
the perimeter is given by

\[ P = a \cdot N_e + b \cdot N_o \]  

(3.24)

were \( N_e \) and \( N_o \) are the numbers of even and odd directions, respectively, with respect to the chain code description of the perimeter, if we are using a 8-directional chain code (Figure 3.5). The value of \( a \) is typically set to 1 and that of \( b \) to \( \sqrt{2} \) (given that the sampling has been carried out with a square lattice).

The diameter of an object \( O \) is defined as

\[ Diam(O) = \max\{ D(p, p') \mid p, p' \in O \} \],

where \( D(p, p') \) is the distance between the points \( p \) and \( p' \).

The curvature of the perimeter can be obtained by calculating the angle between two consecutive line-segments of the polygonal approximation. Assume that the curve (consisting of polygonal line-segments) is \( \{ p_i \}_{i=1}^N \). Then the curvature at point \( p_j \) is given by

\[ c = ||p_{j-1} - 2p_j + p_{j+1}||^2 \]

where \( p_i \in \mathbb{R}^2 \) and \( \| \cdot \| \) denotes the Euclidean 2-norm.

As mentioned, the perimeter consists of all the object pixels with at least one neighbour in the background. The length of the perimeter is the length of the object boundary. If high precision is important, the discrete properties of the values \( a \) and \( b \) in Equation 3.24 have to be taken into account, but this is beyond the scope of this thesis. One easy way to compute the area of an object is to count its pixels. We can approximate the area of an object with the number of pixels belonging to the object. A more accurate measure may however be obtained by using a polygonal approximation and computing the area of the polygon.

Two other useful descriptors are length and width. When an object has been extracted from an image, it is easy to compute the horizontal and vertical extent of the object. The minimum and maximum row number and column number are the quantities needed. A line passing through \( p \) and \( p' \in O \) with \( D(p, p') = Diam(O) \) is called the main axis of the object. For measurements based on a different orientation it is necessary to locate the main axis of the object and measure length and width relative to it. There are several other ways to establish
the principal axis of an object once the boundary of the object is known. One way is to compute a best-fit straight line (or curve) through the points in the object (Castleman, 1996). According to both Hu (1962) and Castleman (1996), the principal axis can, alternatively, be computed from moments. Moments also yield a useful technique applicable to shape analysis (Gonzalez and Woods, 2008; Hu, 1962).

### 3.4.15 Shape Analysis

Shape features can be used to distinguish an object from other objects. The shape measurement can be done independently of or in combination with size measurement.

A group of shape features are called *circularity measures* because they are minimised by the circular shape. The magnitude of the measure reflects how round an object is. A circle of radius $r$ has the area $A = \pi r^2$ and the length of the perimeter is $P = 2\pi r$. Therefore, by taking the number $P$ of border pixels of an object, and the total number $A$ of pixels of an object, we define the quotient (Castleman, 1996; Gonzalez and Woods, 2008):

$$C = \frac{P^2}{4\pi A}.$$  

(3.25)

The feature takes on a minimum value of 1 for a circular shape without holes.

Another technique to capture shape features are *chain codes*. A simple boundary chain code can be described as walking around the boundary of the object recording each step. Each direction has its own number and the series of numbers uniquely describes the perimeter of the object (Castleman, 1996; Gonzalez and Woods, 2008). The differential chain code is based on direction dependency. This means that the number itself is not used, but only the differences between subsequent directions in the chain code. The differential chain code reflects the curvature of the boundary and convexities and concavities show up as peaks, see Figure 3.5.

One mammographic application of chain codes is related to the pectoralis muscle. The anterior margin of the pectoralis muscle may differ in shape between different mammograms. It may be convex and in another mammogram concave. A modified version of chain codes is used in Paper I to determine the convexity vs. concavity of the shape of the pectoralis muscle.

### 3.4.16 Classification

When the image segmentation process has isolated the object and its features have been measured properly, it is possible to determine the type of the object. Properties of the feature measurement are then used to identify and classify an object.

Classifier design consists of establishing the logical structure of the classifier and the mathematical basis of the classification rule (Castleman, 1996).
Commonly, for each object encountered and for each of the classes, the classifier computes a value that indicates by its magnitude the degree to which that object resembles the objects that are typical of that class. There are several methods used for classification, for instance statistical methods and neural networks (Castleman, 1996; Duda et al., 2001; Gonzalez and Woods, 2008; Russell and Norvig, 1995).

### 3.5 Segmentation of the Breast Region

In order to limit the region of search for lesion detection, the breast region may initially need to be segmented from the image. Once the breast region is known, preprocessing and the search for cancer can be limited by the breast border. Several researchers have studied this issue (Bick et al., 1995; Chandrasekhar and Attikiouzel, 2000; Mendez et al., 1996; Lou et al., 2000; Olsén, 2005; Semmlow et al., 1980; Suckling et al., 1995; Wirth and Stapinski, 2003; Yin et al., 1994). One suggested way to correct for the increased density due to the reduced breast thickness, is to incorporate an equalisation of the intensities near the periphery of the breast (Ferrari et al., 2000). This selectively enhances the peripheral region of a mammogram in order to simultaneously display the centre of a breast and the skin line regions without loss in contrast. However, another technique is developed in Paper II. This technique is based on modelling the background and then subtracting it from the mammogram. This results in a more uniform background and solves part of the problem discussed above about displaying skin line regions without loss in contrast.
3.6 Lesion Extraction

One aim of computer-aided diagnosis systems (CAD) is to increase the efficiency and effectiveness of screening procedures by using a computer system as a second reader, to indicate locations of abnormalities in mammograms as an aid to the radiologist. The final decision regarding the likelihood of the presence of a cancer is left to the radiologist, and so is the patient management (Nishikawa, 2007). Lesion extraction, which extracts the lesion from the surrounding tissues, is an essential step in the computerised analysis of mammograms. As mass lesions are usually embedded and hidden in varying densities of parenchymal structures, the task of lesion segmentation is not trivial (Yuan et al., 2007). Figure 3.6 shows a schematic diagram of a computerised detection method for use in screening mammography.

If the radiologist detects a region with an abnormality, he/she decides what course of action should be recommended (i.e. return to screening, return for short term follow-up, or return for biopsy). One example of abnormality is lesion, which is defined as a pathologic change in the tissues.

Thus, another aim of CAD is to extract and analyse the characteristics of benign and malignant lesions seen at mammography in an objective manner in order to aid the radiologist. This has the potential to increase diagnostic accuracy and to reduce the number of false-positive diagnoses of malignancies. This,
in turn, may reduce the number of surgical biopsies and their associated complications. Figure 3.7 shows a schematic diagram of a computerised diagnosis method for use in the mammographic workup of a suspect lesion. Computerised classification methods are also being extended to ultrasound (Hara et al., 2002) and magnetic resonance images of the breast (Gilhuijs et al., 2002).

3.6.1 Extract Masses

Computer-aided diagnosis for characterisation of mammographic masses as malignant or benign has the potential to assist radiologists in reducing the biopsy rate without increasing false negatives (Shi et al., 2008).

In many ways, breast masses, lumps or aggregations of coherent material are more difficult to detect than microcalcifications since, as mentioned above, masses can be mistakenly simulated or obscured by normal breast parenchyma. As an initial step in lesion detection, the breast region may need to be segmented from the image.

