Universitat de Girona

AUTOMATIC MASS SEGMENTATION IN MAMMOGRAPHIC IMAGES

Arnau OLIVER MALAGELADA

ISBN: 978-84-690-8019-1
Dipòsit legal: GI-1188-2007

Universitat de Girona

Department of Electronics, Computer Science and Automatic Control

PhD Thesis

Automatic Mass Segmentation in Mammographic Images

Thesis presented by

Arnau Oliver i Malagelada,

to obtain the degree of:
PhD in Computer Engineering.

Supervisor:
Dr. Jordi Freixenet i Bosch

Girona, Abril 2007

Als qui m'ajuden a tocar el cel amb la mà

Agraïments

Per fi he acabat la tesi! No ho hauria aconseguit sense el suport explícit de molta gent, i tampoc sense el suport en altres moments puntuals de molta altra gent, que durant tots els anys de la meva vida m'han anat aportant informació i m'han anat configurant i refinant aquest caràcter tossut i enfadós, i amb aquell toc d'humor grotesc que segurament tant em caracteritza. Així doncs, la corrua de gent a qui donar les gràcies és llarga, tot i que no tant com els sants de l'Almodòvar.

Aprofitant el tema cinèfil, podria tirar endavant la típica i tòpica metàfora per donar les gràcies. Donar les gràcies a la família, fent servir un tema bíblic. Donar les gràcies als amics d'infantesa, fent servir una pel·lícula de dibuixos animats. Donar les gràcies als amics de joventut i d'universitat, fent servir un serial televisiu. Donar les gràcies als professors que m'han ajudat a estimar l'estudi, fent servir el club dels poetes morts. Donar les gràcies als professors del departament, fent servir la trilogia del padrino. Donar les gràcies als companys que han dinat amb mi durant aquests anys, fent servir un rodolí d'en Pastallé. Donar les gràcies als companys de grup, fent servir alguna d'aquelles pel·lícules americanes sobre societats americanes secretes. Donar gràcies al meu director de tesi, fent servir alguna pel·lícula de mestre i alumne, senyor i esclau. I donar gràcies a la meva nova família, fent servir una pel·lícula romàntica. I donar gràcies a tots aquells a qui m'he deixat, tot esmentant blade runner i black rain. Però no m'agraden les metàfores.

Així doncs, gràcies Papa i Mama per estimar-me durant tots aquests anys i deixar-me fer-vos patir sense parar. Gràcies Joan i Alfred per entretenir-me tot educant-me de petit. I també de gran. Gràcies cunyades Maria Antònia i Maria Mercè, i com no, gràcies nebodes Isaura, Griselda i Maria, i gràces nebot Josep. I també gràcies a tu, Lídia, que m'has, m'estàs, i m'estaràs suportant les meves

neures, tot omplint de joia el meu cor. I també et dono les gràcies, tant a tu com a la teva família, per ser ara també la meva.

També he donar les gràcies als companys molletencs i flaçanencs que de petit vam ser amics, i que durant vuit (en Vadó, l'Albert, en Jordi, l'Alícia...), deu (la Sussanna, la Maria), dotze (en Miquelet) i setze (buf!, eh Joel) anys hem estudiat, hem jugat i ens hem divertit junts. En aquest mateix paràgraf he de donar les gràcies als que vam estudiar als Salesians (en Ricard, en Pere, en Narcís, en Paquete Laureano, l'Albert, en Dani, ...), a l'Autònoma (en Jordi, en Javi, en Figueras, ...) o a la UdG (en Joan, en Joan, l'Emili, en Narcís, en David, ...). També hi ha els companys que no vam coincidir al mateix curs, però que hi he jugat des de petit (en Pitu Maria, en Xiqui, en Carlus, en Jordi, en Jordi, ...). I com no, els companys que encara no he dit i amb qui vaig compartir dos (en Raül, que era andorrà) i quatre anys (l'Albert) de la meva vida.

Ara ja toca donar les gràcies a la gent que ha estat al meu costat durant aquests anys de recerca intensiva i sense parar. Seria injust no començar pel meu director de tesi, en Jordif, i que consti que no ho faig per posar el burro a davant! Sense en Jordi aquests anys hauria estat perdut, no hauria vist la llum. Amb ell al costat, en canvi, sempre hi ha hagut un moment per a l'esperança, i també pel cafè. També gràcies a l'Avi Cufí i a en Joanm, que m'han fet somriure cada cop que estava trist. Gràcies també als altres doctors que corren per aquí i no ens veiem tant, com ara en JB, en Quim, la Marta, en Rafa, en Pere, en Marc, en Pep. I també gràcies en Figui, en Jar, en Lluispa, en Toni, en Tomàs, l'Íngrid, ... per veure'ns i xerrar cada dia. Gràcies també als companys que estan fent o ja han acabat el doctorat: en Jordip, en Carles, l'Ela, en Radu, en Tudor, la Bet, la Silvana, els piscinerus, o els nouvinguts François, Olivier, Marina, Pío, Ricard... i també l'Arman, que deu estar fent la mili. Sense ells, els dinars haurien estat molt avorrits. Gràcies també a l'Anna i la Marta, amb qui anar a demanar material ha estat molt senzill. I gràcies també a la Maria i la Montse, que sempre han tingut un somriure a punt i una mirada encisadora (només) per mi.

Gràcies a en Josep i a l'Elsa, i a l'Hospital Josep Trueta en general. Sense el seu ajut, la seva paciència, i el seu companyerisme no hauria pogut treballar. Gràcies també a l'Albert, per fer la transferència entre les dades mèdiques i les informàtiques.

I també a en Jordi i a en Marcel, pel seu algorisme de segmentació. I gràcies a la resta de projectistes que heu passat per les meves mans. Ja ho diu la cançó: "argo se muere en er arma cuando un projectista se va".

També dono gràcies, amb la mà al cor, als companys de laboratori. L'Anna, voleiant entre Oxford i aquí. I en David, sempre envoltat d'alumnes i alumnes, i de projectistes i projectistes. Gràcies també als doctors de l'anterior fornada: en Xevi, en Xevi i en Xavi. Molt variat, eh? El primer és més conegut per un sobrenom, que no el diré. El segon és el que m'aconsella als moments difícils, m'ajuda pel messenger, i se'n va de viatge amb les meves guies. I gràcies també al tercer, per revisar-me la tesina i també la tesi, per compartir xerrades i amistat. I no vull acabar sense donar les gràcies a en Bobby, vull dir en Robert, que des que va arribar l'he tingut monopolitzat corregint el meu anglès. I el que em queda per aprendre... Sort a tots!

Així doncs, *als que m'han donat tot el seu suport i amor durant tots els meus anys de vida. Als que em van ensenyar a estudiar i a aprendre. Als que vaig conèixer de petit i, tot i estar lluny, ja mai ens hem oblidat. Als que he anat trobant pel camí, ens hem ajudat i mai més ens hem tornat a veure. Als que em van ensenyar que estudiar no és un camí de roses. Als que em van animar a seguir estudiant i a aprendre. Als porreros. Als que durant aquests últims anys m'han financiat com han pogut. Als rivals d'Uno. Als projectistes, els meus deixebles. Als que dinàvem junts i arreglàvem el món. A la que es passa més dies a fora que a casa. Al que algun dia haurà recorregut el món enter amb una sabata i una espardenya. Al corrector personal d'anglès. Al company de copes furtives. Al de la cua llarga, tenista d'interior. Al gran (i gros) Buc Insígnia. A la que em va voler conèixer a fons i mai més ha pogut sortir del pou. En fi, a tots amics meus. Moltes gràcies!*

Acknowledgments

For those of you that didn't get a word from the previous acknowledgments, probably you'll find this more understandable. I should start saying thanks to Dr. Reyer Zwiggelaar, who gave me the opportunity to do a research stay at the University of East Anglia, and whose comments and multiple e-mails about my research improved the thesis (perhaps too much). Further, without him, I will not have a detailed segmentation of the MIAS database, as he convinced Dr. Erika Denton to do such task. Thank you Erika! and also thank you very much Reyer!

I also have to say thank you to Dr. Danika Kragic, who inspired me in the use of eigenfaces and PCA related techniques for both the creation of the templates and the false positive reduction step.

Finally, I would also like to mention here the external examiners of this document, who provided successful comments to improve the quality of this work. Moreover, a brief thanks is directed to those anonymous reviewers of my papers who made great comments to improve my work and the name of whom I will never know.

Abstract

This thesis deals with the detection of masses in mammographic images. Such masses are signs of breast cancer. The contribution of the thesis is the development of a new framework for the detection of masses by taking breast density into account. As a first step, Regions of Interests (ROIs) are detected in the image using templates containing a probabilistic contour shape obtained from training over an annotated set of masses. Firstly, PCA is performed over the training set, and subsequently the template is formed as an average of the gradient of eigenmasses weighted by the top eigenvalues. The training set is clustered into sub-sets of equal size so that PCA can be applied. The template can be deformed according to each eigenmass coefficient. The matching is formulated in a Bayesian framework, where the prior penalizes the deformation, and the likelihood requires template boundaries to agree with image edges in both position and tangents. In the second stage, the detected ROIs are classified into being false positives or true positives using 2DPCA, where the new training set now contains ROIs with masses and ROIs with normal tissue. Mass density is incorporated into the whole process by initially classifying the two training sets according to breast density. Methods for breast density estimation are also analyzed and proposed. The results are obtained using four different mammographic databases (three digitized and one digital). FROC and ROC analysis demonstrate the validity of our approach. The results show a better performance of the approach relative to competing methods. The false positive reduction stage improves the outcome significantly, and the breast density information provides also an improvement.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The aim of this research is the development of a reliable tool to detect early signs of breast cancer in mammographic images. However, the specific techniques developed in this work, as well as the proposed methodology, can be adapted to detect other types of medical diseases. This initial chapter starts with a brief introduction about mammography and CAD systems. In addition, a general overview of the thesis is provided including the description of its structure.

1.1 Breast Cancer

Breast cancer is considered a major health problem in western countries, and indeed it constitutes the most common cancer among women in the European Union [50]. A study developed in 2003 by the American Cancer Society estimates that in the United States between one in eight and one in twelve women will develop breast cancer during their lifetime [1]. This proportion is reduced in our country, Catalonia, where it is estimated that one in sixteen women will develop breast cancer during their lifetime [62].

In the European Community, breast cancer represents 19% of cancer deaths and the 24% of all cancer cases [49]. Nearly 25% of all breast cancer deaths occur in women diagnosed between ages 40 and 49 years. In the United States, for instance, breast cancer remains the leading cause of death for women in their forties [22].

However, although breast cancer incidence has increased over the past decade,

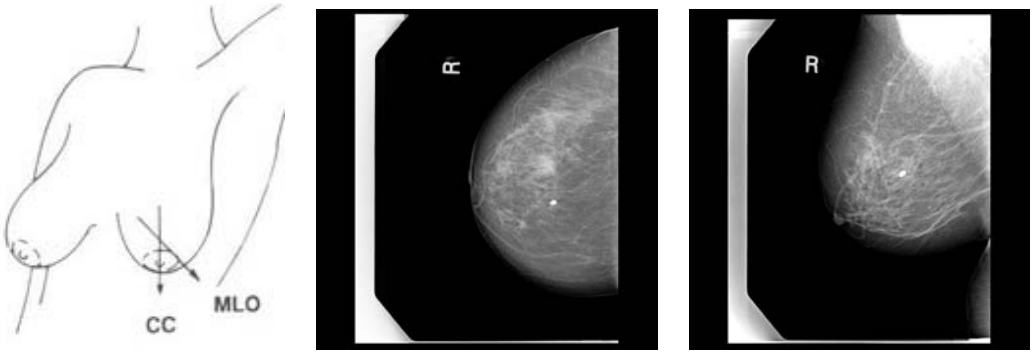
breast cancer mortality has declined among women of all ages [175]. This favourable trend in mortality reduction is considered to be related to the widespread adoption of mammography screening [3, 41, 75, 175, 186, 192] which allows to detect the cancer at its early stages, and to the improvements made in breast cancer treatment [22].

1.2 Mammography

Medical imaging in general, and mammography in particular, lets physicians evaluate a specific area of the body of a patient which might be externally visible. In fact, the *New England Journal of Medicine* [47] regards medical imaging as one of the most important medical developments of the past thousand years, basically due to the fact that it provides physicians with physiology and functionality of organs and cells inside human bodies.

Among the different imaging modalities used for breast cancer detection, mammography remains the key screening tool for the detection of breast abnormalities. In a recent study, Vacek et al. [198] show that the proportion of breast tumours that were detected in Vermont (US) by screening mammography increased from 2% during $1974-1984$ to 36% during $1995-1999$. However, it is also well known that expert radiologists can miss a significant portion of abnormalities [13, 14]. In addition, a large number of mammographic abnormalities turn out to be benign after biopsy [6, 67].

Mammograms capture the low energy X-rays which passes through a compressed breast. Depending on the viewpoint of the X-rays, the images are classified into different categories. Figure 1.1 shows the two most used viewpoints for extracting mammograms: the Medio-Lateral Oblique view (MLO) and the Cranio-Caudal view (CC). Figure 1.1(a) shows physically the viewpoints' directions, while Figures 1.1(b) and 1.1(c) show an example of each mammogram view. It is important to notice that in the MLO views there is one region corresponding to a portion of the pectoral muscle which may be present in the left or the right upper corner of the image. Moreover, some annotations and labels can appear in the images. For instance, the R symbol which appears in Figure 1.1 is used to mark the mammogram as *right breast*.



(a) (b) (c)

Figure 1.1: In mammography there are different viewpoints of the same breast: (a) shows the direction of the two most used views, which produce images like (b) the Cranio-Caudal view, and (c) the Medio-Lateral Oblique view. Note the presence of the pectoral muscle in the upper right corner of image (c).

1.2.1 Mammographic Abnormalities

As Kopans [101] describes in his work, there is a large number of types of abnormalities that can be present in a breast. Among those, signs of breast cancer are normally associated with:

- Asymmetry between images of left and right breasts.
- Distortion of the normal architecture of the breast tissue.
- Presence of micro-calcifications in the breast.
- Presence of masses in the breast.

It is generally accepted that the essential characteristic of a high-quality mammogram is the ability to visualize these four features [5, 177, 208].

Both breasts are usually considered as almost symmetric structures. While exact mirror images are not to be expected when comparing them (usually the first practice of expert physicians), the tissue patterns within each breast should be similarly distributed. An asymmetric area may be indicative of a developing mass or a variation of normal breast tissue.

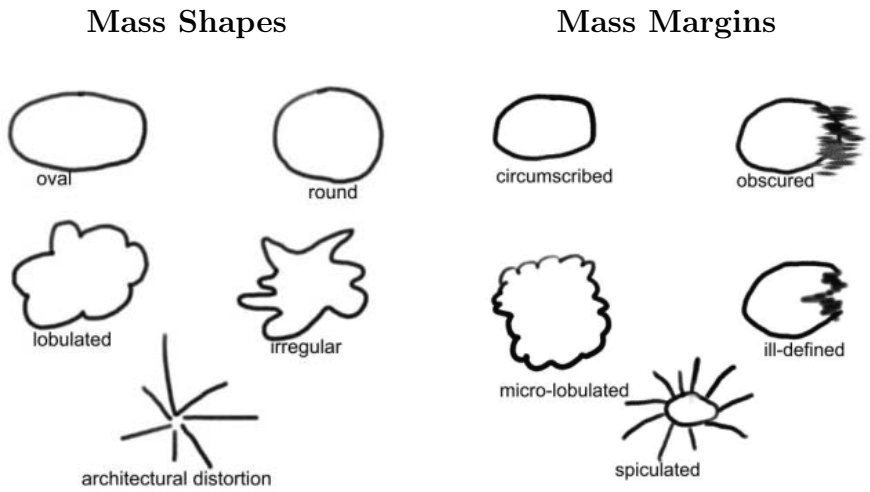
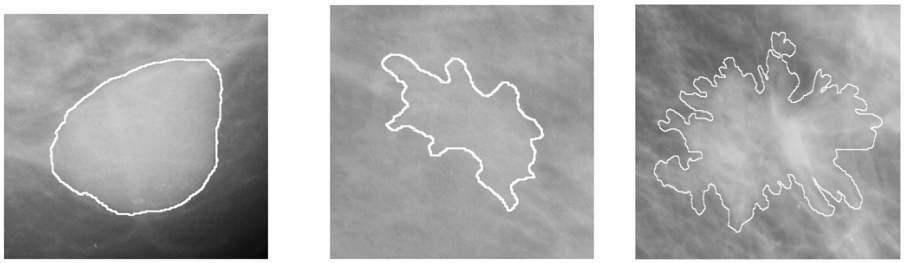


Figure 1.2: The shape and margin of a mass are strong signs of their malignancy/benignancy degree. Image extracted from the web of GE Healthcare [59].