In an attempt to better characterise findings such as masses, most computer detection methods can be described as a pre-processing stage during which a threshold is applied at each pixel based on a feature such as intensity, gradient, line orientation or a combination of features, where remaining pixels can then be grouped yielding lesion candidates. Figure 3.8 shows a typical mass lesion in a mammogram.

Many methods for computerised detection and classification employ mathematical representations of the radiographic features exhibited by mass lesions or clustered microcalcifications (Giger et al., 2000). The characteristics of masses
Figure 3.8: An enlargement of a spiculated mass in mammogram number 202 from the MIAS database.

are determined by their degree of spiculation, margin definition, shape, density, homogeneity (texture), asymmetry, temporal stability, and so forth. Descriptors of these characteristics may also be grouped into gradient-based features, intensity-based features, and geometric features. The extraction of such features may be performed at all pixel locations in an image or just at those locations delineated as a suspect in a pre-processing stage.

One very important feature in the detection and diagnosis of cancerous breast lesions is spiculation. Much work among various investigators concerns mathematical descriptors of spiculation. The aim is to distinguish between true-positive detections and false-positive detections in a mass detection algorithm or in distinguishing between malignant and benign lesions in a mass classification and diagnosis algorithm.

Gradient-based spiculation feature analysis commonly involves the calculation of the gradient relative to either the $x$-direction, $y$-direction or radial direction. For an image function $f(x, y)$ where $(x, y)$ is a point within the image, the Cartesian gradient $\Delta f(x, y)$ is formed as we saw in Section 3.4.1:

$$
\begin{bmatrix}
\frac{\delta f(x,y)}{\delta x} \\
\frac{\delta f(x,y)}{\delta y}
\end{bmatrix},
$$

with a magnitude $m(x, y)$ given as

$$
m(x, y) = |\Delta f(x, y)| = \sqrt{\left(\frac{\delta f(x,y)}{\delta x}\right)^2 + \left(\frac{\delta f(x,y)}{\delta y}\right)^2},
$$

(3.27)
and orientation $\phi(x, y)$ given as

$$
\phi(x, y) = \tan^{-1}
\left( \frac{\frac{\delta f(x, y)}{\delta y}}{\frac{\delta f(x, y)}{\delta x}} \right).
$$

(3.28)

The radial gradient vector has the same magnitude as the Cartesian gradient vector $\Delta f(x, y)$, but the orientation is given as

$$
r(x, y) = \Theta(x, y) - \phi(x, y),
$$

(3.29)

where $\Theta(x, y)$ is the angle of the vector connecting the centre $C$ of the lesion to the point of interest $(x, y)$.

Shape is another important feature in the detection and diagnosis of cancerous breast lesions. Determining the circularity of a shape is used in the detection of mass lesions since masses tend to be rounded in form. Other features in the detection and diagnosis of cancerous breast lesions are density, contrast, and texture (Giger et al., 2000). Deviations from architectural asymmetry of normal right and left breasts (bilateral comparison) are yet another feature.

When judging mammograms, radiologists find it helpful to review cases from prior examinations for the detection of interval changes. Mammograms obtained at different times will typically differ in compression and positioning, thus making temporal subtraction practical only for mass lesions rather than clustered microcalcifications. Various landmarks in the mammogram such as the skin line, nipple location, or the location of large area densities are used on temporal registration of mammograms. The use of correspondences between multiple views of the breast is essential in assessing the presence of an actual lesion to avoid misclassifying gathered parenchyma created only in appearance by overlapping normal parenchymal structure.

Practically, while developing algorithms for the detection of masses, researchers extract more than one feature and use classifiers to merge the features into an appropriate outcome. One example of a method using classifiers for mass characterisation is based on a segmentation algorithm and on the neural classification of several features computed on the segmented mass (Delogua et al., 2007). Extraction of masses plays a key role in most computerised systems. The technique, used in the paper mentioned, is a gradient-based one. In the proposed method, sixteen features based on shape, size and intensity of the segmented masses were extracted and analysed by a multi-layered perceptron neural network trained with the error back-propagation algorithm. Different combinations of features will, in general, yield different classification performances. Computer assisted applications in mammography are often faced with the task of selecting a useful and limited subset of features from many gathered features to predict likely pathology (Giger et al., 2000).

Computerised detection of masses in digital mammograms based on bilateral subtraction images was reported already back in 1991 and 1994 by Yin et al. (1991, 1994). More recent work based on analysis of bilateral mammograms for computerised mass detection systems was reported by Wu et al. (2007). The aim of bilateral comparison analysis is, for each detected object, to perform a
regional registration to define a region of interest and compare the “symmetry”
to the object location on the contralateral mammogram. Multiview comparison
of mammogram is challenging since, as mentioned above, the acquisition proce-
dure (based on compression and positioning) for each view will differ. However,
comparing two mammograms of the same woman is an important aspect to
consider for the mammography system to be successful.

In relation to the discussion above, Qian et al. (2007) emphasise that the
main drawback of CAD methods is the fact that the single-view mammogram
does not allow for full analysis of breast image information. Therefore, they
propose a new fully automatic multiview method. However, their method is
based on constructing an ipsilateral multiview (on the same breast at the same
side of the women) CAD system, in contrast to bilateral comparison, for early
stage breast cancer detection. This approach is important because of the use
of concurrent analysis similar to radiologists’ interpretation. A great deal of
important information can be derived through the comparison of different views
by using concurrent analysis of the same breast over time. Since concurrent
analysis is the main purpose of mammography screening, this is an interesting
innovative approach for computerised mammography systems. However, while
designing these kinds of systems, it is important to take into account that each
image of the ipsilateral pair is affected by a different distortion of the breast
tissue since it may take one or two years between the examinations during
which the mammograms are acquired.

3.6.2 Location and Segmentation of Microcalcifications
The identification of calcifications is a major goal of screening mammography.
Benign calcifications occur in blood vessels, see Figure 3.9. Therefore, the dis-
tinction between ductal and vascular microcalcifications needs to be made if
Towards Automatic Image Analysis for Computerised Mammography

Figure 3.10: Malignant microcalcifications. Image from the Mammographic Unit at the University Hospital of Northern Sweden.

the number of false positives should be kept low (Highnam and Brady, 1999). This is relatively easy for radiologists interpreting mammograms against the background of their knowledge of breast anatomy. It is, however, a challenge for image analysis systems. Thus, expectedly, the development of algorithms for identification of microcalcifications has been the subject of intensive effort in mammographic image processing.

Due to the high x-ray attenuation of calcium, calcifications appear as small white specks in the mammogram. The detection and characterisation of microcalcifications provide additional opportunities for the application of computer vision and artificial intelligence in mammography. Calcifications can be described by shape, area and brightness, based on the morphology of individual calcification, the spatial distribution of calcifications within a cluster, and the heterogeneity of individual features within a cluster (Giger et al., 2000).

The contrast of calcifications relative to the background of parenchymal structures has been utilised in two general ways: searching for regions of increased contrast (with or without pre-processing to enhance such a feature) and performing texture analysis due to the variation in texture patterns with spikes (corresponding to the calcification) and those without spikes. The sizes of clinically relevant calcifications are usually less then $500 \, \mu m$. This characteristic is often taken into consideration when extracting microcalcifications (Giger et al., 2000). Another characteristic taken into account when extracting microcalcifications is edge strength, which is strongly coupled with contrast differences such as linear structures, sharp edges, and topographic peaks.