A distortion in the normal breast architecture (architectural distortion) refers to a derangement or disruption of the normal arrangement of the tissue strands of the breast resulting in a radiating or haphazard pattern without an associated visible centre. This includes spiculations radiating from a point, and focal retraction or distortion of the edge of the parenchyma.

Micro-calcifications are tiny calcifications that range from 50 to several hundred microns in diameter, which usually appear in clusters. In these cases, they are analyzed according to their size, shape, number, and distribution. The general rule is that larger, round or oval shaped calcifications with uniform size have a higher probability of being associated with a benign process, whereas smaller, irregular, polymorphic, branching calcifications heterogeneous in size and morphology are more often associated with a malignant process.

A breast mass, on the other hand, is a localized swelling, protuberance, or lump in the breast, which usually is described by its location, size, shape, margin characteristics, and any other associated findings (i.e. architectural distortion, X-ray attenuation). Depending on a morphologic criteria, the likelihood of malignancy can be established. Figure 1.2 shows schematically different masses according to their shape and margin. Normally, a benign process is associated with the presence of circular or oval shapes, while, in contrast, spiculated masses are more probable to



Circular Shape
Circumscribed Margin

Lobular Shape
Well Defined Margin

Spiculated Shape
Ill Defined Margin

Figure 1.3: Three mass examples with different shape and margin. The last one has an increased probability to be malignant.

be the sign of a malign process. The margin refers to the border of a mass, and it should be examined carefully because it is one of the most important criteria in determining whether the mass is the result of a benign or malign process. Radiologists classify the margin among five classes [76]:

- Circumscribed margins, which are well defined and sharply demarcated with an abrupt transition between the lesion and the surrounding tissue.
- Obscured margins, which are hidden by superimposed or adjacent normal tissue.
- Micro-lobulated margins, which have small undulating circles along the edge of the mass.
- Ill-defined margins, which are poorly defined and scattered.
- Spiculated margins, which are marked by radiating thin lines.

The probability to find a malignancy mass is normally ordered according to this classification. The more ill-defined and spiculated the margin, the higher the probability to be associated with a malignant process. It should be clear that these morphological aspects can be very subtle and difficult to diagnose, even for an expert radiologist. Figure 1.3 shows three different regions of interest, including expert annotations.

1.3 Digital Mammography

As a consequence of the actual digital revolution, traditional film-based hospitals are themselves converting to digital hospitals, where patient medical records, chart information, and test results are easily available electronically for physicians from anywhere in the hospital [137]. In that sense, full-digital mammography is gaining importance compared to the nowadays still conventional film-screen mammography, due to the fact that digital acquisition, digital storage, and digital display processes may be separated and individually optimized. Let me describe in more detail these steps.

1.3.1 Image Acquisition

Digital detectors offer higher quantum efficiency and higher resolution than traditional screen-film receptors [179]. These will translate into both lower dose and improved image quality mammograms.

Digital detectors for mammography can be categorized as indirect or direct conversion detectors according to how the X-ray is captured. In indirect conversion methods (the earliest ones) the energy of the X-rays is captured by a scintillator which converts it to light. Subsequently, this light is captured by an array of thin-film diodes which convert it to electronic signal which, in turn, is captured using thin-film transistors. The main problem of these systems is that the created light by the scintillator scatters to its neighbourhood, and the same X-ray is captured by different transistors. On the other hand, in direct conversion systems the same photoconductor that captures the X-ray is able to directly generate an electronic signal. Thus, scattering is less severe in these systems.

Berns et al. [11] compared the acquisition time between screen-film and digital mammography acquisition. While the former averaged 21.6 minutes, the latter averaged 14.1 minutes, which is a highly significant 35% shorter time. However, the total interpretation time of the images by the experts averaged 1.4 minutes for screen-film mammography and 2.3 minutes for digital mammography, a highly significant 57% longer interpretation time.

1.3.2 Image Storage

Once the image is acquired it has to be stored. This is accomplished by the Digital Imaging and Communications in Medicine (DICOM) [44] standard developed by the National Electrical Manufacturers Association. This standard handles not only the storage process but also the print and transmission of the information in medical imaging.

The DICOM format differs from all the rest in the fact that it groups together the information of the patient and of the image source (the mammograph in our case) as well as the full image or images. Thus, the images coming from a single study are never mistakenly separated neither from their information. The image data can be compressed using a variety of standards, like JPEG, JPEG Lossless, JPEG 2000, and Run-length encoding (RLE). A discussion about its convenience is given in the article of Avrin et al. [5]. They conclude that the benefits of using compression techniques compared to increasing storing capabilities of the current and future technology are not that significantly clear.

Moreover, the DICOM standard also handles the communication protocol enabling, therefore, the integration of the different imaging devices of the hospital. Thus displays, scanners, servers, workstations, printers, and network hardware can be integrated in a fully digital system, usually referred as the Picture Archiving and Communication System (PACS). Typically a PACS network consists of a central server storing the DICOM database and the set of clients which provides or uses the images.

1.3.3 Image Display

Once the digital mammograph has been acquired the image and the PACS database has received and stored it, the image can be sent to the screening workspace, where usually, one or more experts will analyze and diagnose the case. However, in contrast with the typical film-screen image, the experts will have to watch the image on an electronic display (also called soft-copy display).

In contrast to (static) film reading, soft-copy offers new opportunities. For instance, there is experimental evidence that alternating the current and prior mammograms on the same display allows better evaluation of temporal changes than

conventional display of images next to each other, in particular if images are properly registered in the spatial and gray-scale domains [199]. However, the main drawback is that a faulty or inadequately calibrated or improperly set up display can compromise the overall quality of a diagnostic procedure [143].

Clinical studies show that radiologists can perform equally well with soft-copy reading of digitized screening mammograms as with reading of conventional films, where differences in sensitivity and specificity were not significant [149, 165]. Moreover, both studies show that, again, there was no significant difference in the speed of interpretation (in contrast to the already cited work of Berns et al. [11]).

1.4 Computer-Aided Systems

The idea of computer systems aiding radiologists to detect breast cancer is not recent [209]. However, the nowadays rapid development of full digital mammographic systems has being accompanied by the natural increase of such systems. A Computer-Aided System (CAD) is a set of automatic or semiautomatic tools developed to assist radiologists in the detection and/or evaluation of mammographic images [14, 54].

1.4.1 The Benefits of CAD

Back in 2001 Freer and Ulissey [54] using a database containing $12,860$ patients conclude that the use of CAD in the interpretation of screening mammograms can increase the detection of early-stage malignancies. However, results published in 2005 using a database of $6,111$ women claimed to show that CAD, in its present form, is not effective in that there was no significant difference observed in cancer detection rates with CAD compared with reading mammograms without CAD [92].

These are only two examples of an still open debate (see for instance the work of Malich et al. [116] for more examples), with people saying that actual CADs are effective and people not agree with such a statement [136]. People agree that CADs are available to detect cancers clinically missed. However the main drawback of such systems is the number of false positives obtained, which makes the radiologist not really trust them [136]. This is a major issue for the low number of malignancies

within the screening population, which is supposed to be around 6 out of 1,000 screened cases [188].

The full benefits of computer-aided systems will be seen in the future, when both the internal algorithms for abnormality detection and diagnosis and the preparation of the experts in reading mammograms with CAD will be improved. We are sure that CAD will become the standard of care.

1.4.2 Commercial CADs

Since 1998 four CAD mammographic systems have received FDA approval in the United States. The pioneer was the ImageChecker of R2 Technologies Inc. [159]. In January 2002 two new CADs were approved: the CADx Second Look and the iCAD MammoReader [82]. However, these two companies are now merged[1], and only the Second Look is still available. Finally, the last one to join this selected group has been the Kodak Mammography CAD Engine [99], in 2004. However, it is almost sure that more commercial CADs will be available in the near future.

1.5 Scope of the Research

The Computer Vision and Robotics group (VICOROB) of the University of Girona has been working in mammographic image analysis since 1996 in two main directions: the study and development of algorithms to detect and characterize micro-calcifications and clusters of them and, on the other hand, the analysis and implementation of mammographic image registration techniques.

The use of selected shape-based features was proposed in order to classify clustered micro-calcifications as benign and malignant lesions [117]. The computerized analysis of micro-calcifications was divided into four steps: 1) digitization of mammograms and enhancement of images, 2) detection and localization of suspicious areas using a region growing segmentation algorithm based on Shen proposal [174],

[1]Reuters News Service informed in 2003 "iCad Inc., which sells medical imaging systems, agreed to buy closely held rival CADx Systems Inc. and its parent company, Qualia Computing Inc., to gain new technology. (...) The companies are working on a joint coordinated basis, they said, and the deal is expected to close in two weeks".

3) extraction of shape-based features for every segmented micro-calcification, and 4) analysis of the features using case-based reasoning techniques. Moreover, and due to its demonstrated relevance for issuing a diagnosis [57], the characterization of micro-calcifications clusters was also studied. It has been observed in a great number of malignant diagnosed mammograms, that the only indicator used to issue a diagnosis was the number of micro-calcifications and their distribution inside every cluster.

Following this work, and in cooperation with the Hospital Universitari Josep Trueta of Girona and the Universitat Ramon Llull of Barcelona, the Computer Vision and Robotics group developed the HRIMAC project [119]. The main goal of this project was to develop a web-based computing tool to allow radiologists to assess the diagnosis of breast cancer from digitized mammograms. Next subsection describes in more detail the architecture of this system.

On the other hand, the group has also investigated mammographic image registration [120, 121]. Image registration is based on aligning images of the same object (taken at different times or views) in order to extract relevant information. For instance, image registration can be applied to mammographic images to assess internal breast changes. Thus, it is possible to detect the development of possible abnormalities in a particular area of interest.

1.5.1 HRIMAC Project

HRIMAC is designed as a content-based image retrieval CAD system. This kind of CADs aims to return similar previously diagnosed cases to a given new one. The analysis of these returned cases might help the radiologist in issuing his/her diagnosis. Thus, given a mammographic image, HRIMAC searches its database in order to provide the most related cases, according to some specific criteria, which are the shape of micro-calcification and its distribution in clusters.

As it is shown in Figure 1.4, the HRIMAC's architecture can be divided into two main blocks: an on-line block and an off-line block (the grey box). In the on-line block the radiologist is studying a new mammogram of a woman. Therefore, in a fast way, the expert wants/needs to find similar analyzed cases to this one. These similar cases are the basis of the off-line block, because they had been previously

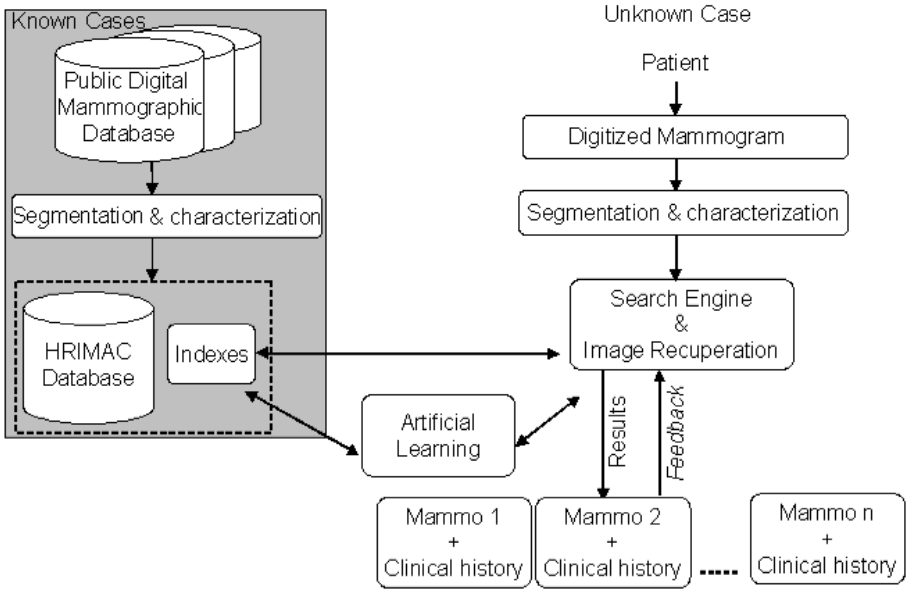


Figure 1.4: Typical architecture of a CAD system that retrieves similar diagnosed cases compared to an unknown one. The grey square symbolizes the off-line block, meanwhile the rest of figure belongs to the on-line block.

characterized, and stored into a huge database of cases.

In more detail, the system is composed by:

- **Easy Interface**. The interaction between the computer and the radiologist has to be as simple as possible. Moreover, the interface provides a large number of tools to help, for example, to detect abnormalities, to measure areas of interest, to compare mammograms, etc.

- **Mammographic Database**. A CBIR CAD system contains a database of mammograms previously diagnosed, in order to be able to return the most similar cases to the radiologist. This database constitutes the knowledge of the system. As more known (and correctly diagnosed) cases are contained in the database, the accuracy of the system is expected to improve.

- **Unknown Case**. The unknown case is the new case that the radiologist wants to diagnose.

- **Segmentation**. Segmentation refers to the detection of abnormalities in the mammogram. As shown in Figure 1.4 this is the first step in both blocks. However, while the segmentation in the on-line block has to be relatively fast, the time to segment the image in the off-line block is less important.

- **Characterization**. The characterization of mammograms is done according to the features extracted from the abnormalities. Note that in mammography, colour is not a discriminative feature, as the images are grey-level. Thus, typical used features are related with histogram information [90, 216] although texture features are nowadays gaining importance [63, 221].

- **Database of Characteristics** or HRIMAC Database. This is the database constructed from the characterization of the database of mammograms. It is required in order to reduce the computational cost of the search comparison. Furthermore, this database is indexed in order to avoid unnecessary comparisons.

- **Image Retrieval**. As it is shown in Figure 1.4, the image retrieval step is referred to the interaction between both blocks. When the new case has been characterized, this block has to find similar cases in the database of characteristics, and to return the corresponding mammograms.

- **Relevance Feedback**. The goal of this final step is to refine the retrieved result. When the system returns similar cases to the original query it may occur that not all of them are relevant. Thus, in a second step, the user indicates to the system the interesting ones. In the next retrieval, the system (hopefully) will provide more relevant images.

1.6 Objectives of the Thesis

As explained above, the *HRIMAC* project was developed as a CAD system to assess the diagnosis of breast cancer. However, in this previous work, no algorithm for mass segmentation was proposed neither studied. This was a serious drawback to make a reliable tool for the radiologists. Thus,

the main objective of this thesis is the proposal of a new framework for the detection of masses of any shape, margin, and size using a single mammographic view.

However, it should be clear that our goal does not include the characterization nor the diagnosis of the masses found.