Isolated calcifications are not clinically significant, and thus a clustering criterion has been incorporated into computerised detection methods by many investigators (Jiang et al., 1996a,b; Nishikawa et al., 1993). Malignant clustered microcalcifications are shown in Figure 3.10.
It is difficult, however, to determine whether a lesion with clustered microcalcification is malignant or benign on mammograms. In many years, CAD tools for identifying the histological classification of clustered microcalcifications in order to assist radiologists' interpretation as a “second opinion” have been developed (Nakayama et al., 2006).

Recently, much research work regarding CAD system for finding microcalcifications incorporates pattern classification techniques to improve earlier work. For instance, (Ge et al., 2006) used several pattern classification techniques such as neural networks and linear discriminant analysis to identify microcalcification clusters automatically on full field digital mammograms. This is expected, since pattern classifications techniques are well suited for these contrastive and relatively easily defined objects compared to, for instance, masses.

In a very interesting paper, Sahiner et al. (2006) developed a new technique to improve the accuracy of computerised microcalcification detection by using the joint two-view information on cranio-caudal (CC) and mediolateral-oblique (MLO) views. After cluster candidates were detected using a single-view detection technique, candidates on CC and MLO views were paired using their radial distances from the nipple. This approach is not only interesting because their results indicate that the correspondence of cluster candidates on two different views provides valuable additional information for distinguishing false positives from true microcalcification clusters. In addition, an automatic two-view comparison is, as mentioned in Chapter 2, desirable to incorporate into a quality assessment system for mammography.

### 3.6.3 Extraction of Curvilinear Structures

Ductal structures in a mammogram take the form of curvilinear structures (CLS) that are slightly brighter than their surroundings. The breast contains many curvilinear structures that correspond mammographically to CLS, such as milk ducts, blood vessels, supporting tissues, called Cooper’s ligaments, parenchymal tissue, and edges of the pectoral muscle (Rangayyan et al., 2007). As seen in Section 2.1, the parenchyma is held in place by Cooper’s ligaments, and the whole structure can move non-rigidly as the breast is compressed. The presence of CLS is an important factor in the detection of abnormalities in mammograms for several reasons. On the one hand, some lesions are characterised by the presence of certain types of CLS, such as spicules, in the mammographic image (for example, spiculated masses and architectural distortion), or by the asymmetric disposition of the oriented texture in the breast image. On the other hand, some lesions, such as circumscribed masses, may be obscured by superimposed CLS; the resulting altered appearance could lead to misdiagnosis. The segmentation, identification, and removal of such structures potentially facilitate a wide range of mammographic image processing applications, such as mass detection and temporal registration. Thus, the ability to detect and classify CLS could enhance the performance of CAD algorithms. Several authors have focused on the detection of linear mammographic structures for the detection of ducts over the years (Bakić et al., 1998; Highnam and Brady, 1999; Bakić et al., 2002; Evans...
et al., 2002; Abdel-Mottaleb et al., 2004)

The identification and removal of CLS has been shown to lead to increased accuracy of detection of signs of breast cancer, such as architectural distortion (Rangayyan and Ayres, 2006). According to Rangayyana et al. (2007), the small number of methods that have been proposed in the literature for the detection of CLS have demonstrated limited success. Furthermore, the authors emphasised that there existed a need for the development of improved detection and analysis techniques for accurate discrimination of spicules against blood vessels and ducts.

3.7 Clinical Aspects of Computerised Aids

Computer analysis in mammography has been discussed for many years. Many researchers have been working intensively on developing methods to interpret mammograms. Back in 1967, Winsberg et al. (1967) proposed the automation of the reading of the radiographs by means of optical scanning and computer interpretation with the aim to demonstrate the feasibility of this method.

For maximum diagnostic yield and a high level of reliability in mammography, independent double reading by two radiologists is recommended (Thurfjell et al., 1994). As mentioned before, the main purpose of CAD systems is to assist the human reader in the detection of breast cancer. The development of CAD systems seems to be necessary because of the financial and logistical problems associated with the double reading of radiologists (Malich et al., 2006). The CAD systems on the market are highly sensitive and are characteristically more sensitive than the radiologist, mainly in the detection of malignant microcalcifications. Furthermore, several studies demonstrate that CAD systems to a significant extent detect early signs in an earlier stage compared to radiologists (Nishikawa et al., 1998; Birdwell et al., 2001; Freer and Ulissey, 2001; Malich et al., 2006). The CAD effect on radiologists seems to depend additionally on the experience and training of the radiologist in analysing mammograms by using CAD. This will effect on accuracy of using CAD.

Furthermore, radiologists need training in how to use CAD systems in order for screening using CAD to be successful. Despite several technical limitations so far, such as the low reliability and the high costs of this technology, especially in relation to the number of additional cancers detected, CAD systems are undergoing fast further development with continued improvements of the available software versions.

3.7.1 Other Modalities used in Mammography Examinations

In addition to x-ray imaging, there are several further promising imaging technologies for mammography, namely magnetic resonance imaging, ultrasound and more recently digital breast tomosynthesis and breast computed tomography.
Magnetic Resonance Imaging

Gilhuijs et al. (2002) developed a clinical system for computerised delineation, rating and classification of breast lesions depicted in contrast material-enhanced magnetic resonance images (MRI) obtained in women with increased lifetime risk of breast cancer. According to the authors, the system demonstrated potential to help exclude malignancy with high confidence and reproducibility with positive predicative value that is acceptable in screening. The system was developed to assist in the identification of subgroups of lesions with characteristics that can be correlated accurately with a low likelihood of malignancy. Contrast material-enhanced MRI of the breast is reported to be the most sensitive (90%) modality for invasive breast cancer. It has, however, demonstrated lower and variable specificities, and the effectiveness of contrast material-enhanced MRI for screening of asymptotic disease in women at increased lifetime risk of breast cancer is being investigated currently in various multi-institutional trials (Gilhuijs et al., 2002).

Contrast MRI, with its multiplanar imaging capabilities, is being used to screen women at high risk for breast cancer and also for other specific diagnostic situations, but it is expensive and access is limited (Lindfors et al., 2008).

Ultrasound

Sonography (imaging with sound) is a principal method to diagnose breast cancer, especially after high frequency ultrasound devices were invented. Sonography is harmless to the human body and provides real-time diagnosis of internal organs. It has incomparable superiority since it can distinguish and discover masses with little side effects. The entrance of digital and 3D ultrasound scanners have brought about increased diagnostic efficiency and easier diagnostic methods in breast ultrasound compared to the 2D ultrasound. Hara et al. (2002) developed a CAD system for masses on 3D breast ultrasound images and estimated the preliminary performance with 200 cases. Diagnosis by using ultrasound images is strongly recommended for younger women whose breasts tend to be dense (Tian et al., 2007).

However, the sonographic images have some limitations. For instance when the intensity range is narrow, the breast textures and masses are obscure. The contrast becomes low and the masses usually cannot be distinguished. Furthermore, there is a limitation in detecting microcalcifications. Tian et al. (2007) have developed a ultrasound image enhancement algorithm to overcome some of these limitations. They validate the effectiveness and usefulness their algorithm and investigate whether and how its use improves the accuracy of mass detection and classification by clinical trials. Their results show that the new technique can remarkably increase diagnostic rate.

Digital Breast Tomosynthesis

Digital breast tomosynthesis (DBT), a form of limited-angle tomography, has emerged as a new and promising 3D modality in breast imaging to improve the
accuracy of breast cancer screening. Several researchers are working extensively on this technique (Reiser et al., 2006). Results indicate that computerised mass detection in the sequence of projection images for DBT may be effective despite the higher noise level in those images (Reiser et al., 2006).

Breast Computed Tomography

Lindfors et al. (2008) have developed a dedicated breast computed tomographic (CT) system capable of performing a cone-beam CT of the breast. The cone beam of a breast computed tomography (CT) scanner takes 360-degree views of the breast anatomy, without having to compress the sensitive tissue. According to the authors, this system provides full tomographic imaging of the breast for either screening or diagnostic breast examination. Some technical challenges remain compared to images with screen-film mammograms, but breast CT is promising and may have potential clinical applications (Lindfors et al., 2008).

3.8 Mammographic Image Databases

Medical computer systems which aim to assist, support, or replace human operators must undergo comprehensive testing since the systems may play a critical role in patient management. Therefore, in computer assisted mammography, a common digital database would be a positive step towards achieving consistency in performance comparison and the objective testing of algorithms (Chandrasekhar and Attikiouzel, 1998). However, there are many difficulties, both intellectual and practical, with the establishing of such a database (Suckling et al., 1994). Currently, there are several independent databases used within computer-aided mammography research. Two of them are described below, the MIAS- and the DDSM databases. These databases are also used as the data sets in all the papers in this thesis.

In common databases, it is difficult to reliably estimate the accuracy of a CAD scheme. Creating a common database or using a quantitative measure to characterise databases are possible solutions to this problem. However, none of these solutions is available at present. Therefore, the method used for selecting cases, and histograms or mean and standard deviations of relevant image features should be reported whenever performance data are presented (Nishikawa et al., 1994).

3.8.1 The MIAS Database

The Mammographic Image Analysis Society has produced a digital mammographic database (Suckling et al., 1994), which is used in computer assisted mammography research. This database contains 322 images, all taken with the mediolateral oblique (MLO) view, representing 161 bilateral mammogram pairs. The database is divided into seven categories, which include normal image pairs (204 mammograms) and abnormal image pairs (118 mammograms).
containing: microcalcifications (25), circumscribed masses (20), spiculated lesions (21), ill-defined masses (15), architectural distortion (20), and asymmetric densities (17). All mammograms from the MIAS database were digitised to a spatial resolution of 50 µm pixel edge. According to Suckling et al. (1994), all the mammograms have been carefully selected from the United Kingdom National Breast Screening Programme to comply with the highest criteria regarding quality of exposure and patient positioning. Interestingly, an evaluation of the diagnostic quality based on the quality criteria described in Section 2.5 shows that many of mammograms are inadequate with respect to diagnostic quality.

In particular, the mammograms in the database contain artefacts such as plastic tape and scratches. Even though the images are taken with the MLO view, the pectoralis muscle is not always visible.

### 3.8.2 The DDSM Database

The Digital Database for Screening Mammography (DDSM) was constructed to provide the research community with a source of data that can be used to rigorously compare different image analysis techniques. Heath et al. (1998) gave a brief description of the development of the database, some information regarding artefacts and ground truth, but also how to access the mammograms. Although the films have been cleaned before scanning them, artefacts appear in the digitised images, caused by dust or scratches on the film or introduced by the scanner. Heath et al. (1998) pointed out that any practical computer assisted screening system that uses digitised films will need to handle these kinds of artefacts and that even direct digital image systems must cope with pixel non-uniformity and dust from the imaging system. The database was completed in 1999 and contains 2620 cases from several North American hospitals, four-view mammographic screening examinations, grouped as normal, benign, and cancer. The cases were assigned to volumes according to the severity of the finding. Heath et al. (2000) provide an overview of the complete database, unnamed patient information such as for instance age, breast density and screening date, information on usable software for access and use of the database, and a brief discussion of the performance evaluation of CAD algorithms. The DDSM differs from the MIAS database both in diagnostic quality aspects and in the number of views of each woman. The DDSM contains high quality digitised copies of four mammograms taken during a screening examination, one MLO and one CC view on each breast. Several volumes of full-resolution images are available for ftp transfer.

### 3.9 Future Prospects

The field of computer-aided mammography has evolved rapidly over the past two decades and has become a sophisticated technical medical speciality. Progress has been based on the interplay between improved techniques for breast imag-
Towards Automatic Image Analysis for Computerised Mammography

ing, massive advances in computing power and usability allied to substantial improvements in image processing, and an improved understanding of the relationship between breast images and breast disease.

Already back in 1999, Highnam and Brady (1999) argued that reliable mammographic image processing must be based on a physical model of how the image is formed. Furthermore, it is of importance that the image processing specialists work closely together with the clinicians.

Digital mammography has made significant progress over the last few years (Nishikawa, 2007). The increasing use of computers to present, access, and process images, but also the availability of Picture Archiving and Communication Systems (PACS) based on high bandwidth computer networks, facilitate the rapid electronic transfer of images between distant sites. Furthermore, the efficiency and cost advantages of storing images electronically rather than on film must be taken into account. All of these arguments are in favour of an increased use of digital mammographic images.

Research is going on to present “artificial intelligence” software capable of helping the radiologist to identify possible areas of cancer in the breast. In the proceedings of the International Workshop on Digital Mammography from 1994 to 2002 (Gale et al., 1994; Doi et al., 1996; Karssemeijer et al., 1998; Yaffe, 2000; Peitgen, 2002; Pisano et al., 2004; Astley et al., 2006), the focus has turned from the microcalcification and mass detection in digitised hard copy reading to the soft copy interpretation. It is also noticeable that different reasoning techniques are discussed more frequently today. Today, pattern classification and pattern recognition is more involved in the algorithms. Over the years, the major developments have been in automated reasoning systems, particularly those based on uncertain information. The non-image based data the clinician takes into account, such as patient history, age, appearance, and symptoms are incorporated into the computerised diagnosis aid to improve the automatic interpretation of the images by making a judgement based on reasoning. Studies of different 3D techniques have become more common over the years. Full field digital mammography is one contribution to this progress, and computer-aided diagnosis methods are becoming automated. The technical aspects have changed focus from detector systems to image display problems. The easy transmission of digital images from one place to another will be beneficial for women living in rural areas. It may also be possible to perform real time consultations in difficult cases between different hospitals. This introduces another modality, namely, telemammography.

In theory, digital mammography offers significant benefits compared with conventional film-screen mammography, not only in image quality and image management but also in detection aid in diagnosis by the use of CAD. However, to compete with conventional mammography, a digital system must meet the quality requirements established for the conventional system and offer significant advantages that diminish existing weaknesses.
Chapter 4

Segmentation Evaluation

4.1 Background

Performance evaluation is essential for providing a scientific basis for image analysis in general and in medical image analysis in particular. Image analysis usually refers to processing of images by computers with the aim to distinguish and classify objects that are present in an image. Image segmentation is, as we have seen before, the process of identifying and delineating objects in images. Identifying is the high-level process of roughly determining the location of an object of interest in the image. Delineation is the low-level process of determining the position (i.e. the precise spatial extent) of the object in the image. Image segmentation is a fundamental problem in image analysis and is one of the most critical tasks in automatic image analysis. In spite of several decades of research and many segmentation methods that have been proposed, segmentation remains a challenging problem in image analysis and computer vision (Udupa et al., 2006). In this ongoing work, an important question to answer is how to test these methods and evaluate their robustness with respect to a given real-life application (Martina et al., 2006).