We split the main goal of the thesis into a set of sub-objectives. In this sense, **the first sub-goal is a qualitative and quantitative study of the different proposals for mass segmentation**. From this review, we notice that there is not a single method providing the best segmentation results in all cases. Moreover, we demonstrate that the algorithms depend on the density of the breast, and also, on the shape and size of the mass.

Hence, we want to design a new mass detection tool which takes the above three parameters into account. Note that, while the density of the breast is an information which can be known a priori, the shape and size of the mass is an information known a posteriori, once the mass has been found.

We divide the construction of the detection algorithm into two different steps: firstly we will design the algorithm without the breast density information, and subsequently we will introduce this information. Thus, **the second sub-goal is the development of a new algorithm for mass detection which introduces mass shape and size information**.

The result of this algorithm will be a set of regions, some of them being really mass and others depicting only normal tissue (a false positive). In this sense, **the third sub-goal is the design of an algorithm for false positive reduction in order to increase detection accuracy**.

The following step is the introduction of the breast density into the proposed algorithms. Thus, **the fourth sub-goal is the construction of an algorithm capable of effectively classifying the breast according to its internal density** before the application of the mass detection algorithm.

Finally, **the fifth and last sub-goal is the introduction of the computed breast density information into both the mass detection and the false positive reduction algorithms**.

1.7 Thesis Outline

We summarize in the next subsection the proposed framework designed according to the mentioned objectives. Subsequently, the overview of this document is described.

1.7.1 A New Framework for Mass Detection

Figure 1.5 shows a graphical scheme of the proposed framework. The algorithm has been designed as a supervised solution, where the system begins learning a set of parameters from a database of already studied cases. These parameters are the breast tissue, the shape, and the size of the lesions. Moreover, it learns to distinguish between RoIs containing masses and RoIs of normal tissue.

When the algorithm is able to distinguish and classify the breasts/RoIs, it is ready for searching masses in a new mammogram. Roughly, the first step is to know the density of the new breast. Using this information, and the learned shape and size of the masses, it makes use of a Bayesian template matching scheme to find suspicious regions in a mammogram. As those suspicious regions can be a mass or normal parenchyma, a false positive reduction algorithm is applied. This last step is performed using the RoI discrimination learned earlier.

Note that the steps of this framework are the sub-objectives of the thesis. Thus, we have studied and developed a new algorithm to detect masses in a mammogram, another one to make the false positive reduction, and a new algorithm to classify the breasts according to their internal tissue.

We also want to mention here that the proposed template matching algorithm, as well as the false positive reduction scheme, could also be extended and applied to classify other types of mammographic abnormalities, like RoIs containing micro-calcifications or architectural distortions, or even to other types of medical diseases.

1.7.2 Document Overview

The thesis is structured according to the mentioned objectives. Thus, the following chapter is an extensive survey of the mass segmentation methods found in the literature. A qualitative and quantitative analysis is performed, extracting the corres-

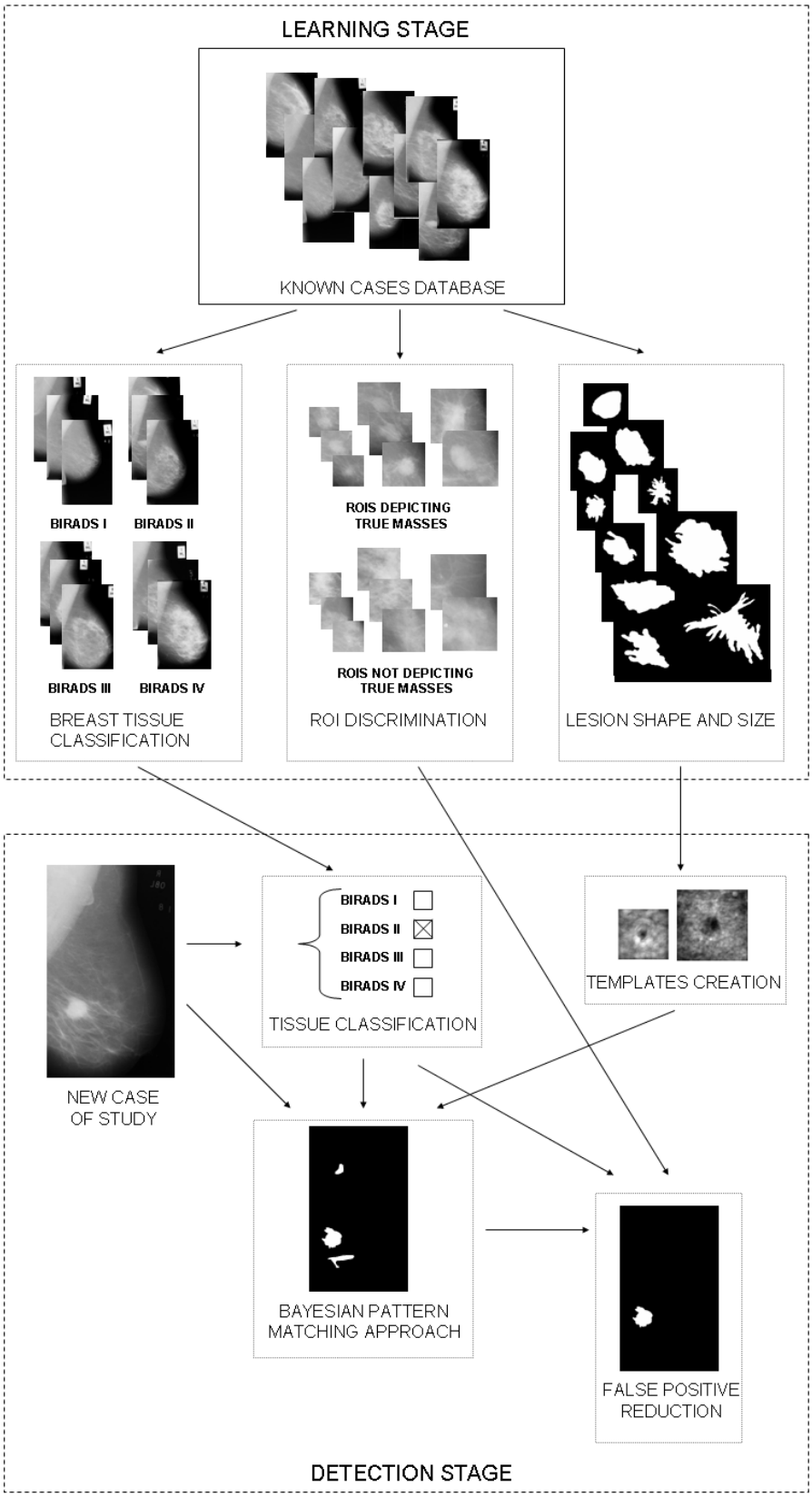


Figure 1.5: Scheme of our proposal.

ponding conclusions. As already explained, these supported the influence of breast tissue and the lesion shape and size on the performance of the algorithms.

Chapter 3 covers the classification of the breasts according to their internal tissue. It starts with a brief survey of the few existing methods dealing with this issue and subsequently, a new algorithm is proposed and exhaustively evaluated using two different databases and the opinion of three different radiologists.

Chapter 4 describes the developed algorithm for finding masses in a mammogram, without taking the breast tissue into account. Briefly, a learning stage is firstly developed in order to acquire the knowledge of the shape and the size of mammographic masses. Following, the algorithm looks for masses in a new mammogram following a template matching scheme. Results are evaluated over different databases and are given using Receiver Operating Characteristic (ROC) and Free-Response Receiver Operating Characteristic (FROC) analysis. The conclusion is that the method has large accuracy but with the penalty of obtaining a number of false positives.

Thus, in Chapter 5 a proposal for false positive reduction based on the statistical analysis of the RoIs is elucidated and extensively evaluated using different databases and ROC and FROC analysis. The evaluation is done in two steps: firstly the algorithm is evaluated using a set of manually segmented RoIs, while secondly the overall performance of the system (Bayesian template matching plus false positive reduction) is analyzed.

Chapter 6 describes how to introduce the breast tissue information into both algorithms. FROC and ROC analysis demonstrates the usefulness of such information. Finally, the thesis concludes with Chapter 7, where the conclusions and the ways in which further work is directed are covered. Moreover, in this last chapter, a list of the publications related to this thesis is included.

Moreover, three appendixes are given. The first one is related to the breast profile segmentation. Firstly, the proposal used in this work for segment the mammograms is explained. Secondly, a new algorithm which better adjusts to the external boundary of the mammogram is presented. The second appendix is focused on explaining the main characteristics of the databases of mammograms used in this work, which are four different databases in total. The last appendix explains the strategies used to evaluate the proposals.

Chapter 2

A Review of Automatic Mass Segmentation Techniques

Segmentation or abnormality detection is the initial step in mammographic computer-aided diagnosis (CAD) systems. In this chapter we review different approaches to the segmentation of mammographic masses, describing their main features and highlighting the differences among them. The key objective is to point out the advantages and disadvantages of these approaches. In contrast with other reviews, which only describe and compare different approaches qualitatively, in this chapter we also provide a quantitative comparison. Using a public database, the performance of eight relevant methods are compared using Receiver Operating Characteristic (ROC) and Free-Response Receiver Operating Characteristic (FROC) analysis.

2.1 Introduction

A segmentation algorithm, in a mammographic context, is an algorithm used to detect something, usually the whole breast or a specific kind of abnormalities, like micro-calcifications or masses. It is generally accepted that the detection of masses is technically more difficult than the detection of micro-calcifications, because masses can be simulated or obscured by normal breast parenchyma [6, 176]. Moreover, there is a large variability in these lesions, which is reflected in the morphology variation (shape and size of the lesions), and also in the large number of features that have

been used to detect and classify them.

The objective of this chapter is to review and classify the different mass detection algorithms found in the literature. The initial classification of the algorithms is done according to the number of images that the algorithms deal with. Thus, in Section 2.3 the algorithms which look for masses using a single mammogram are described and classified from a computer vision viewpoint, reflecting the strategy and the features used. It is important to notice that most of the algorithms are only able to detect a specific kind of mass, usually circular or spiculated masses. On the other hand, Section 2.4 briefly describes the algorithms which used more than one image to detect masses. They are classified according to the kind of images they used: comparison of both CC and MLO views of the same breast, comparison of the left and right breasts of the same woman, or also the temporal comparison of the same mammogram.

In Section 2.5 two key methods have been selected from each strategy and re-implemented in order to evaluate and compare their performance over the same set of MLO mammographic images, which can be clearly divided into two subsets: one containing masses and the other one with normal cases. The resulting conclusions are explained in Section 2.6.

Before starting the review of the mass segmentation strategies, however, the next section briefly describes different proposals for segmenting the breast profile. This is a necessary step in order to correctly focus the mass detection algorithms, otherwise, the annotations and background noise will introduce a huge number of outliers.

2.2 Breast Profile Segmentation

Although segmentation of the breast from the background is not the main goal of this review, this kind of segmentation is a fundamental step in mammogram image analysis because the techniques which will be discussed in Sections 2.3 and 2.4 are only applied to the breast area. For this reason, we briefly review different approaches to breast profile segmentation found in the literature.

The aim of breast profile segmentation is to separate the breast from other objects in the mammogram with a minimum loss of breast tissue. In general two independent

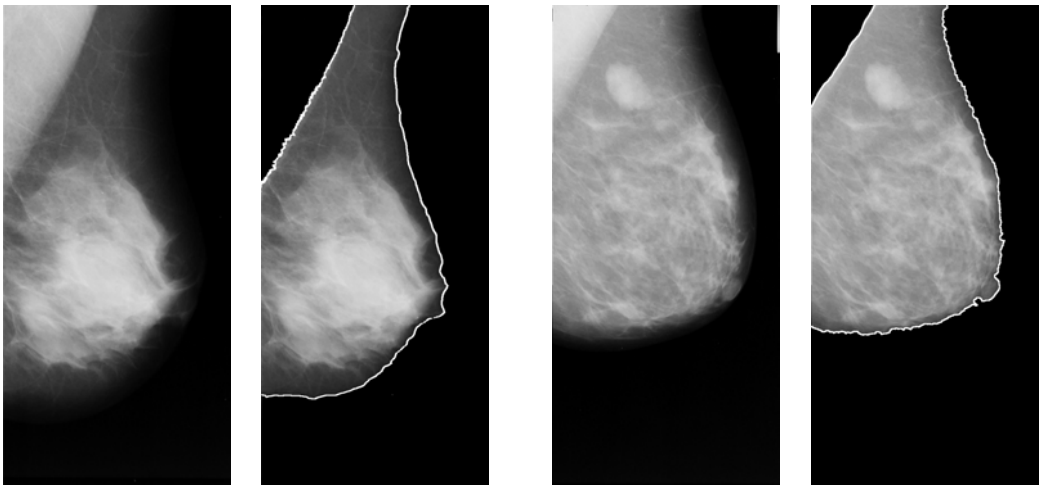


Figure 2.1: Breast profile segmentation of two mammograms using the algorithm explained in Appendix A.2, which is the used in this work.

steps are performed. The first one aims to segment the background and annotations from the whole breast area, while the second one involves separating the pectoral muscle (when present) from the rest of the breast area. Approaches to the breast-background segmentation range from simple histogram thresholding followed by smoothing [73, 107] to polynomial modeling [28], or, more recently, active contour approaches [51]. Typical strategies to segment the pectoral muscle have been based on straight line estimation using a Hough transform [89, 104] or direct detection using Gabor filters as edge detectors [52].

In this work, a new approach has been designed for such task. Firstly, an automatic thresholding algorithm is used to separate the area composed of the breast and the pectoral muscle from the background of the image. Subsequently, a region growing algorithm allows to locate the muscle and extract it from the breast. The result of this approach is a visually correct segmentation of the breast, although some pixels belonging to the skin-line are misclassified as background. However, this set of pixels is not relevant in posterior steps. Figure 2.1 shows two examples of the performance of these algorithms. This approach is described in detail in Appendix A.2.

To avoid the misclassifications of the pixels with low grey-level value located near to the skin-line a new algorithm has also been designed applying edge detection and

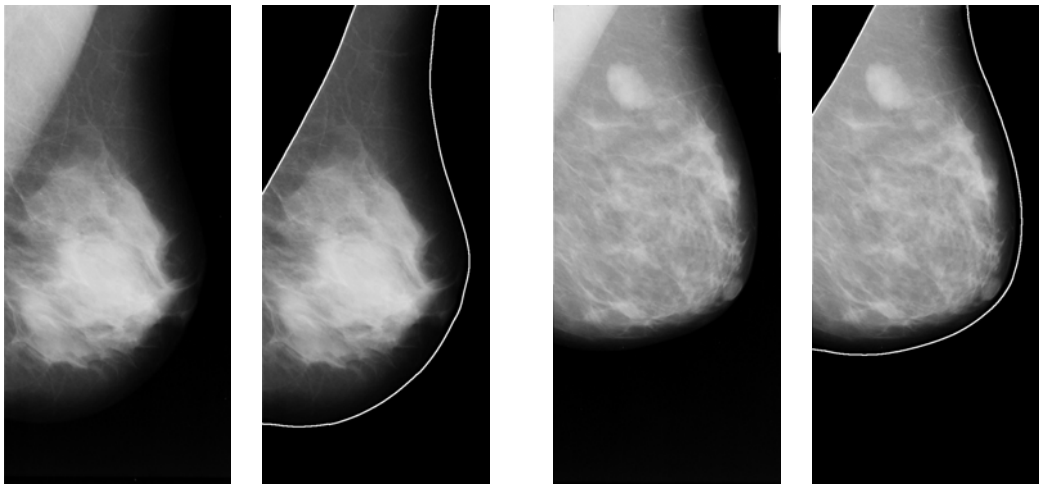


Figure 2.2: Breast profile segmentation of two mammograms using the algorithm described in Appendix A.3. This algorithm segments all pixels in the breast, although the boundary seems far away of it.

scale space concepts. Thus, the main edges of the image are correctly located and subsequent post-processing is used to isolate the skin-line. Figure 2.2 shows the performance of this algorithm in combination with the pectoral muscle segmentation proposed by Ferrari et al. [52]. Appendix A.3 gives more details of this skin-line detection algorithm.