Focusing on more specific areas such as medical image analysis, researchers in the area have long sought to extract contours of different body organs and tissue types from medical images of various modalities. One important step toward establishing validity and clinical applicability of medical image segmentation algorithms is to strive towards an objective evaluation of these algorithms based on a large set of clinical data. However, one common way, among many researchers, to evaluate algorithms is to use phantoms or idealistic representations of real data for comparison of their algorithms. Predicting performance on real data, based on such results, may be difficult. Segmentation evaluation is a difficult task and comparing the performance of one algorithms against other algorithms is hard even when the algorithms are evaluated on real clinical data. The performance evaluation of segmentation algorithms is indispensable and thus an important subject in the study of segmentation.
Segmentation algorithms can be evaluated analytically or empirically. The analytical methods examine and assess the segmentation algorithms themselves by analyzing their principles and properties. The empirical methods indirectly judge the algorithms by applying them to test images and measuring the quality of segmentation results. Various empirical methods have been proposed (Warfield et al., 2004; Ölsén and Georgsson, 2008).

Several authors have pointed out that the reliability of the state-of-the-art performance analysis of many real-life segmentation algorithms is limited by several problems (Chalana and Kim, 1997; Hoover et al., 1996; Martin et al., 2006; Udupa et al., 2006; Zhang, 1996): the data sets are too small, there are different data sets used for different estimations of performance; appropriate ground truths (or surrogates) are difficult to acquire; the performance metrics are poorly defined; and large costs in terms of time and effort are involved in collecting and hand-segmenting data.

4.2 Analytical Methods

The analytical methods evaluate the segmentation algorithms directly by considering for instance the principles, requirements, utilities, complexity of algorithms. The principle of the analytical methods is that they evaluate segmentation algorithms based on theoretical characteristics and thus avoid the concrete implementation of these algorithms. Therefore, the results from an analytical evaluation are exempted from the influence caused by the arrangement of evaluation experiments that the empirical methods suffer from (Zhang, 1996). However, analytical methods depend on the general theory for image segmentation and if this theory is insufficient the evaluation will be difficult. Therefore, the analytical methods work only with some particular models or desirable properties of algorithms. Assume that there exist a particular model of the segmentation problem. Then the behaviour of the algorithm on such an image can be analysed (mathematically) in terms of the parameters of the image and the algorithm. Some properties of segmentation algorithms can be deduced from their description such as for instance the processing approach. In theory, algorithms can be evaluated analytically. However many important properties cannot be determined, because the optical segmentation result is not clearly defined.

4.3 Empirical Methods

Since the accuracy of segmentation is often the major concern in real applications, empirical methods are mainly used to determine the correctness of segmentation algorithms in the sense that the resulting segmentation is “the right one”. In order to empirically evaluate the correctness of the algorithm it needs to be implemented. It is also necessary for this evaluation method to have test images and to calculate proper quality measures. This makes this evalu-
Segmentation Evaluation

Evaluation more complicated to apply then analytical methods. On the one hand, empirical methods can be considered be somewhat extent limited, since there is only one property studied. On the other hand, they can evaluate different types of segmentation algorithms, since they only need the input images as well as the resulting output segmentation, no matter which type of algorithms is used. This, in turn, makes empirical evaluation methods relatively general.

Zhang (1996) classifies the empirical methods into two types: goodness methods and discrepancy methods. Following this kind of reasoning, in goodness methods, some desirable properties of segmented images, often checked by visual inspection by a human, measure how “good” the segmented image is extracted, based on the opinion of the human observer. In other words, the performance of segmentation algorithms in focus is judged by the values of goodness measures. Humans are more qualitative in their judgement, whereas computerised algorithms are more quantitative. To use the assistance of humans (knowledgeable in the particularly domain) is common and will remain essential in any practical image segmentation method. However, the incorporation of the knowledge of the domain experts into the computer algorithms has remained a challenge.

4.3.1 Goodness Measures

Empirical methods based on measuring goodness involve evaluating the performance of algorithms by judging the quality of segmented images. This method requires certain quality measures. These measures are subjective and established according to human intuition about what conditions should be satisfied by an “ideal” segmentation (for example, a pretty picture). In the evaluation methods of this kind the goodness measure is computed based on the segmented image without the a priori knowledge of the correct segmentation. In other words, the applications of these evaluation methods do not use any reference image.

4.3.2 Discrepancy Methods

In discrepancy methods reference images are needed that represent the ideal or expected segmentation results. Such an image is then used for comparing the segmentation result produced by a segmentation algorithm and ideal result image by measuring their disparity. In practical segmentation applications, segmenting images is complex, and for fully automatic algorithms some error is inevitable and must be tolerated. In cases where the test images are real images, manually segmented images (with the help of visual inspection by a domain expert) are often used as reference images. The disparity between an actually segmented image and a reference image is the best expected result which can be used to assess the performance of algorithms where the true segmentations (ideal results) are impossible to procure. Both the actually segmented and the reference images are obtained from the same input image. These reference image is often referred to as ground truth, golden truth or gold standard. From now on the term ground truth will be used to refer to it. For many real applications
the ground truth must be manually provided via a visual inspection. In other applications, it is possible to use synthetic images as ground truths. The ground truth can then be obtained by using an appropriate image generation procedure. Discrepancy methods take into account the difference (measured by various discrepancy parameters) between the actually segmented image and the ground truth. The resulting discrepancy measure will give information about how the actually segmented image differ in relation to the ground truth image. Then, of course, a high measure indicates the lower performance of applied segmentation algorithms.

Discrepancy measures can be divided into several groups. There are measures based on the number of incorrectly segmented pixels, on the position of incorrectly segmented pixels, on the number of objects in the image, on the feature values of segmented objects and on several other quantities.

The discrepancy measures based only on the number of incorrectly segmented pixels do not take into account the spatial information of these pixels. Such approaches may give the same discrepancy measure values even if the images are segmented differently. For many tasks a discrepancy measures based on pixel position is more valuable for segmentation evaluation. One way to address this problem, is to use the distance between the incorrectly segmented pixel and the nearest pixel that actually belongs to the incorrectly segmented class.

Discrepancy based on the number of objects in the image is that the correct number of objects of each class, given by the reference image, should be determined. Discrepancy based on feature extraction of objects is to measure if the extracted feature values from the objects in an image is the same as the feature values of the objects in the reference image. The ultimate goal of image segmentation in the context of image analysis is to obtain measurements of object features.

4.3.3 Ground Truth

Ideally, if the ground truth were known, we would have the ideal segmentation and would compare the segmentation output from the computer with this truth/result. If this would be the case for a large number of clinical data sets, then the test would be to see if the hypothesis that the segmentation output from the computer is not statistically different from the ground truth will hold. However, this is rarely the case in medical applications. In this domain, segmentation evaluation is a difficult problem subject to research.

The reference image or the ground truth referred to in the explanation of empirical discrepancy methods can be divided into three different categories outlined below:

**Manual delineation**  The term manual delineation is used when object boundaries are traced or regions are outlined manually by experts (e.g. Figure 4.1). In order for this ground truth to be successful, it is necessary to get multiple repetitions of segmentation by multiple experts of the same object. One way to combined these are simply by averaging the multiple manual delineations (Udupa
Figure 4.1: Since, our aim is to develop robust automated extraction methods for extracting the anatomical features required for an automatic patient positioning assessments in mammography, the correct outline of the breast border is important. Consider the mammograms no. 212, 234 and 241. In the a-column the extraction of the breast border produced by our proposed method is outlined. In the b-column we can see the outlines of the breast border provided of two of the experts.
et al., 2006). Another alternative is to use the method suggested in Warfield et al. (2004) wherein an expectation-maximisation algorithm is proposed, that computes an estimate of the ground truth segmentation from a group of expert segmentations. A third alternative is to use the method proposed in Paper V. In addition Martina et al. (2006) propose two new distance-based measures in order to evaluate well and incorrectly segmented pixels by taking into account both the location of the borders and the expert certainty.