2.3 Mass Segmentation Using One Single View

Segmentation using a single mammographic image relies on the fact that pixels inside a mass have different characteristics from the other pixels within the breast area. The characteristics used can be simply related to intensity values and to (local) texture measures, or otherwise related to morphological features like the size and margins of the mass, as Figure 1.2 schematizes. In addition, some approaches take the distribution of spicules associated with masses into account. Both aspects can be treated independently or sequentially. The columns of Table 2.1 classify proposed methods according to the characteristics they use for the segmentation, while the properties used in the optional subsequent classification processes (benign/malignant discrimination) are not taken into account.

Segmentation techniques can be divided into supervised and unsupervised approaches. **Supervised segmentation**, also known as model-based segmentation, relies on prior knowledge about the to be segmented and optional background regions. The prior information is used to determine if specific regions are present within an image or not. **Unsupervised segmentation** consists of partitioning the image into a set of regions which are distinct and uniform with respect to specific properties, such as grey-level, texture or colour. Classical approaches to solve unsupervised segmentation are divided in three major groups [58]:

- **Region-based methods**, which divide the image into homogeneous and spatially connected regions.
- **Contour-based methods**, which rely on finding the boundaries of regions.
- **Clustering methods**, which group together those pixels having similar properties and might result in non-connected regions.

The rows of Table 2.1 show the reviewed works arranged according to their classifications. In subsequent subsections, a detailed description of the methods in each category (region-based, contour-based, clustering, and model-based methods) is given.

2.3.1 Region-Based Methods

Region-based segmentation relies on the principle of homogeneity, which means that there has to be at least one feature which remains uniform (plus/minus an error measure) for all pixels within a region. Two basic strategies of region-based methods are the well known *region growing* and *split and merge* approaches.

2.3.1.1 Region Growing and Related Methods

More than 30 years have already passed since Zucker reviewed region growing algorithms [218]. Region growing is based on the propagation of an initial seed point according to a specific homogeneity criterion, iteratively increasing the size of the

		Author	Year	Texture	Gradient	Grey-Level	Shape
Unsupervised Segmentation	Region	Huo [81]	1995			√	√
		Rangayyan [162]	1997			√	
		Guliato [65]	1998			√	
		Kupinski [102]	1998		√		√
		Qi [151]	1998			√	
		Petrick [145]	1999		√	√	
		Kwok [103]	2002	√			
		Zheng [214]	2003			√	√
		Wei [206, 207]	2005		√	√	
	Contour	Kobatake [97]	1996			√	
		Petrick [147, 146]	1996		√	√	
		Kobatake [96, 95]	1999		√	√	
		Sahiner [170, 168]	2001		√	√	
		Timp [193]	2004		√	√	
	Clustering	Brzakovic [21]	1990			√	
		Kobatake [98, 94]	1994			√	
		Li [111]	1995	√		√	
		Sahiner [169, 167]	1996		√	√	
		Undrill [197]	1996	√	√	√	
		Chen [30]	1997	√			
		Matsubara [125, 124]	1998			√	
		Goto [64]	1998		√	√	
		Qian [153, 154, 152]	1999	√			
		Heath [72]	2000	√			
		Velthuizen [201]	2000			√	
		Mudigonda [132]	2001		√	√	
		Saha [166]	2001			√	
		Zheng [215]	2001	√			
		Li [112]	2002	√		√	
		Hassanien [69]	2004			√	
		Catarious [26]	2004			√	
		Bellotti [9]	2006			√	
Supervised Segmentation	Model	Lai [105]	1989			√	√
		Stathaki [181]	1994	√			
		Karssemeijer [89]	1996		√		
		Polakowski [150]	1997		√	√	
		Jiang [86]	1998		√		√
		teBrake [190, 191]	1999		√	√	
		Constantidis [37, 38]	1999				√
		Zwiggelaar [223]	1999		√	√	
		Hatanaka [70]	2001		√	√	
		Li [110]	2001	√		√	
		Baydush [7]	2003			√	
		Tourassi [195]	2003			√	
		Campanini [25]	2004			√	
		Cheng [32]	2004	√			

Table 2.1: Single image based mammographic segmentation techniques indicating underlying (region, contour, clustering or model based) approach and used feature information (texture, gradient, grey-level or shape).

region. Since those early days, region growing has received a number of improvements, mainly due to the integration of boundary information in the algorithm. As Freixenet et al. [56, 134] reviewed, this information can be introduced before the growth step, using for example a controlled seed placement [10] or during it, like in active region algorithms [55, 217].

Region growing algorithms have been widely used in mammographic mass segmentation. Since the early nineties, people from the University of Chicago studied the introduction of shape information into the homogeneity criterion. With the aim to integrate the radiologists experiences, Huo et al. [81] developed a semi-automatic region growing approach, in which the growing step was automatically computed after a radiologist had manually placed the seed point. Later, Kupinski and Giger [102] compared this initial approach with two improved versions. The first one incorporated the Radial Gradient Index, which is a measure of the average proportion of gradient which are radially directed outwards (for a circular region the radial gradient index is equal to one). The second one was based on a probabilistic method in which the probability of belonging to one region was modeled by a non-Gaussian distribution (using a kernel distribution), whilst the background was modeled using a uniform probability. They showed that this last parametric version performed better compared to both other approaches. A different approach was proposed by Guliato et al. [65], who implemented a fuzzy version of the region growing algorithm. In contrast with the methods from Chicago, this approach was pixel based (the homogeneity criterion is evaluated for each pixel), and no prior shape information was considered. Guliato et al.'s method was based on considering the uncertainty present around the boundaries of a tumour region, with the aim to preserve the transition between mass and normal tissue. An alternative approach was proposed by Petrick et al. [145], who introduced gradient information into the region growing algorithm with the objective to reduce merging between adjacent and overlapping structures. Initially, the algorithm selects seeds using local maxima in the grey-scale image. In a subsequent step, a gradient image is constructed by using a frequency-weighted Gaussian filtering. With this image, the thresholds of the regions bounded by the edges are extracted. In common with the Chicago approach, Petrick et al. aggregate groups of pixels with similar characteristics (using thresholds), but, in contrast, they do not use shape information in the homogeneity criterion.

Other researchers spent their efforts improving the region growing algorithm by identifying the optimal set of initial seeds. Qi and Snyder [151] used Bézier splines to interpolate histograms, from which they extracted threshold values at local maxima. Zheng et al. [214] used as starting point an edge image. This image was obtained by subtracting two blurred images obtained by using Gaussian filtering of the original image using a large difference in kernel size.

2.3.1.2 Split and Merge Methods

The split and merge technique [31] is the other classical region-based segmentation method. As the name indicates, the process consists of recursively splitting the image until all regions confirm to a homogeneity criterion. In an accompanying step, all adjacent regions satisfying a second homogeneity criterion are merged. However, to our knowledge, for mammographic mass segmentation, this approach has only been used by Rangayyan et al. [162] who, beginning with a hand-selected region of interest containing a single mass, used this technique to approximate its boundary by polygons.

2.3.2 Contour-Based Methods

Image segmentation techniques based on edge detection have been in use since the early work of Roberts [164]. However, identifying regions on the basis of edge information is far from trivial, since algorithms for edge detection do not usually possess the ability of the human vision system to complete interrupted edges using experience and contextual information. Therefore, sometimes edges are detected which are not the transition from one region to another and correctly detected edges often have gaps in them at places where the transitions between regions are not abrupt enough. Hence, detected edges may not necessarily form a set of closed connected curves that surround distinct regions.

As indicated in Table 2.1, there is only a limited number of publications trying to segment mammographic images using edge-based methods, which is mainly due to the difficulty of extracting the boundary between masses and normal tissue. Typical algorithms for finding edges are based on filtering the image in order to enhance relevant edges prior to the detection stage. The earliest approaches for mass

segmentation are based on such methodology. The detection of edges in Petrick et al. [146, 147] was based on a Gaussian-Laplacian edge detector, after the image was enhanced by the adaptive density-weighted contrast enhancement filter. A different approach is described by Kobatake and Yoshinaga [97], which starts with a sub-image containing a possible mass lesion. The algorithm looks for spicules using gradient information in three steps: firstly, the morphological line-skeletons are extracted in order to detect long and thin anatomical structures (like spicules). Secondly, a modified Hough transform is used to extract lines passing near the centre of the mass, and finally the algorithm automatically selects candidates based on the number of line-skeletons that satisfy the second step.

In recent approaches, edge information has been used to refine initial segmentation results. Examples are Kobatake et al. [95] and Sahiner et al. [168, 170], who used active contour models (snakes) as a final step of their algorithms. Timp and Karssemeijer [193] found the best contour of the mass by an optimization technique based on dynamic programming. Their approach used both edge based information as well as a priori knowledge about the grey-level distribution of the region of interest around the mass. They demonstrated a better performance of their method in comparison with an implemented version of the region growing algorithm inspired on the already mentioned work of Kupinski and Giger [102], and the discrete contour model inspired on the work of Viergever and Lobregt [203].

There are related approaches, which are based on the detection of spicules and the use of statistical analysis of gradient-orientation maps [89, 191]. However, due to the necessity to perform a posterior classification step we consider these as model-based approaches.

2.3.3 Clustering and Thresholding Methods

Clustering methods are one of the most commonly used techniques in image segmentation, as discussed in the review by Jain et al. [83]. Based on this work, clustering techniques can be divided into hierarchical and partitional algorithms, where the main difference between them is that *hierarchical methods* produce a nested series of partitions while *partitional methods* produce only a single one. Although hierarchical methods can be more accurate, partitional methods are used in applications

involving large data sets, like the ones related to images, because the use of nested partitions becomes computationally prohibitive. However, partitional algorithms have two main disadvantages: 1) the algorithm has to know, a priori, the number of clusters that are in the image, and 2) the fact that clustering algorithms do not use spatial information inherent to the image.

A traditional partitional clustering algorithm is the k-Means algorithm [114], which is characterized by easy implementation and low complexity. For mass segmentation purposes, this algorithm has been used by Sahiner et al. [167, 169] in order to generate an initial segmentation result. As we have described in the above section, Sahiner et al. improved this segmentation using edge information. In contrast, Li et al. [112] used a generalization of k-Means that included spatial information to refine an initial segmentation (the initial result is achieved by using adaptive thresholding).

Alternative clustering algorithms used for mammographic mass segmentation are the Fuzzy C-Means (FCM) algorithm [12], the Dogs and Rabbit (DaR) algorithm [126] and the Expectation Maximization (EM) algorithm [42]. FCM was used with different objectives in the works of Velthuizen [201] and Chen and Lee [30]. While Velthuizen used it to group pixels with similar grey-level values in the original images, Chen and Lee used it over the set of local features extracted from the application of a multi-resolution wavelet transform and Gaussian Markov random fields analysis. Moreover, the output of the FCM was the input to an EM algorithm based on Gibbs random fields. On the other hand, the DaR algorithm was used by Zheng and Chan [215], and in contrast to FCM which improves k-Means using a fuzzy approach of the energy function, this algorithm performs a more robust seed placement, resulting in a stable clustering algorithm [126]. Other clustering approaches are based on prior assumptions (models) of the image, as for example, the algorithm of Li et al. [111] which is similar to the EM approach proposed by Chen and Lee [30], but they formulated the segmentation as a Markov Random Fields model. Like the k-Means, their algorithm is iterative and alternates between estimating the mean intensity and the pixel label.

One of the earliest approaches to mass segmentation was the work of Brzakovic et al. [21] which was based on a multi-resolution fuzzy pyramid linking approach, a data structure in which the input image formed the basis of the pyramid and

each subsequent level (of lower resolution) was sequentially constructed. The links between each node and its four parents were propagated using a fuzzy function to upper levels. They demonstrated that this algorithm was directly correlated with the isodata clustering algorithm [21]. It has to be noted, that with this strategy, spatial information (region information) is taken into account.

Thresholding Methods

Like Fu and Mu [58], we consider threshold methods as a special case of partitional clustering methods, where only two clusters are considered. Threshold methods have been widely used for mass segmentation. For instance, Matsubara et al. [124, 125] used different grey-level threshold values depending on the type of tissue of the breast based on histogram analysis. More recently, Mudigonda et al. [132] used multilevel thresholding to detect closed edges. In this approach a concentric group of contours represents the propagation of density information from the central-core portion of an object or tissue region in the image into the surrounding tissues. This algorithm can be regarded as a region growing approach, where in each iteration neighbours with similar grey-level values are grouped (the works of Huo et al. [81], as well as Petrick et al. [145], described in subsection 2.3.1, follow a similar strategy). The main drawback of this approach is the assumption that masses have (more or less) uniform density compared to the local background.

In some cases the thresholding is not applied directly to the mammographic image, but to an enhanced version of the original image. For example, Kobatake et al. [94, 98] applied an iris filter designed to enhance rounded opacities and to be insensitive to thin anatomical structures. Using adaptive thresholding they detected round masses. Another example is by Saha et al. [166], who first enhanced the image by a scale-based fuzzy connectivity method, and subsequently thresholded the image to detect masses.

Instead of enhancing the image, a different approach is to first extract some (texture) features from the image and threshold them in a posterior step. For instance, Undrill et al. [197] thresholded images using Laws masks, while Heath and Bowyer [72] developed a new mass detection algorithm which was based on the use of an Average Fraction Under the Minimum (AFUM) filter. This filtering is designed to find the degree to which the surrounding region of a point radially decreases in

intensity. The final step is to threshold the image to identify suspicious regions.

2.3.4 Model-Based Methods

As model-based segmentation we include those methods that firstly train the system to detect specific objects. Subsequently, the system has to be able to detect and classify new images depending on the presence or absence of the object. Model-based segmentation methods are closely related to general pattern recognition problems such as pattern matching. In pattern matching, the training is based on images containing the object to detect. Pattern matching has been used in segmentation of mammographic images by Lai et al. [105] and by Constantinidis et al. [37, 38]. The main drawbacks of these approaches are the large variation in the shapes of masses and the use of a synthetic model to find real masses. Recently, Tourassi et al. have improved the performance of such approaches by using mutual information as a similarity metric to determine if a (real) query RoI depicts a true mass [195].

Most of the mass segmentation model-based methods train the system on gradient information. Due to the training step we do not classify such approaches as being pure edge-based methods. Training images cover examples with and without the object present. Thus:

- From images containing the object, the system learns the probable location of the object and the variation in shape and size of the object, a mass in our case.

- From images not containing the object, the system can learn features that represent normal tissue.

Based on both training aspects the system learns what features to look for when presented with a new image. One of the early approaches using this strategy was the work of Karssemeijer and te Brake [89]. They first detect spicules using second order Gaussian derivatives operators. If a line-like structure is present at a given site, the method provides an estimation of the orientation of these structures, whereas in other cases the image noise will generate a random orientation. With this information they constructed two new features that formed the input for the

classification stage. In more recent work [190, 191], they improved this algorithm using a multi-scale approach.

Other approaches using edge information are by Jiang et al. [86] and Polakowski et al. [150]. The former was based on the enhancement of the spicules using morphological operations and, subsequently, two features representing the concentration of spicules were used to train a classifier based on a discrimination function. In the work of Polakowski et al. [150], the edges of the image were extracted by subtracting two smoothed versions of the original mammogram. Two Gaussian filters with different standard deviation were used.

On the other hand, Zwiggelaar et al. [223] and Li et al. [110] used statistical approaches to model masses. The former detected spiculated lesions by the union of two techniques: the first one models the centre of the mass using a directional recursive median filter, while the second one models the surrounding pattern of linear structures applying a multi-scale directional line detector. The combination of both methods results in a probability image. Therefore, the detection is performed by thresholding these resulting probability images. On the other hand, Li et al. applied firstly an image enhancement algorithm using morphological filtering. Subsequently, they employed a finite generalized Gaussian mixture (FGGM) distribution to model the histogram. They incorporated an EM algorithm to determine the optimal number of image regions and the kernel shape in the FGGM model. The final step was the use of a Bayesian relaxation labeling approach to perform the selection of suspicious masses.

Three recent approaches [7, 25, 32] are based on using neural network classifiers and formulate the problem of segmentation as a classification of RoIs being suspicious or not. The features used for the training step are a set of known RoIs containing masses and a set of random samples from normal tissue.

2.4 Mass Segmentation Using Two or More Images

The comparison of different mammographic images of the same woman is a commonly used practice in breast radiology. Usually, this comparison is between:

- Left and right mammograms.
- Two mammographic views (CC and MLO) of the same breast.
- Same view mammograms taken at different times.

This practice is justified from a clinical point of view by many studies. Kopans, for example, makes two important observations when comparing different mammograms of the same woman [101]:

- Though one breast may be larger than the other, internal structures are quite symmetric over broad areas.
- Overlapping tissue structures that form summation shadows and normal tissue variations on the mammogram highlight unimportant asymmetries.

In order to distinguish masses and asymmetric breast tissue, clinicians take a range of characteristics of abnormal areas, such as size, density, and shape into account.

The following subsections provide a more detailed description of the various approaches to mass segmentation from multiple mammographic images.

2.4.1 Comparison of Left and Right Mammograms

The comparison between left and right breasts is based on the fact that both breasts have similar internal structure, as it is shown in Figure 2.3, where mammograms corresponding to left and right breasts of a woman are displayed as radiologists are used to. A thoroughly evaluated method for comparing mammographic images is known as *bi-lateral subtraction*. Both images are first aligned and subsequently subtracted. The alignment of both breasts is the critical component in this process. There are two main approaches to this:

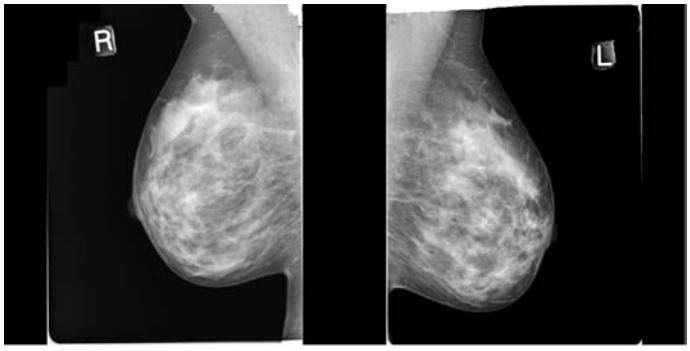


Figure 2.3: Right and left mammograms of a woman.

- The use of anatomical features, like the position of the nipple, regions or assumptions over the compression of the breasts [4, 33, 60, 128, 129, 163, 213].
- The use of the breast profile for aligning both breasts [61, 90].

The advantage of the former techniques over the latters is that the distortion of the internal structures is taken into account, thus registering correctly the breast interior. It should be noted that imaging conditions (breast compression and exposure, X-rays energy, etc.) are likely to differ for the left and right mammograms.

There are alternatives to this alignment and subtraction approach. Kok-Wiles et al. [100] represent the bright zones of the breast as a tree-like-structure, which forms the basis for the comparison process. Stamatakis et al. [180] developed a *Multiple Image Comparison* approach. This method starts with one pair of mammograms, then the system computes eight new pairs of images using a set of grey-level features. Finally each of those pairs is bilaterally compared with each other and the resulting difference images are recombined into a final result.

2.4.2 Comparison of Two Mammographic Views

The comparison of two views of the same breast is known as *ipsilateral comparison*. Figure 2.4 shows CC and MLO views of the same breast. There are two main strategies depending on when the segmentation is performed: after or before establishing correspondence between both views. Sun et al. [184, 185] segmented both views independently and subsequently compared the results. In contrast, Highnam

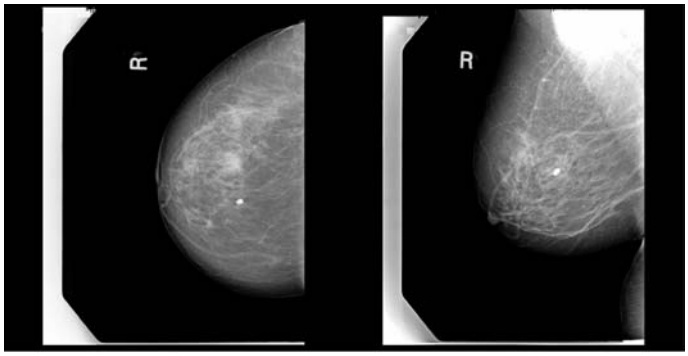


Figure 2.4: CC and MLO views of the same breast.

et al. [78] proposed an alternative approach, consisting of initially obtaining correspondence and subsequently segmenting the images. Specifically, they proposed a simple breast acquisition model where they draw a point in one view and map this to a line in the other view. Local information is used to reduce the line to a point, to obtain a point to point correspondence between views.

2.4.3 Temporal Comparison of Mammograms

Finally, a comparison can be performed between mammograms of the same breast but taken at different time intervals. Thus, with such technique, radiologists can evaluate how a suspicious region has evolved. Sanjay-Gopal et al. [172] first identify regions of interest on the most recent mammogram. Subsequently, the nipple location is used to align mammograms and locate the segmented region in the previous mammogram. Vujovic and Brzakovic [204] proposed an algorithm for identifying the potential control points (for instance, crossing points) and establishing the correspondence between them. Using a similar methodology, Martí et al. [121, 122] extracted salient control points from internal linear structures in order to establish the correspondence. The latter provides a more robust approach to the extraction of control points using features such as orientation and width of the structures, instead of just their positions.

Strategy	Id	Method	Based on
Region	$a1$	Region Growing	Petrick et al. [145]
	$a2$	Region Growing	Kupinski and Giger [102]
Contour	$b1$	Laplacian	Petrick et al. [146, 147]
	$b2$	Skeletons	Kobatake and Yoshinaga [97]
Clustering	$c1$	K-Means	Sahiner et al. [167, 169]
	$c2$	Fuzzy C-Means	Velthuizen [201]
Model	$d1$	Pattern Matching	Lai et al. [105]
	$d2$	Contour + Classifier	Karssemeijer et al. [89]

Table 2.2: Compared mass segmentation methods.

2.5 Evaluation of Mass Segmentation Methods

In Section 2.3, mass segmentation approaches using one single view have been qualitatively analyzed and classified as region, contour, clustering, or model-based methods. However, based only on this analysis, we can not estimate the robustness of the algorithms with respect to different mammographic cases (different breast tissue, different lesion types, etc.). Therefore, in order to study how the performance varies for the different strategies, and with the aim to extract reliable conclusions, we have quantitatively compared the different approaches. We have selected and implemented two of the most representative algorithms of each class (see Table 2.2).

The algorithms have been evaluated on a set of 120 mammograms from the MIAS database [183]. Of this set, 40 mammograms contained masses, whilst the other 80 represented normal cases. A comparison of the performance between all the algorithms is provided to determine their main advantages and constraints. The performance has been analyzed using ROC and FROC curves [130].

2.5.1 Evaluated Mass Segmentation Methods

As shown in Table 2.2, each selected method is symbolized using the first letter of the class and a number to distinguish between each algorithm. This description will be useful when presenting results and comparing algorithms. The following subsections provide a detailed description of the selected algorithms.

2.5.1.1 *a1*: Region Growing Including Gradient Information

This algorithm is inspired by recent work of Petrick et al. [145]. Similar to their other earlier work [146, 147], the algorithm starts by preprocessing the image using a Density-Weighted Contrast Enhancement (DWCE) filter. This filter is based on two filtered images of the original mammogram $F(x, y)$:

- A density image $F_D(x, y)$, which is an smoothed version of the mammogram obtained by applying a Gaussian filter.
- A contrast image $F_C(x, y)$, which is found by subtracting the original image and another smoothed version of this image, obtained using a different σ value.

The density image is filtered again using a non-linear filter K_M and used to define a multiplication factor which modifies the corresponding pixel in the contrast image. This way, it allows the local density value of each pixel to be weighted by the local contrast. This intermediate image referred to as F_{KC} can be analytically described as:

$$F_{KC}(x, y) = K_M(F_D(x, y)) * F_C(x, y) \tag{2.1}$$

where $*$ denotes the convolution operation. This image is used to define a second multiplication value using another non-linear filter K_{NL}, which is multiplied again by the weighted contrast of the corresponding pixels:

$$F_E(x, y) = K_{NL}(F_{KC}(x, y)) * F_{KC}(x, y) \tag{2.2}$$

The resulting image $F_E(x, y)$ is the output of the DWCE filter.

The main aim of preprocessing is to enhance possible mass lesions in the image. Once the image is filtered, morphological erosion techniques [40] are used to identify local maxima, which are the seeds of a subsequent region growing algorithm which is used to expand them using grey-level and gradient information. The gradient image is obtained using Frequency-Weighted Gaussian (FWG) filtering, which is based on the following decomposition:

$$F(x, y) = F_F(x, y) + F_{sub^+}(x, y) + F_{sub^-}(x, y) \tag{2.3}$$

where $F_F(x,y)$ is a smoothed version of the image $F(x,y)$ resulting from the application of a Gaussian filter with mean 0 and standard deviation 10, and

$$F_{sub^+}(x,y)\begin{cases} F(x,y)-F_F(x,y), & F(x,y)>F_F(x,y) \\ 0, & \text{Otherwise} \end{cases} \tag{2.4}$$

$$F_{sub^-}(x,y)\begin{cases} F_F(x,y)-F(x,y), & F(x,y)<F_F(x,y) \\ 0, & \text{Otherwise} \end{cases} \tag{2.5}$$

This filtering is repeated twice. The first iteration reduces the gradients within the breast, whilst the second one eliminates gradients in the background. Hence, the result of this decomposition is an enhancement of the contrast between the breast structures and the background. Subsequently, applying a Sobel filter produces the gradient image of the original mammogram with a significant amount of background eliminated. Finally, as a result of this additional information, the region growing algorithm has a limited number of regions to grow.

2.5.1.2 *a2*: Probability Driven Region Growing

Kupinski and Giger [102] compared three region growing approaches: traditional region growing using only grey-level information, region growing using gradient information, and region growing using probability information based on grey-level. The latter approach provided the best results and therefore algorithm *a2* is based on that approach.

In the *a2* algorithm, the probability of pixel grey-levels given a partition L_i is modeled as:

$$P(f(x,y)\mid L_i,\sigma_l^2)=\begin{cases} N(f(x,y);f(\mu_x,\mu_y),\sigma_l^2), & (x,y)\in L_i \\ z(f(x,y)), & (x,y)\notin L_i \end{cases} \tag{2.6}$$

where $N(f(x,y);f(\mu_x,\mu_y),\sigma_l^2)$ is a normal distribution centred at the seed point grey-level $f(\mu_x,\mu_y)$ and variance σ_l^2, whereas $z(f(x,y))$ is a function estimated for each breast using the grey-levels of all its pixels. This is based on kernel density estimation, which is an extension of histogram analysis. An Epanechnikov kernel [127]

is used to estimate the pixel distribution. The probability of the image given a partition L_i is:

$$p(I \mid L_i, \sigma_l^2) = \prod_{(x,y)\in I} p(f(x,y) \mid L_i, \sigma_l^2) \tag{2.7}$$

The partition L_i used will be the one that maximizes this probability, i.e:

$$p(I \mid L_{final}, \sigma_l^2) = argmax_i\{p(I \mid L_i, \sigma_l^2)\} \tag{2.8}$$

Although Kupinski and Giger [102] applied this method to a Region of Interest (RoI) manually selected by an expert, we have slightly modified it with the aim to apply it to whole mammograms. We have automatically placed seeds over the image at high-intensity regions, following an approach proposed to detect micro-calcifications [118].

2.5.1.3 *b1*: Laplacian Edge Detector

The *b1* algorithm is inspired by Petrick et al. [146, 147], who used a Laplacian edge detector in order to find closed regions in an enhanced image of the mammogram. Similar to the *a1* algorithm, the DWCE filter is applied as a enhancement process to highlight potential masses. Applying an edge detector makes it possible to select these regions.

2.5.1.4 *b2*: Based on Skeleton Approach

Kobatake and Yoshinaga [97] used skeleton analysis to detect spicules. The *b2* algorithm first extracts the skeleton in regions of interest using different structuring elements to enhance lines in different directions. Subsequently, *b2* detects the main lines of the skeleton using a variation on the Hough transform. Using the skeleton information and the main lines, *b2* calculates the probability that pixels belong to a mass. Although in the original work the authors proposed an iris filter in order to enhance the image and automatically detect the regions of interest [98], we have used the same philosophy of the use of skeletons over the whole image.

2.5.1.5 *c1*: k-Means

The popular k-Means clustering algorithm, first proposed by MacQueen [114], is an error-based minimization algorithm, where the minimizing function is the sum of squared error:

$$e^2(I, \Xi) = \sum_{k=1}^{K} \sum_{i \in C_k} ||p_i - c_k||^2 \tag{2.9}$$

In this equation, Ξ represents the partition of the image I, c_k is the centroid of cluster C_k, and p_i is each pattern of the image (each pixel). Two factors have made the k-Means one of the most popular clustering algorithms: it has linear time complexity and it is easy to implement [83].

In mammography, the k-Means algorithm has been applied by Sahiner et al. [167, 169], who used the intensity of the pixels as features. Hence, the suspicious regions will be those regions with higher average grey-level. In our implementation the algorithm works with additional features. The aim of the first one is to prevent disconnected regions and, as suggested Jain et al. [83], we use a smoothed version of the original mammogram. In addition, we have included texture features derived from co-occurrence matrices [68] and Laws filters [108]. From co-occurrence matrices, for distances one to five and angles 0°,45°,90° and 135°, the following statistics have been extracted: contrast, energy, entropy, and homogeneity. The other texture features are based on Laws energy filters of size five.

As has been discussed in Section 2.3.3, the k-Means approach starts by randomly selecting a pre-determined number of seed points. In our experiments, this number can vary from 5 to 25. However, we have observed that best performances are reached when over-segmenting the images. In such cases, the location of a mass is indicated by concentric regions of decreasing intensity.

2.5.1.6 *c2*: Fuzzy C-Means

The Fuzzy C-Means (FCM) clustering algorithm [12] is an extension of the k-Means algorithm. The main difference is that Fuzzy C-Means allows each pattern of the image to be associated with every cluster using a fuzzy membership function (in contrast with k-Means, where each pattern belongs to one and only one cluster). The introduction of such fuzziness has two effects [83]:

- The FCM algorithm has a poor convergence performance compared to k-Means.
- The FCM, in general, has an improved segmentation performance.

In our implementation, the function criterion minimized by the algorithm is defined by:

$$e^2(I, \Xi) = \sum_{k=1}^{K} \sum_{i=1}^{N} u_{ik}^m ||p_i - c_k||^2 \tag{2.10}$$

where u_{it} represents the membership of pattern p_i to belong to cluster k, which is centred at

$$c_k = \frac{\sum_{i=1}^{N} u_{ik}^m p_i}{\sum_{i=1}^{N} u_{ik}^m}, \tag{2.11}$$

N is the number of patterns in the whole image (i.e. the number of pixels), K the number of clusters, which has to be known a priori, and m the degree of fuzzyness (a number greater than 1).