Acquiring ground truth by manual delineation has several shortcomings. As mentioned earlier, it is very costly and time and effort consuming to have several expert physicians skilled in the particular domain hand segment multiple data sets multiple times (Udupa et al., 2006; Chalana and Kim, 1997; Hoover et al., 1996; Martina et al., 2006; Zhang, 1996). Second, it can suffer from high intra- and inter-operator variation (Martina et al., 2006). Third, the precision of manual delineations depends on the sharpness of the object, the window level settings for image display, the computer monitor and its settings (if the outline if performed at the screen via a computer), and even on the experts vision characteristics. Lastly, when object regions/boundaries are fuzzy or very complex, manual delineation becomes ill-defined.

Given the various problems with ground truths (see for example Figure 4.2 and Figure 4.3) and given that manual delineation is an accepted method for acquiring ground truth, it makes sense to examine how we may overcome some of the drawbacks of manual outlining. Therefore, the effort should be to produce one surrogate ground truth that is governed by the underlying characteristics of the delineations provided by the different ground truths. This surrogate ground truth can then be used for evaluation.

**Mathematical phantoms** Mathematical phantoms are created to represent the ideal truth as realistically as possible in terms of image blur, relative tissue contrast and heterogeneity, noise, and background inhomogeneity in the scenes. These phantoms are only feasible in applications where it is possible to represent the ideal truth in a realistic way by constructing a model of the truth.

**Simulated scenes** Simulated scenes is the approach of using the method of mathematical phantoms described above to generate scenes. These scenes are then manipulated with the aim to generate realistic scenes. Then the performance of a segmentation algorithm on the particular scene is evaluated. The main drawback of this approach is that it is difficult to devise deformations and the associated changes in intensity characteristics that are realistic. Another way of using simulated scenes is when it is desirable to emulate the process of image acquisition as realistically as possible. Not all objects might be simulated with their realistic properties, which mean that the resulting scenes do not depict the same challenges as real-world data.
4.4 Segmentation Evaluation in Image Analysis for Computerised Mammography

As we have seen, a quality assessment system for computerised mammography requires both the study of how to emulate the radiologists’ decision making and how to solve the required image analysis tasks. The decision related problems such as to quantitatively describe the quality criteria are, in reality, hard (if not impossible) to solve. Rules that determine the exact meaning of the terms used need to be defined in each possible case. Humans tend to approximate between discrete points in the continuum of possible cases, these points being known to them due to earlier experience in similar situations. This approximation is possible due to the flexibility in the definition of the notions that constitute the rules, in connection with the flexibility of human reasoning.

In order to get a better understanding about the way in which human experts assess the quality of mammograms, a questionnaire (Appendix A) was developed by the author. Another purpose of the questionnaire was also to gather ground truth about the five major landmarks (Appendix B): pectoralis muscle, glandular tissue, position and character of the nipple, the breast border and the presence of any skin folds. The questionnaire forms the basis for a study in which a number of radiologists and radiographers, from several countries in Europe, evaluated 200 randomly selected mammograms (cases) from two different standard databases, i.e. the Digital Database for Screening Mammography (DDSM) (Heath et al., 2000) and Mammographic Image Analysis Society Database (MIAS) (Suckling et al., 1994).

As we have seen, in real-life applications such as for instance computerised mammography systems it difficult to create an objective ground-truth. One of the principal reasons is that the inter-expert variations among manually created segmentations are very high.

Figure 4.2 shows an example of the typical inter-expert variations between the manually segmented regions of the glandular tissue disc, in the same mammogram, provided by five experts in mammography. By observing Figure 4.2, it can be concluded that an automated segmentation algorithm cannot be tested against a ground truth from just one expert, because the evaluation would be highly subjective, with an almost random result. Paper V proposes a method...
striving to create an objective performance measure, based on markings provided by experts.

Another example of why the performance evaluation cannot be based on a marking by expert radiologists is that they fail throughout the database to outline the nipple when it is in profile on the breast border. Furthermore, the experts also fail throughout the database to outline the breast border. As we know from Chapter 2, these two are also important anatomical landmarks and therefore must be extracted with high accuracy. In the histogram-equalised version in (the right image of) Figure 4.3, we can see that there actually exist nipples which need to be extracted well and Figure 4.1 clearly visualises how high the variation, among the experts, is.

### 4.5 Discussion

The analytical and empirical evaluation methods are different and are suitable for different kind of applications and segmentation algorithms. Empirical methods are considered more suitable and useful than the analytical methods for performance evaluation of segmentation algorithms. Since currently no general segmentation theory exists, the feasibility of analytical evaluation is limited (Zhang, 1996). As discussed before, among the empirical methods, the discrepancy methods are considered to be more objective than the goodness methods. However, discrepancy evaluation methods have the disadvantage that...
a ground truth is required for comparison. Generally, for a complete evaluation and comparison of segmentation techniques, a set of performance measures is necessary. Some researchers try to create performance measures for segmentation evaluation in complex domains where no ideal truth exists. The difficulties that arise are related to gathering manual processed images and establishing the surrogate ground truth by merging the manually acquired results for the same image. How to form such a set will be a promising research subject in segmentation evaluation.

An evaluation method should evaluate segmentation algorithms in a quantitative way and on an objective basis. A quantitative study can provide a quantitative measure which can be used to reflect the correctness of the algorithm. For an evaluation to be as objective as possible, it is necessary to exempt or at least minimise the influence of humans in the process and provide consistency and unbiased results. Generally, analytical methods are more ready to apply, but they often provide only qualitative properties of algorithms (Zhang, 1996). Empirical methods are normally considered quantitative as the values of quality measures can be numerically computed. However, Zhang (1996) considers goodness methods less suitable for an objective evaluation of segmented algorithms since they are based on subjective measures of image quality. Furthermore, he claims that discrepancy methods can be both objective and quantitative. This is partly true. If the ground truth available is acquired by several experts outlining, for instance, some anatomical landmarks, the evaluation will be biased on the ground truths generated, and the evaluation will be subjective. If the ground truth is generated by some realistic phantom, then the evaluation can also be considered objective.

In empirical evaluation methods, the assessment of the delineation accuracy is often based on treating all aspects of the region corresponding to the ground truth or the surrogate truth of true delineation with equal weight. This might be correct in the absence of any prerequisites. However, such approaches do not address the fact that some areas of the object may be more important than others. In some domains, such as the medical domain, there are anatomical landmarks that are more important than others. Therefore, it is important to be careful with landmark identification and weighing in evaluation of an algorithm’s recognition performance based on the importance of the particular landmark.

As mentioned earlier, several researchers in image analysis have studied the evaluation of segmentation algorithms which use markings from a group of experts as ground truth. This experience made many of them emphasise the importance of an objective ground truth due to the large variation among the experts markings. During discussions several experts have communicated that this is actually the case in practice as well. They also realise the problems this involves, especially while developing objective computer-aided tools.