For mammographic mass segmentation purposes, the Fuzzy C-Means algorithm has been applied by Velthuizen [201] and Chen and Lee [30] who again only used grey-level features. As for the *c1* approach, the *c2* implementation is based on additional features. To perform a realistic comparison, we used the same features for both algorithms.

2.5.1.7 *d1*: Pattern Matching Approach

Pattern matching starts by defining a template, in our case, a tumour-like template. The definition of the template is based on the approach of Lai et al. [105], who defined the tumour by three characteristics: brightness contrast, uniform density and circular shape. In our implementation, the template can vary between 3 and 200 pixels in diameter. Figure 2.5 shows a 5-pixel radius template. The circular patch of ones in the centre represents a tumour area having uniform density. The ring of zeros represents the "don't care" area to account for some of the shape variability. Finally, the outer edge of the template is filled with minus ones to represent the dark background. One of the drawbacks of this algorithm is its poor performance in detecting spiculated masses [105].

			-1	-1	-1			
		-1	-1	0	-1	-1		
	-1	-1	0	1	0	-1	-1	
-1	-1	0	1	1	1	0	-1	-1
-1	0	1	1	1	1	1	0	-1
-1	-1	0	1	1	1	0	-1	-1
	-1	-1	0	1	0	-1	-1	
		-1	-1	0	-1	-1		
			-1	-1	-1			

Figure 2.5: A tumour-like template for matching with tumours of five pixels in diameter [105].

In contrast with the original work, where the authors used a cross-correlation metric to measure the similarity among the image patches and the template, in this work we used a mutual information based metric. This similarity measure was inspired on the work of Tourassi et al. [195], where they used it to retrieve similar RoIs in a CBIR system. As shown in [139], the results obtained using this probabilistic metric outperforms the ones obtained using the cross-correlation metric.

Given two images A and B, the mutual information is expressed as:

$$MI(A,B) = \sum_{x}\sum_{y} P_{AB}(x,y) \cdot log_2(\frac{P_{AB}(x,y)}{P_A(x)P_B(y)}) \tag{2.12}$$

where $P_{AB}(x,y)$ is the joint probability of the two images based on their corresponding pixels values and $P_A(x)$ and $P_B(y)$ are the marginal probabilities of the variables x and y which are the image pixel values, and are obtained from the corresponding normalized histograms. To obtain a compatible template we calculated the mean of all pixels in the breast. Subsequently, in the template, -1's are replaced with pixels values inferiors to the mean, 0's with the value of the mean, and 1 with values superiors to it.

2.5.1.8 *d2: Classifier Approach*

The final approach uses classification and has been inspired by the work of Karssemeijer and te Brake [89]. The *d2* algorithm finds possible masses from the detection of spicules using second order Gaussian derivatives operators. If a line-like structure

is present at a given site, the method provides an estimation of the orientation of these structures, whereas in other cases the image noise will generate a random orientation. With this information two new features are constructed. The first feature represents the total number of pixels pointing towards the centre, while the second one estimates if these directions are circularly oriented. With these two features and a set of classified mammograms *d2* trains a binary decision tree. Subsequently, the decision tree can be used in segmenting unseen mammograms.

2.5.2 Evaluation Methodology: ROC and FROC Curves

In mammography, the most common evaluation methodology is to compare the results obtained by the algorithms to those obtained by a set of experts, which is considered as ground-truth.

The performance of computer-based detection techniques can be summarized using Receiver Operating Characteristic (ROC) curves [130]. A ROC curve indicates the true positive rate (*sensitivity*) as a function of the false positive rate (1−*specificity*). When no useful discrimination is achieved the true positive rate is similar to the false positive rate. As the accuracy increases, the ROC curve moves closer to the upper-left-hand corner, where a higher sensitivity corresponds to a lower false positive rate. A measure commonly derived from a ROC curve is the area under the curve A_z, which is an indication for the overall performance of the observer [20]. For an ideal classifier the A_z value is equal to one, or as a percentage (as used in the remainder of this thesis) equal to 100. It should be noted that ROC analysis is a pixel based assessment.

Region based analysis can be summarized using Free Response Operating Characteristic (FROC) curves [130]. This is similar to ROC analysis, except that the false positive rate on the x-axis is replaced by the number of false positives per image. In this case a definition of a detected region is needed and a typical approach expects a 50% overlap between the annotated and detected regions to indicate a true positive. See also Appendix C.3 for more information about both kinds of evaluations.

Test Set	Fatty	Glandular	Dense	Total
Circumscribed	9	8	3	20
Spiculated	6	7	7	20
Normal	28	28	24	80
Total	43	43	34	120

Table 2.3: Test set of mammograms.

2.5.3 Mass Segmentation Results

All the algorithms have been tested on a set of 120 mammograms extracted from the MIAS database [183] (see Appendix B.2 for more information about this database). Forty of the mammograms contain a mass (see Table 2.3 for a breakdown of the dataset). All abnormal mammograms have been manually segmented by a radiologist and the annotations are used in following sections as ground-truth data.

The first experiment is related to the capability of the algorithms to distinguish mammograms with and without masses, i.e. the capability of the algorithms to detect masses. Therefore, FROC analysis is used in the evaluation.

The second experiment evaluates the accuracy with which the masses have been detected, and here ROC analysis is used to compare the various approaches. For each algorithm and each mammogram, we compute the value of A_z. Thus, mean and standard deviation values for A_z are given for each algorithm.

2.5.3.1 Capability to detect masses

This aspect of the evaluation mimics the radiologist in detecting abnormalities (in this case abnormal masses). The FROC curves of the eight algorithms based on the 120 mammograms are shown in Figure 2.6. In general, all the implemented approaches have a tendency to over-segment and hence produce a large number of false positives at high sensitivity rates.

Algorithms $d2$, $b2$ and $d1$ show the best trade-off between sensitivity and false positives per image. This improved performance might be due to one aspect that these three approaches have in common, which is the incorporation of directional distribution information. This aspect is expected to reduce the number of false positives which are likely to have a more random distribution than true positive regions. This indicates that spatial information is essential to reduce false positive regions

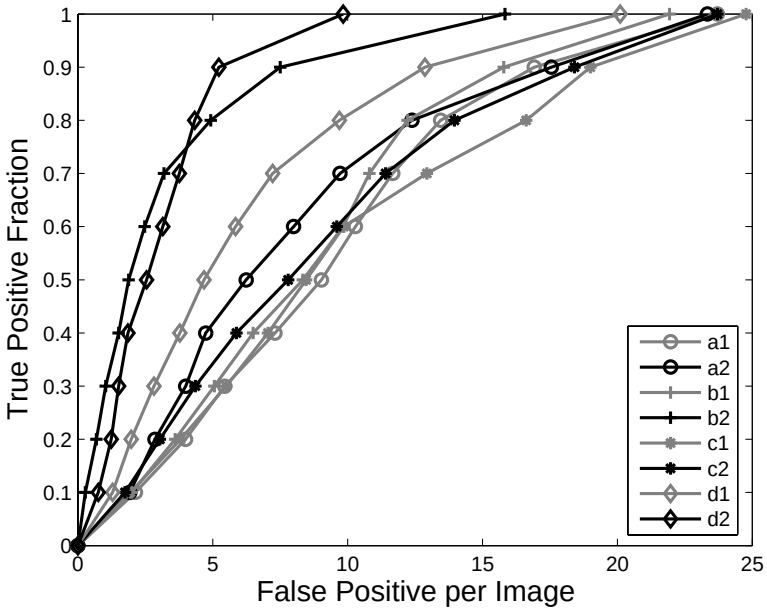


Figure 2.6: FROC analysis of the algorithms over the set of 120 mammograms.

to more acceptable levels.

Computational Cost

We briefly evaluate and compare the computational cost of the algorithms. Although the cost is not a relevant feature, it could be taken into account if the aim of the system is to develop an interactive tool, where the segmentation stage should be relatively fast. Table 2.4 shows the comparison of the execution times for the 8 implemented algorithms. As the MIAS database contains mammograms of different sizes, we compute the mean time of segmentation (and its standard deviation) of the algorithms at each size. It should be noted that the large standard deviation in the segmentation time is caused by variation in breast area.

Algorithms $a1$ and $b1$ are extremely fast in comparison with the rest. This is due to the fact that both algorithms are based on the use of filters which affect the whole image, whilst the other algorithms are pixel-based. On the other hand, the necessity of applying different size and shapes of the template in the pattern matching ($d1$) algorithm makes it extremely slow compared to the rest. In addition, the necessity of algorithm $d2$ to search a neighbourhood makes it also relatively slow.

Algorithms	MIAS Image Sizes			
	Small	**Medium**	**Large**	**Extra**
a1	11 ± 1	14 ± 3	18 ± 5	34 ± 6
a2	31 ± 1	40 ± 3	51 ± 4	96 ± 6
b1	2 ± 1	3 ± 1	4 ± 1	8 ± 3
b2	17 ± 1	22 ± 2	28 ± 4	49 ± 5
c1	20 ± 5	34 ± 9	56 ± 12	97 ± 24
c2	68 ± 27	95 ± 29	146 ± 40	294 ± 70
d1	1766 ± 143	2211 ± 155	2767 ± 183	4794 ± 259
d2	1096 ± 123	1756 ± 160	2504 ± 199	4056 ± 244

Table 2.4: Segmentation time in seconds of the algorithms for the four mammogram sizes (*small*, *medium*, *large*, *extra*) present in MIAS database. Note that each mammogram is downsampled by a factor of 4×4.

2.5.3.2 Accuracy of the detection

Once the mammograms that contain masses have been detected, the algorithms have to be capable of precisely identifying the position and borders of them. This capability is here evaluated using ROC analysis, with the emphasis on performance with respect to the different morphological aspects detailed in MIAS annotations: the lesion shape (circular mass or spiculated one), lesion size, and breast tissue type (glandular, dense or fatty). In addition, an evaluation of the influence of the number of clusters in clustering algorithms is presented.

Lesion Shape Influence

The lesion shape has a strong influence on the performance of the segmentation algorithms. Table 2.5 shows the values of A_z when segmenting the 40 mammograms with masses. In all cases the algorithms show more accurate detection on circumscribed masses. However, due to the small number of samples this difference is not significant.

All algorithms are more accurate in detecting circular masses than spiculated ones. For algorithms $a1$ and $b1$ this is due to the fact that both algorithms apply enhancement filters and thin spicules are likely to be removed. Algorithm $a2$ shows a larger difference which might be caused by the fact that the original algorithm was applied to Regions of Interest (RoIs), while our implementation is applied to the whole image. On the other hand, note that the efficiency of algorithms $c1$ and

Algorithms	Lesion Shape: Circular	Lesion Shape: Spiculated
a1	86.6 ± 6.9	81.5 ± 11.7
a2	86.3 ± 17.3	76.3 ± 23.7
b1	83.5 ± 14.8	81.3 ± 11.5
b2	90.7 ± 5.8	85.6 ± 14.2
c1	85.2 ± 6.9	84.0 ± 13.1
c2	85.8 ± 6.9	84.2 ± 13.2
d1	84.6 ± 10.8	83.6 ± 7.9
d2	90.9 ± 6.3	86.3 ± 12.8

Table 2.5: Influence of the lesion shape for the segmentation algorithms. The results show mean and standard deviation A_z values.

$c2$ is hardly affected by the lesion shape and it can be attributed to the nature of clustering, which only takes pixel feature similarity into account independently of neighbouring pixels. The pattern matching approach ($d1$) loses some accuracy if the lesion is spiculated, because the proposed templates are circular. Finally, algorithms $b2$ and $d2$ were originally developed to detect spiculated masses, but in our implementation show higher accuracy for circumscribed masses. These last two algorithms show the overall best performance on both circumscribed and spiculated masses.

Lesion Size Influence

The influence of the lesion size on the algorithms accuracy is summarized in Table 2.6. It shows that the $d1$ approach works well for small masses, but when the size of the masses increases, the algorithms performance decreases. This is due to the fact that as the size of the masses increases, the variation in their shape also increases. The opposite is true for algorithms $c1$ and $c2$, where performance increases with the size of the lesion. Both k-Means ($c1$) and FCM ($c2$) tend to produce homogeneous clusters [83]. We can also see that the use of filters and statistical approaches (algorithms $a1$, $b1$ and $d2$) follow the same behaviour as the clustering-based methods.

Finally, note that algorithm $b2$ increases its performance with the size of the lesion until a maximum is reached. Then, its performance decreases. This is due to the use of skeletons. When the mass is small, its skeleton is less important for the

Algorithms	Lesion Size (cm^2)					
	< 1.0	$1.0 - 2.0$	$2.0 - 3.0$	$3.0 - 4.0$	$4.0 - 7.5$	> 7.5
a1	80.6 ± 5.7	79.5 ± 10.6	83.1 ± 5.3	84.8 ± 4.6	88.1 ± 6.3	90.1 ± 5.1
a2	89.7 ± 6.3	81.3 ± 12.4	84.7 ± 12.6	78.3 ± 15.5	83.3 ± 8.7	84.0 ± 7.3
b1	68.9 ± 10.4	79.6 ± 10.0	83.3 ± 5.4	85.3 ± 4.7	88.4 ± 6.3	90.3 ± 5.3
b2	84.9 ± 10.7	84.9 ± 9.1	88.1 ± 5.7	90.2 ± 8.3	90.0 ± 7.4	89.0 ± 2.9
c1	80.7 ± 9.2	82.3 ± 9.4	86.8 ± 3.0	87.2 ± 5.8	88.3 ± 7.7	91.5 ± 4.9
c2	81.5 ± 10.3	83.2 ± 9.0	86.9 ± 3.6	87.4 ± 6.3	88.7 ± 8.4	92.0 ± 4.6
d1	93.1 ± 3.4	91.7 ± 9.2	83.0 ± 9.1	82.3 ± 6.6	83.0 ± 5.3	78.3 ± 8.5
d2	84.7 ± 9.1	85.3 ± 8.0	89.9 ± 4.2	88.5 ± 8.1	90.5 ± 9.3	92.6 ± 4.8

Table 2.6: Influence of the lesion size (in cm^2) for the segmentation algorithms. The results show mean and the standard deviation A_z values.

detection process. If the mass size increases, its skeleton becomes easier to detect. For large masses the skeleton becomes more difficult to detect and hence of less use in the detection process. It should be noted that this decrease in performance for larger masses is not significant.

Breast Tissue Influence

The accuracy of the 8 algorithms depending on breast tissue classification is summarized in Table 2.7. Note that most of the algorithms have superior performance in fatty breasts. This is clearly true for the $d1$ and *a2* algorithms, which reduce their accuracy by 10% for dense breasts. The reason for this can be found in the fact that in the glandular and dense breasts of the MIAS database, the difference between mass and normal tissue is less clear when compared to fatty breasts. In addition, most algorithms have better performance when dealing with fatty breasts when compared to breasts of increasing density tissue.

The two exceptions to the above rules are algorithms $b2$ and $d2$. This is because these algorithms use contour information as a basis for the detection process and as such have a better performance when increased intensity changes are present.

Number of Clusters Influence

As indicated when describing the clustering approaches, over-segmentation of the image results in improved mass detection. Table 2.8 shows the result of segmenting the image using algorithms $c1$ and $c2$ but with different numbers of initial clusters,

		Breast Tissue		
		Fatty	**Glandular**	**Dense**
Algorithms	***a1***	85.2 ± 19.7	82.8 ± 7.8	84.3 ± 5.0
	a2	87.5 ± 15.0	79.7 ± 22.0	76.2 ± 26.2
	b1	84.0 ± 6.8	82.8 ± 7.5	82.9 ± 12.8
	b2	87.6 ± 14.1	89.8 ± 8.0	88.0 ± 6.4
	c1	85.9 ± 14.4	85.0 ± 8.0	84.5 ± 5.2
	c2	86.2 ± 14.3	85.6 ± 8.2	85.0 ± 6.7
	d1	88.8 ± 8.7	82.6 ± 8.2	78.5 ± 9.8
	d2	89.6 ± 13.7	90.0 ± 6.5	86.4 ± 6.5

Table 2.7: Influence of the breast tissue for the segmentation algorithms. The results show mean and the standard deviation A_z values.

		Number of Clusters				
		5	**10**	**15**	**20**	**25**
Alg.	***c1***	84.3 ± 5.8	87.5 ± 6.9	88.1 ± 5.8	88.4 ± 5.5	88.5 ± 4.6
	c2	82.2 ± 6.8	87.9 ± 4.1	87.8 ± 3.4	88.7 ± 3.5	89.1 ± 3.2

Table 2.8: A_z dependence on the number of clusters; A_z grows with the number of clusters until it stabilizes.

pointing out that the performance of the algorithms increases with the increase in initial set of seeds. However, there is a point where the increase becomes small and the A_z value is stable. However, this point is not the same for all mammograms. In addition, the segmentation time also increases with the number of initial seed points. In general, an initial set of $15 - 20$ seeds gives the best trade-off between processing time and detection performance.

2.6 Discussion

We have presented and reviewed different approaches to the automatic and semi-automatic segmentation of mammographic masses. We have described several algorithms, pointing out their specific features. Specific emphasis has been placed on the different strategies and a classification of these techniques has been proposed. We have seen that few algorithms are contour-based, probably due to the fact that masses often have not a definite one. Moreover, few algorithms take shape information into account, due again to the morphology of the masses, as masses can appear in a great diversity of shapes and sizes.

Further, we have evaluated 8 of the most frequently used strategies. These methods have been fully evaluated using ROC and FROC analysis with a common database. The annotations, which were used as the gold standard, were provided by an expert mammographic radiologist. It should be made clear that none of the investigated approaches provides the best segmentation on all forty mass containing mammograms. When taking the best segmentation per mammogram into account the A_z values for circular and spiculated masses are 96.0 ± 3.6 and 92.8 ± 5.7, respectively. The equivalent A_z values for the fatty, glandular and dense mammograms are 95.9 ± 5.9, 94.7 ± 2.8 and 91.9 ± 5.0, respectively. Both these results are as expected, since it is more difficult to detect spiculated lesions and it is more difficult to detect masses on a dense (and to a lesser extend glandular) mammographic background. The picture is more complicated with respect to the size of the annotated mass, where for the same range as used in Table 2.6 the A_z values are 96.4 ± 4.3, 94.5 ± 7.4, 92.9 ± 2.4, 94.1 ± 3.0, 93.6 ± 6.2, and 95.7 ± 3.6, respectively. For the smallest masses this is achieved by algorithm $d1$, whilst at the large end of the scale it is algorithm $d2$ that provides the best segmentation. For the middle size range the best segmentations are ascribed to a mix of the algorithms with a majority achieved by algorithm $b2$. It should be clear that for all these cases the mean A_z value is high, which clearly shows an improvement on the individual approaches. One logical continuation of this line of reasoning is to use a combination of segmentation results to provide an improved segmentation approach.

As shown in Section 2.5.3 lesion shape, size and tissue type strongly influence the performance of the algorithms. Few algorithms make use of breast tissue information. Moreover, using FROC analysis we have seen that the algorithms tend to over-segment the image, obtaining a large number of false positive regions. In our opinion, the number of false positives can be reduced by incorporating pixel neighbourhood information for tissue classification to assist in the segmentation process.

Most of the model-based algorithms require the use of a classifier which implies training the system. Currently this is becoming less of a problem as more manually segmented mammograms can be found in the public domain. It should be noted that the quality of annotations is variable and the use of mammographic images from various sources makes normalization essential.

In summary, the reviewed segmentation techniques are still in need of improve-

ment. Results demonstrated that the pattern matching approach using mutual information is an adequate solution if the mass is small. However, it seems less adequate for larger masses, and this is due to the fact that such masses appear in a great variety of sizes and shapes. On the other hand, the classifier based approach seems to be an all-round mass segmentation approach which copes equally well with the small and the large masses. This assessment is based on both region (FROC) and pixel (ROC) based classification.

None of the studied techniques provides perfect segmented results, probably showing that more than a single mammogram have to be used to detect masses. Thus, integration of segmentation results from ipsilateral, bilateral and temporal mammograms is expected to bring improvements in the final result. Work in this field of research has generated interest in the last few years.

Chapter 3

Breast Density Classification

As shown in the previous chapter, the internal density of the breast is a parameter that clearly affects the performance of all the studied segmentation algorithms. Surprisingly, most of these segmentation algorithms do not take this information into account, although density information can be extracted before using such algorithms. In this chapter we review different methods for computing such parameter, and we also present and exhaustively evaluate a new strategy.

3.1 Introduction

Mammographic Computer Aided Diagnosis (CAD) systems are being developed to assist radiologists in the evaluation of mammographic images [14, 54]. However, recent studies [80], as well as the results presented in the previous chapter, have shown that the sensitivity of these systems is significantly decreased as the density of the breast increases, while the specificity of the systems remains relatively constant. In addition, it is well-known that there is a strong positive correlation between breast parenchymal density in mammograms and the risk to develop breast cancer [211]. As Taylor [189] suggested, the development of automatic methods for classification of breast tissue are justified, at least, by two factors:

- To permit better use of the time and skills of expert radiologists by allowing the difficult mammograms to be examined by the most experienced readers.

- To increase the scope for computer-aided detection of abnormalities by concentrating on the easier (fatty) mammograms.

In this thesis, we concentrated on the second factor, not only to detect breast cancer in "easy" mammograms, but also to establish an optimal strategy to look for mammographic abnormalities, as will be shown in Chapter 6.

The origins of breast density classification are the work of Wolfe [211], who showed the relationship between mammographic parenchymal patterns and the risk of developing breast cancer, classifying the parenchymal patterns in four categories. Since the discovery of this relationship, automated parenchymal pattern classification has been investigated, as is explained in the next section. One of the main variations in those publications is that they classified the breast density using various numbers of density categories/scales [133]. However, the American College of Radiology (ACR) Breast Imaging Reporting And Data System (BIRADS) [2] is becoming a standard on the assessment of mammographic images, not only in the US, but world-wide. In this standard, breasts are classified in four categories according to their density (see Figure 3.1 for mammogram examples):

- BIRADS I: the breast is almost entirely fatty.
- BIRADS II: there is some fibrogandular tissue.
- BIRADS III: the breast is heterogeneously dense.
- BIRADS IV: the breast is extremely dense.

In this chapter we review different approaches to automatically classify the breast according to their internal tissue and, moreover, we present a new approach for classifying them according to BIRADS categories. The proposed approach assumes that mammograms belonging to different BIRADS categories are represented by tissue with different texture features. One of the novel aspects of the proposal is the use of an initial fatty versus dense tissue segmentation in order to group pixels with similar tissue characteristics. Subsequently, extracting and comparing texture features from each cluster, the system learns how to differentiate mammograms belonging to each class.

The remainder of this chapter is structured as follows: Section 3.2 shows a survey of the methods found in literature, explaining the main strategies and showing the key points of each method. Section 3.3 describes the proposed segmentation and classification method. Experimental results indicating the validity of the developed approach are presented in Section 3.4, where a quantitative comparison among the reviewed strategies is also done. Finally, discussion and conclusions are given in Section 3.5.

3.2 A Survey on Automatic Breast Density Classification

Two main trends have been followed in breast density classification: mammographic density quantification and mammographic tissue classification. Mammographic density quantification is related to obtaining a single value in order to quantify the breast dense tissue. This value can be obtained using the common 2D views [74, 166, 173, 178, 182, 189] or otherwise using 3D information, in which case, it is called a volumetric measure [16, 79, 200]. However, in medical practice, such quantitative analysis seems unnecessary. In fact, radiologists mainly estimate the breast density by visual judgment of the mammographic imaged tissue. According to this judgment, and using a determined classification, the breast is assigned to their corresponding class. Thus, automatic tissue classification methods try to imitate such visual judgment, learning from the radiologists experience.

This qualitative description of the breast density introduces large intraobserver and interobserver variations in the estimated classification, thus obtaining in general different qualitative descriptions. Although there exist a number of different lexicon/scales for breast tissue classification [133], nowadays, the commonly used is the BIRADS lexicon [2].

In the literature, different approaches based on the use of only histogram information have been proposed for classifying breast tissue [88, 216]. However, in our experience and using public databases, it is clear that histogram information alone is not sufficient to classify mammograms according to BIRADS categories [138, 222]. To illustrate this, the third row of Figure 3.1 shows the respective histograms of

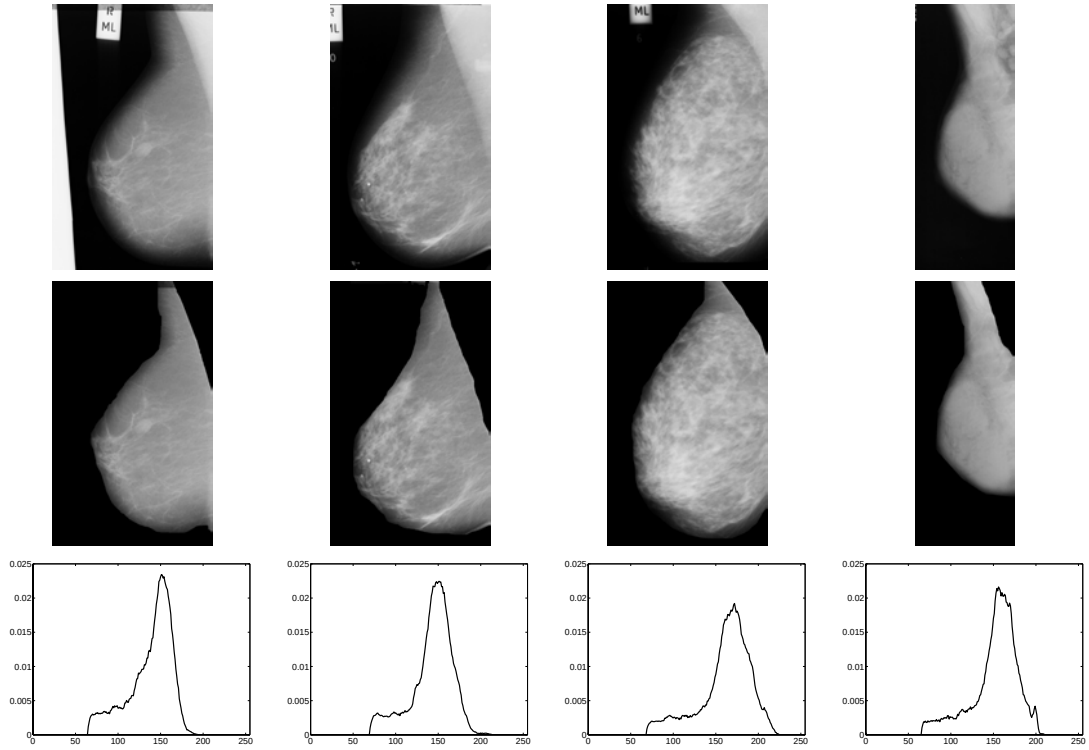


Figure 3.1: From top to bottom the four columns show the original mammogram, the segmented breast area, and the associated histogram, respectively. From left to right this shows four similar histograms, each of increasing BIRADS category: from BIRADS I (first column) to BIRADS IV (last column).

four different mammograms, each belonging to a different BIRADS class. Note that although the mammograms belong to different classes, the four histograms are quite similar both in the mean grey-level value and the shape of the histogram.

Thus, several researchers have focused their attention on the use of texture features to describe breast density. Miller and Astley [131] investigated texture-based discrimination between fatty and dense breast types applying granulometric techniques and Laws texture masks. Byng et al. [23] used measures based on fractal dimension. Bovis and Singh [18] estimated features from the construction of spatial grey level dependency matrices. Recently, Petroudi et al. [148] used textons to capture the mammographic appearance within the breast area. Zwiggelaar et al. [219, 220] segmented mammograms into density regions based on a set of co-occurrence matrices and the subsequent density classification used the relative area

	Author	Year	Segmenting Features
Quantification	Taylor [189]	1994	Fractal, grey-level
	Suckling [182]	1995	Grey-level
	Heine [74]	2000	Grey-level
	Sivaramakrishna [178]	2001	Grey-level
	Saha [166]	2001	Grey-level
	Blot [16]	2005	Volumetric
	Highnam [79]	2006	Volumetric
	Selvan [173]	2006	Grey-level
	Van Engeland [200]	2006	Volumetric

	Author	Year	Extracted Features	Classifier	Lexicon
Global Extraction	Magnin [115]	1986	Co-occurrence	N/A	Wolfe (4)
	Caldwell [24]	1990	Fractal Analysis	Bayesian	Wolfe (4)
	Tahoces [187]	1995	Grey-level, Fourier	LDA	Wolfe (4)
	Boyd [19]	1995	Histogram	Bayesian	SCC (6)
	Byng [23]	1996	Fractal Analysis	Bayesian	SCC (6)
	Zhou [216]	2001	Histogram	Rule-based	BIRADS (4)
	Bovis [18]	2002	Co-occurrence	kNN	BIRADS (4)
	Petroudi [148]	2003	Textons	kNN	BIRADS (4)
Local Extraction	Miller [131]	1992	Granulometric	Bayesian	Wolfe (4)
	Byng [23]	1996	Histogram	Bayesian	SCC (6)
	Karssemeijer [88]	1998	Histogram	kNN	Wolfe (4)
	Blot [15]	2001	Co-occurrence	kNN	MIAS (3)
	Zwiggelaar [220, 219]	2003	Co-occurrence	kNN	SCC (6)
	Gong [63]	2006	Textons	Rule-based	Wolfe (4)
	Martin [123]	2006	Histogram	Rule-based	BIRADS (4)

Table 3.1: Table summary of the reviewed work on breast density classification. The upper block shows works which quantify the density of the breast. The works of bottom block classify the breasts according to the lexicon shown.

of the density regions as the feature space.

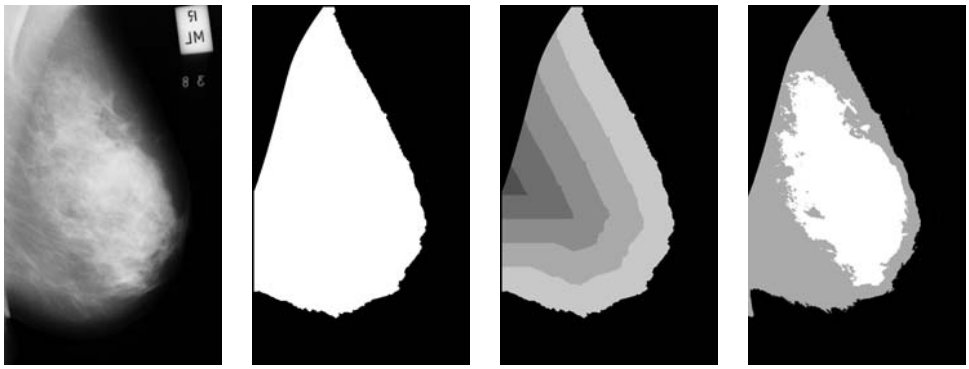
Table 3.1 shows different proposals for breast classification. In the table, the works are classified according to their objective: breast density quantification or classification. Moreover, some other characteristics of the works, as the year, features, the type of classifier, and the number of categories used are shown. Note that among all the previous analyzed works, only the ones developed by Bovis and Singh [18] and Petroudi et al. [148] classified breasts according to BIRADS categories. Moreover, the classification algorithms are further separated into approaches that extract the features treating the global breast as a single region, and approaches that extract features segmenting the breast according to some parameters, for example, the

distance to the skin-line [88].