Therefore, is important to discuss the problem of identifying ground truth with large but realistic inter-variation and to put effort in developing methods for estimating a surrogate ground truth based on several given segmentations provided by a group of experts.
Chapter 5

Summary of Contributions

As seen in Chapter 2, making an accurate diagnosis imposes a great number of demands on the technical system but also on the personnel working within the mammography unit. Building a computer system that automates this process will ease some of the stringent demands, but is in itself challenging and requires a thorough specification of the requirements.

There are many issues to consider when constructing a computer-aided mammographic system intended to support the radiologist in the screening process. One of the issues is automatic determination of the diagnostic quality of mammograms. If a mammogram is of poor quality, there is no reason to use it in the computer-aided screening system. The radiologist does this evaluation of quality while viewing the mammogram. If the mammogram exhibits too poor quality, it has to be retaken. Automated determination of adequacy in mammograms is not necessarily a part of a fully computer-aided system. In the present situation, it is rather justified to facilitate for the radiographer by providing an evaluation of quality of the mammograms taken before the patient is sent home.

A system for automated evaluation of quality of mammograms should basically take the mammogram as input, segment the desired area for a specific part of the mammogram and work out some relevant measures, make an evaluation of all the quality criteria, and then present an output value indicating the overall quality (Figure 5.1). Such a decision aid system must be reliable and should be integrated into the digital mammography acquisition module. It is desirable not only to make an assessment concerning the accuracy (based on quality) of the mammogram but also to suggest possible causes of inadequacy. Figure 5.2 on page 76 shows the technical criteria presented in Section 2.4, and Figure 5.3 on page 77 shows the influence diagram compiled from the positioning quality criteria presented in Section 2.5. All these criteria need to be evaluated and combined for the final quality output of a particular examination from the quality assessment system. There is very little research done within the field of automatically evaluating diagnostic adequacy in mammograms.

There are several challenging computer-related problems which need to be solved. One of these problems is to quantitatively describe linguistic variables
such as glandular tissue should be well spread. Others are more related to image analysis, such as the problems to extract and distinguish between different tissue types and to extract and analyse other anatomical characteristics within the mammogram. Image segmentation is a fundamental problem in image analysis and is one of the most critical tasks in automatic image analysis systems. As Martina et al. (2006) point out, in computerised mammography an important question to answer is how to test these methods and evaluate their robustness with respect to the mammography system. In this evaluation procedure, we need an ideal result to compare the developed segmentation algorithms with. In the medical domain, this usually involves human domain experts who define estimates of the ideal result. However, in mammography it turns out that these estimates cannot be considered as the ideal result because they suffer from interobserver variability. One way to overcome this problem would be to develop methods for computing a surrogate based on the estimations provided by domain experts in mammography. The most appropriate way of computing a computer-generated surrogate of segmentations created by a group of experts, is subject for research (Warfield et al., 2004).

In the following sections, the contents of the papers included in this thesis is summarised. Paper I-IV (Olsén, 2003, 2005; Olsén and Georgsson, 2005; Olsén and Mukhdoomi, 2007) address the extraction of quality criteria from mammograms, which are presented in Chapter 2. Paper V (Olsén and Georgsson, 2008) addresses overall issues and problems related to segmentation evaluation.

5.1 Detecting the Pectoralis Muscle (Paper I)

Among the existing mammographic views described in Section 2.5, the mediolateral oblique (MLO) view is the most important one since it best visualises the tissue adjacent to the chest wall and the axillary tail, where cancer occurs in the majority of cases (Kimme-Smith et al., 1997). An important criterion for the assessment of the quality of MLO mammograms is the position, shape and visibility of the pectoralis muscle. For this reason, a method for pectoralis muscle analysis was developed, which is presented in Paper I. It involves two
Summary of Contributions

major steps: 1) Detection and extraction of the pectoralis muscle and 2) classification of its shape (which, ideally, ought to be convex; see Section 2.5). The algorithm is based on using the Hough transform to approximate a straight line representing the anterior edge of the muscle. The main idea of this approach is to divide the muscle into subregions representing the muscle at sequential different locations and then analyse the slope of the lines found by the Hough transform in each subregion. The proposed method is reliable and yields very good results. In a test based on 155 images randomly chosen from the MIAS database, the shape was correctly classified in over 95% of the cases (according to a manual examination).

The major contribution of this research is the algorithm that automatically analyses the shape of the muscle and computes measurements needed to determine how well the quality criteria concerning the muscle are satisfied.

5.2 Extracting the Breast Border (Paper II)

The first segmentation procedure involves extracting the principal feature of a mammogram: The breast border, also known as the skin-air interface. This is done by segmenting the breast and the non-breast into distinct regions.

In Paper II, a robust method for extracting the breast border is developed. Issues regarding the extraction of landmarks and the problem of using experts as ground truth are also addressed. The strength of the method and also the main contribution, compared to similar methods, is that it makes use of an iterative approach to reduce the effects of the background, produces smooth edges, and automatically finds thresholds. It has been tested on the entire MIAS database (322 images) with a performance of 94% estimated by comparing the results with the provided ground truth.

5.3 Detecting the Nipple (Paper III)

The nipple is an important anatomical feature. Its extraction is a challenging task since the contrast near the border of the breast, and thus near the nipple in mammograms, is very low. Once the images have been segmented into background and object, the next step is to locate the nipple. This step is important since the glandular tissue, where most cancer develops, converges to the nipple. Therefore, the nipple is a good starting point when judging mammograms and searching for lesions. Radiologists compare corresponding regions bilaterally (Georgsson, 2003), and the nipple is the most important landmark in those comparisons. As seen in Section 2.5, the nipple is also important concerning mammography adequacy (Olsén, 2002).

A common method of finding the nipple, used by radiologists and within computer-aided mammography, is based on the geometrical assumption that the nipple is located at the point on the breast border that has the furthest distance perpendicular to the pectoralis muscle. This assumption has become
widely accepted and is used by others as well (Yam et al., 2001). However, according to Chandrasekhar (1996), this assumption does not necessarily hold. Olsén and Georgsson (2003) investigate how well this geometric assumption holds, and the results show that it is a rather good approximation of the location of the nipple. It turns out to be highly unlikely that the estimation obtained in this way will be further away from the true location than four times the average extension of the nipple. Furthermore, the estimation is far more sensitive to the extracted skin line and the stretching of the breast than to the detection of the pectoralis muscle.

There are some methods for locating the nipple and extracting the breast border presented in the literature. Chandrasekhar and Attikiouzel (1997) developed a promising nipple detection method based on the observation that the iso-intensity lines deviate towards the skin line at the position of the nipple. Later, the same authors presented a promising breast border extraction method (Chandrasekhar and Attikiouzel, 2000).

In (Olsén, 2004), these two methods were implemented and tested. However, it was found that both, in order to be useful for fully automatic methods, needed to be improved with respect to both robustness and generality and must also be made more automatic. Based on the results of the research presented in (Olsén and Georgsson, 2003) and the conclusion in (Olsén, 2004), further research was conducted in order to make it easier to find the position of the nipple automatically.

As a result of this former research, Paper III (Olsén and Georgsson, 2005) describes the development of a method that automatically restricts the search area to the region where the nipple most likely is located (both along the breast border and into the breast tissue). Thereafter, the nipple detection method of (Chandrasekhar and Attikiouzel, 1997) is applied, which turns out to yield acceptable results. In addition, the problem of using an expert panel as the ground truth is addressed.

### 5.4 Extracting the Glandular Tissue (Paper IV)

Breast cancer in females often occurs in fibroglandular parts of the breast tissue. Due to several reasons, the majority of which are related to the development of techniques for computer-aided diagnosis of breast cancer, there has been an increasing need for developing an automated segmentation algorithm for extracting the glandular tissue. A first step in that process can be to locate and segment the glandular tissue disc. The next step would be to analyse the different structures within that disc. The location and delineation of the glandular tissue disc is also important, as a quality assurance, in order to determine if the entire glandular tissue disc is depicted in the mammogram.

To make this procedure fully automatic is a highly difficult segmentation task since many different tissue types within the three-dimensional breast superimpose when being projected down to a two-dimensional image plane. The novel approach for extracting the glandular tissue disc proposed in Paper IV
extracts regions depicting fibroglandular tissue in mammograms in a robust and reasonably accurate way. This approach is based on iterative global histogram analysis of the glandular regions in mammograms. The algorithm has been evaluated with promising results based on an ensemble of ground truths provided by a group of mammography experts, and its performance seems to match that of medical professionals specialised in mammography. Furthermore, it can be concluded that this algorithm is able to calculate a good approximation of the location and delineations of the glandular tissue disc.

However, since the most appropriate way of comparing computer generated segmentations to segmentations created by a group of experts is unclear so far, evaluation of segmentation algorithms for mammograms is very complex. Another main observation and contribution of Paper IV is the fact that, in fields where only uncertain ground truths are available, formation of an objective evaluation method is as important as the construction of the segmentation routines that are being evaluated.

5.5 Identifying Ground Truth Based on Multiple Markings by Domain Experts (Paper V)

Performance evaluation is essential for providing a scientific basis for image analysis in general and for medical image analysis in particular. In order to evaluate the performance of segmentation algorithms for image analysis, the results need to be compared with a ground truth. The ground truth which is used is often based on markings by human domain experts. Ground truths are frequently considered to be identical to the ideal result. However, depending on the application domain, this simplification may be highly inappropriate. Furthermore, Udupa et al. (2006) pointed out that it is usually impossible to establish an ideal segmentation result. Therefore, we need to choose a surrogate for the ideal result.

In earlier work (Olsén and Georgsson, 2006), we propose a method to relate ground truths provided by an expert panel to each other in order to establish levels of agreement. In other words, we measure how well each ground truth fits into the ensemble given. Of course, it would be highly desirable to be able to take an ensemble of ground truths and turn it into one ground truth. This is what we call a surrogate ground truth representing the best combination of the individual ground truths. In Paper V (Olsén and Georgsson, 2008), which extends (Olsén and Georgsson, 2006), we present an algorithm that combines an ensemble of markings into one surrogate ground truth.

To summarise, the major contribution of Paper V is the ALGSII algorithm which combines the outlines given by an ensemble of domain experts to yield one outline of an object, which is then considered to be the surrogate ground truth. This surrogate ground truth represents the common characteristics of the outlines provided by experts in the field and can therefore be assumed to be close to the ideal truth that should be used for segmentation evaluation.
Figure 5.2: Influence diagram for the technical quality criteria.
Figure 5.3: Influence diagram for the position quality criteria.
Chapter 6

Future Work

The aims of the post-research are to deepen the understanding of initiated work and to make new contributions within the field of image analysis, still keeping the focus on medical imaging. The expertise gained during the past four years yields a reliable basis for future research in this field.

To date, the algorithms developed within the research have been implemented and tested, and the tests have revealed that they work efficiently and yield satisfactory results. However, a more general investigation of the properties, advantages, and limitations of these algorithms are interesting subjects for future research. Such algorithmic topic investigations are particularly interesting for those algorithms which are of a general nature and should, thus, be made available for other applications as well. Of course, the future development of computer-aided mammography and especially the questions concerning quality assessment systems will be an interesting area that will remain some focus. Apart from this, the focus will also be on the part of the questionnaire concerning the radiologist decision making. Most important and interesting is the continuation of the work proposed in Paper V concerning the identification of ground truth useful in segmentation evaluation for real-life problems.

6.1 The Decision Making Process of Experts

During the post-licentiate studies made for this thesis, extensive effort has been put on the ground truths provided by experts in mammography. The compilation of the part of the questionnaire concerning the markings surfaced many interesting questions involving segmentation evaluation. The purpose of that part of the questionnaire was to gather ground truths useful to evaluate the algorithms developed in the quality assessment project. However, the compilation of that part made us suspect that the provided ground truths were unfeasible as they were provided. Our suspicion was also confirmed in an investigation carried out in the beginning of the work on Paper V. We also, as described several times, investigated and proposed, in Paper V, how to merge these provided
ground truths into one surrogate ground truth.

However, this leaves us with the part concerning the decision related problems such as to quantitatively describe the quality criteria. This is, in reality, difficult (if not impossible) to solve. Rules that determine the exact meaning of each term used in the quality criteria need to be defined in each possible case.

Preliminary results of the compilation of the study show that there is a large variation among the experts’ decisions concerning the diagnostic quality. The problem of how to handle this variation poses interesting research questions regarding decision making. Interesting and relevant questions that that the work in this thesis leaves open are: Is there a significant difference between radiologist and radiographers with regard to their answers? Are there national differences between radiologists? What does the difference depend on? How do we merge the answers? Should we merge the answers? And so on...

An example of the complexity involved and the resulting inconsistency in assessing mammogram adequacy are illustrated by the fact that in 78% of all cases, presented to five different examiners, the binary decisions on the matter of whether or not to recall the patient for further examination differed. Hopefully, the evaluation will improve the understanding of how the radiologists value the quality criteria in order to make a final decision concerning the diagnostic quality of the mammogram in question.

### 6.2 Assessing Ground Truth Useful in Segmentation Evaluation for Real-Life Problems

In order to investigate the performance of an image analysis system it is desirable to estimate the error of the computed result. In order to calculate this error the true value (ideal result) is needed to be known.

Further work will continue our work on the problem of identifying and formalising the problem of identifying ground truth (Paper V). Unique data for that study has, as we have seen, been collected using the questionnaire described in Chapter 4. More specifically, for further investigation are: What is the proper way to perform empirical segmentation evaluation in the absence of the ideal result? In which way should we identify the surrogate ground truth based on ground truth given by experts in a real-life application? To answer these questions it is interesting to further evaluate our proposed measure in Paper V, based on accuracy, limitations and feasibility. An ongoing work is the comparison between ALGSII (Olsén and Georgsson, 2008), STAPLE (Warfield et al., 2004) and voting (Ma et al., 2006; Warfield et al., 2004). Five experts provided five different ground truth for five anatomical landmarks based on 200 mammograms. This gives us in total 5000 delineations of anatomical landmarks provided as being the ground truth to use for segmentation evaluation. Preliminary results show that voting and ALGSII are superior to STAPLE for identifying the surrogate ground truth. However, the performance comparison between voting and ALGSII is so far unclear and subject for future work.
References


Carlson, S. (2002). Personal communication with chief radiologist Stina Carlson at the University Hospital of Northern Sweden.


Towards Automatic Image Analysis for Computerised Mammography


Olsén, C. and Georgsson, F. (2008). Identifying ground truth based on multiple markings by domain experts. AIIM. Submitted for review to AIIM.