3.3 A New Proposal for Automatic Breast Density Classification

The first step in our approach is the segmentation of the breast profile, which is done using the algorithm explained in Appendix A.2. As explained, this segmentation provides a minor loss of skin-line pixels in the breast area. In this case, these pixels are also deemed not to be relevant for tissue estimation and, in addition, the relative number of potentially affected pixels is small. The second row in Figure 3.1 shows examples of the breast segmentation.

Therefore, once the breast is segmented, our approach will go beyond the use of histogram information obtaining a set of features for characterizing the mammogram. From the reviewed literature, we can distinguish two related strategies. Bovis and Singh [18] extracted a set of features using the global breast area, hence assuming that the breast is composed of a single texture. As shown in the results of Figure 3.2(a,b) (and in the mammograms shown in Figure 3.1), in many cases this is hard to justify. On the other hand, Karssemeijer [88], and subsequently Blot and Zwiggelaar [15], divided the breast into different regions according to the distance between pixels and the skin-line, as is shown in Figure 3.2(c). The main idea for such approach is the assumption that a strong correlation between tissue density and distance to the skin line will exist. However, note from Figure 3.2 (and again in the mammograms shown in Figure 3.1) that using this strategy it seems that tissue with the same appearance (texture) is divided over different regions, as well as tissues with different appearance are merged in the same region. In contrast with these approaches, our proposal is based on the segmentation of the breast in order to group those pixels with similar tissue appearance, as is shown in Figure 3.2(d). Subsequently, extracting a set of features from each region, different classifiers are trained and tested. A quantitative evaluation of these strategies is provided in Section 3.4.3.



(a) (b) (c) (d)

Figure 3.2: Three different strategies for dividing a mammogram (a) into regions: (b) whole breast area, (c) based on the distance between pixels and the skin-line [15, 88], and (d) based on clustering pixels with similar appearance.

3.3.1 Finding Regions with Similar Tissue

From observing mammographic images one can conclude that pixels from a similar tissue have similar grey-level values, as can be seen in Figure 3.1. Hence, as our aim is to cluster those pixels into meaningful regions, the Fuzzy C-Means algorithm (see Section 2.5.1.6) is used to group them into two separate categories: fatty tissue and dense tissue. Beforehand, and with the aim to avoid effects from microtexture that could appear in some regions, the breast region is smoothed by using a median filter of size 5×5. From our experiments, this filter size is a good compromise between noise reduction and texture preservation of mammographic tissue.

When using partitional clustering algorithms, like Fuzzy C-Means, the placement of the initial seed points is one of the central issues in the variation of segmentation results [83]. Despite their importance, usually seeds for these algorithms are randomly initialized. As we only consider two classes in our approach, the Fuzzy C-Means is initialized using histogram information, with the aim to obtain representative instances of both classes. Hence, we initialized the two seeds with the grey-level values that represent 15% and 85% of the accumulative histogram of the breast pixels of each mammogram (representing fatty and dense tissue, respectively). Although these values were empirically determined, the obtained segmentations do not critically depend on them. Moreover, some mammograms do not have clearly

determined dense and fatty components. In these cases, the segmentation result is one cluster grouping the breast tissue and the other cluster grouping regions with less compressed tissue (an elongated region, like a ribbon, following the skin-line). In these cases, the breast texture information is in the breast tissue cluster, while the ribbon does not provide significant information to the system.

3.3.2 Extracted Features

The result of the Fuzzy C-Means algorithm is the division of the breast into (only) two clusters. Subsequently, a set of features for both classes can be directly extracted from the original images (no preprocessing/filtering was applied). Here we used a set of morphological and texture features.

As morphological features, the relative area and the four first histogram moments for both clusters were calculated. Note that the four moments of the histogram are related to the mean intensity, the standard deviation, the skewness and the kurtosis of each cluster. On the other hand, a set of features derived from co-occurrence matrices [68] were used as texture features. Here we use four different directions: 0°, 45°, 90°, and 135°, and three distances equal to 1, 5, and 9 pixels. Note that these values were empirically determined and are related to the scale of textural features found in mammographic images. Co-occurrence matrices are not generally used as features, rather a large number of textural features derived from matrices have been proposed [68]. For each co-occurrence matrix the following statistics were used: contrast, energy, entropy, correlation, sum average, sum entropy, difference average, difference entropy, and homogeneity features.

As each of these features were extracted from each class, we deal with 226 features in total, 10 from morphological characteristics and 216 from textural information.

3.3.3 Classification

The classification of mammograms according to BIRADS categories was performed in three different ways: by using the k-Nearest Neighbours algorithm, a Decision Tree classifier, and a Bayes classifier based on the combination of the first two algorithms.

3.3.3.1 k-Nearest Neighbours Classification

The k-Nearest Neighbours classifier [45] (kNN) consists of the assignment of an unclassified vector using the closest k vectors found in the training set. In this classifier, the Euclidean distance is used. Due to the fact that kNN is based on distances between sample points in the feature space, features need to be normalized to avoid that some features are weighted more strongly than others. Hence, all features have been normalized to unit variance and zero mean. Moreover, kNN presents another inherent problem, which is the uniform weighting of features regardless their discriminant power. In order to solve this problem we have included a feature selection step which automatically selects the set of the most discriminant features. Here we have used the Sequential Forward Selection (SFS) algorithm [93], which is a widely known technique that selects a local optimum solution in a computationally attractive way. SFS starts by selecting the best single feature and, in an iterative process, subsequent features are selected one at a time which in combination with the already selected ones, maximizes an Euclidean distance based criterion function.

3.3.3.2 Decision Tree Classification

The second classifier used is a decision tree. A decision tree recursively subdivides regions in the feature space into different subspaces, using different thresholds in each dimension to maximize class discrimination. Ideally, for a given subspace the process stops when it only contains patterns of one class. However, in practice, sometimes it is not possible or computationally prohibitive to use such stopping criterion, and the algorithm stops when most of the patterns belong to the same region. In this work we have used the C4.5 decision tree [157], which is an extension of the ID3 decision tree [156] and naturally deals with continuous data.

Both the ID3 and C4.5 algorithms are based on the use of the ID3 information criterion [45] to simultaneously determine the best attribute to split on and at which threshold it has to be split. This criterion is based on the information gain, which measures how well a given attribute separates the training examples according to their target classification at each step. The main difference of both algorithms is that, while ID3 used the information gain itself, C4.5 used the gain ratio.

In our implementation, each leaf corresponds to a test T of the kind: $A_f \leq t$

(feature A is lesser or equal to the given threshold t). This test only has two possible outcomes: true or false, and the issue is to find the best threshold to partition the data. The splitting criterion used is the gain ratio.

The information gained by a test T with $|T|$ outcomes is defined by:

$$Gain(D,T) = I(D) - E(D,T) \tag{3.1}$$

where $I(D)$ is the information associated with the partitions over the set of patterns D (the database of mammograms) and $E(D,T)$ the entropy of the given partition, which is defined below. Mathematically, $I(D)$ is also defined by the entropy:

$$I(D) = -\sum_{c=1}^{|B|} p(D,B_c) log_2(p(D,B_c)) \tag{3.2}$$

where $p(D,B_c)$ is the probability that a pattern/mammogram belongs to the c^{th} class (in this work, $|B| = 4$, as there are four BIRADS classes). On the other hand, $E(D,T)$ refers to the entropy defined by the partition, which is calculated as:

$$E(D,T) = \sum_{i=1}^{|T|} \frac{|D_i|}{|D|} \; I(D_i) \tag{3.3}$$

where D_i are the partitions of the data D defined by the test T (the true and false instances). Thus, the information gain is simply the expected reduction in entropy caused by partitioning the examples according to an attribute. This is the criterion used by the ID3 decision tree.

As the above criterion is maximal when there is one case in each class, a new term is used to penalize this situation. Thus, the split information ($Sp(D,T)$) is defined as the potential information obtained by partitioning a set of cases and knowing the class which a pattern falls in, and is given by:

$$Sp(D,T) = -\sum_{i=1}^{|T|} \frac{|D_i|}{|D|} \; log_2 \left(\frac{|D_i|}{|D|} \right) \tag{3.4}$$

The gain ratio criterion G_{RC} used in the C4.5 decision tree is defined by:

$$G_{RC} = \frac{Gain(D,T)}{Sp(D,T)} \tag{3.5}$$

This ratio is determined for every possible partition, and the split that gives the maximum G_{RC} is selected.

In order to obtain a more robust classifier, the boosting procedure described in [158] is used. The underlying idea of this machine-learning method is to combine simple classifiers to form an ensemble such that the performance of the simple ensemble member is improved. To achieve this, boosting assigns a weight to each instance of the training data, reflecting their importance. Adjusting these weights causes the learner to focus on different instances, leading to different classifiers. Thus, to construct the first decision tree, the weight of all the instances are initialized with the same value. Subsequently, the training data is classified by using this initial tree and an error rate is obtained, by counting the amount of misclassified data. Using this error rate, the weights are recalculated such that those belonging to misclassified instances are increased, and those belonging to correct classified instances are decreased. This process is repeated until either all instances are correctly classified or convergence is achieved. The final classification aggregates the learned classifiers by voting, where each classifier's vote is a function of its accuracy (the error rate).

3.3.3.3 Combined Bayesian Classification

Finally, we constructed a third classifier as a combination of the two classifiers described above with the aim to achieve improved results, because, as we will show in the results, kNN and ID3 classifiers provide complementary information. This third classifier is based on the Bayes rule [45] estimation. When a new case is studied, it is classified according to the classic Bayes equation:

$$P(x \in B_c | A(x)) = \frac{P(A(x)|x \in B_c)P(B_c)}{\sum_{l=1..4} P(A(x)|x \in B_l)P(B_l)} \tag{3.6}$$

Translating this formula into words, we consider the probability of a mammogram x, with set of features $A(x)$, to belong to the class B_c as the posterior probability. The prior is the probability of the mammogram to belong to a class before any observation of the mammogram. If there were the same number of cases for each

class, the prior would be constant (for four categories, as is the case for BIRADS classification and hence $l = 1..4$, the constant value would be equal to 0.25). Here we used as the prior probability the number of cases that exists in the database for each class, divided by the total number of cases. The likelihood estimation is calculated by using a non-parametric estimation, which is explained in the next paragraph. Finally, the evidence includes a normalization factor, needed to ensure that the sum of posteriors probabilities for each class is equal to one.

Combining the kNN and C4.5 classifiers is achieved by a soft-assign approach where binary (or discrete) classification results are transformed into continuous values which depict class membership. For the kNN classifier, the membership value of a class is proportional to the number of neighbours belonging to this class. The membership value for each class B_c will be the sum of the inverse Euclidean distances between the k neighbouring patterns belonging to that class and the unclassified pattern:

$$P_{kNN}(A(x)|x \in B_c) = \sum_{j \in kNN \ \wedge \ j \in B_c} \frac{1}{1 + dist(A(x), A(j))} \tag{3.7}$$

Note that with this definition, a final normalization to one over all the membership values is required. On the other hand, in the traditional C4.5 decision tree, a new pattern is classified by using the vote of the different classifiers weighted by their accuracy. Thus, in order to achieve a membership for each class, instead of considering the voting criteria we take into account the result of each classifier. Adding all the results for the same class and normalizing all the results, the membership for each class is finally obtained.

3.4 Results

Two public and widely known databases were used to test the proposed method: the MIAS database [183] and the DDSM database [71]. As shown in Appendixes B.2 and B.3, whilst the latter has its density classified using BIRADS categories, the former only uses three classes. As we want to classify the breast in BIRADS categories, three mammographic experts (two from the Hospital Dr. Josep Trueta of

κ	Agreement
< 0	Poor
$[0, 0.20]$	Slight
$[0.21, 0.40]$	Fair
$[0.41, 0.60]$	Moderate
$[0.61, 0.80]$	Substantial
$[0.81, 1.00]$	Almost Perfect

Table 3.2: Common interpretation of κ values [106].

Girona and the other one from the Norfolk and Norwich University Hospital) have classified all the MIAS mammograms according to the BIRADS lexicon.

The evaluation of the automatic and manual density classification is presented in the form of confusion matrices [45]. For each confusion matrix we include the kappa (κ) coefficient [36]. This is used by means of estimating agreement in categorical data, and is computed as:

$$\kappa = \frac{P(D) - P(E)}{1 - P(E)} \tag{3.8}$$

where $P(D)$ is the proportion of times the model values were equal to the actual value (the diagonal terms) and $P(E)$ is the expected proportion by chance. A κ coefficient equal to one means a statistically perfect model whereas a value equal to zero means every model value was different from the actual value. Table 3.2 shows a commonly used interpretation of the various κ values [106]. See Appendix C.2.1 for more information about this procedure.

3.4.1 MIAS Database

The method was applied to the whole set of 322 mammograms contained in the MIAS database [183]. Three expert mammographic readers classified all the images in the MIAS database according to the BIRADS categories (the correlation between the original triple MIAS and BIRADS classification is discussed in [133]).

In screening mammography, it is common to obtain expert agreement; here a similar approach is used and consensus between the individual expert classification is used. Table 3.3 shows the confusion matrix for the classification of the three radiologists and the consensus opinion. This consensus is determined by selecting

Consensus \ Expert A (78%, $\kappa = 0.70$)	B-I	B-II	B-III	B-IV
B-I	85	2	0	0
B-II	43	60	0	0
B-III	1	17	70	7
B-IV	0	0	0	37

Expert B (89%, $\kappa = 0.85$)			
B-I	B-II	B-III	B-IV
85	2	0	0
1	93	9	0
0	17	72	6
0	0	0	37

Expert C (72%, $\kappa = 0.61$)			
B-I	B-II	B-III	B-IV
59	28	0	0
0	58	45	0
0	0	88	7
0	0	10	27

Table 3.3: Confusion matrices for three mammographic expert radiologists and their consensus opinion.

as the final class, the class where two or three radiologists agreed (majority vote). If the three experts classified the mammogram in different classes, the median value is selected as the consensus opinion. The results in Table 3.3 show divergence in the opinion of the radiologists, directly illustrating the difficulty of the problem we are dealing with. This disagreement indicates the need to remove inter-observer (inter-operator) variability by the development of automatic methods.

Using the κ values the agreement of Expert A and C with the consensus opinion fall in the *Substantial* category, whilst the agreement of Expert B and the consensus opinion belongs to the *Almost Perfect* category (i.e. the classification by Expert B is almost equal to the consensus). Compared to the consensus, Expert C shows a slight bias towards the higher BIRADS classes than the other two experts, while Expert A shows a slight bias towards the lower BIRADS classes.

Instead of using the majority vote to provide the consensus classification, it is possible to use an expectation maximization approach like STAPLE [205]. In this case, STAPLE produced a consensus that was very close to the majority vote results, with only two mammograms being classed differently. This has minimal effects on the results: the maximum difference on the overall classification results being $\pm 0.3\%$, while for the individual BIRADS classes this increases to $\pm 1.1\%$ (and here positive changes for one BIRADS class are matched by negative changes for one of the other BIRADS classes). For the remainder of the chapter we have used the majority vote results as the consensus classification results.

In order to test the proposed method, we performed two experiments related to the experts classification. Firstly, training the classifiers based on the ground truth as provided by the individual experts, we can evaluate the correlation between the methods and each radiologist. The second experiment was performed training the classifier using as ground truth the consensus between all three experts. In this

case, we would expect an improved agreement as the inter-observer variability is minimized.

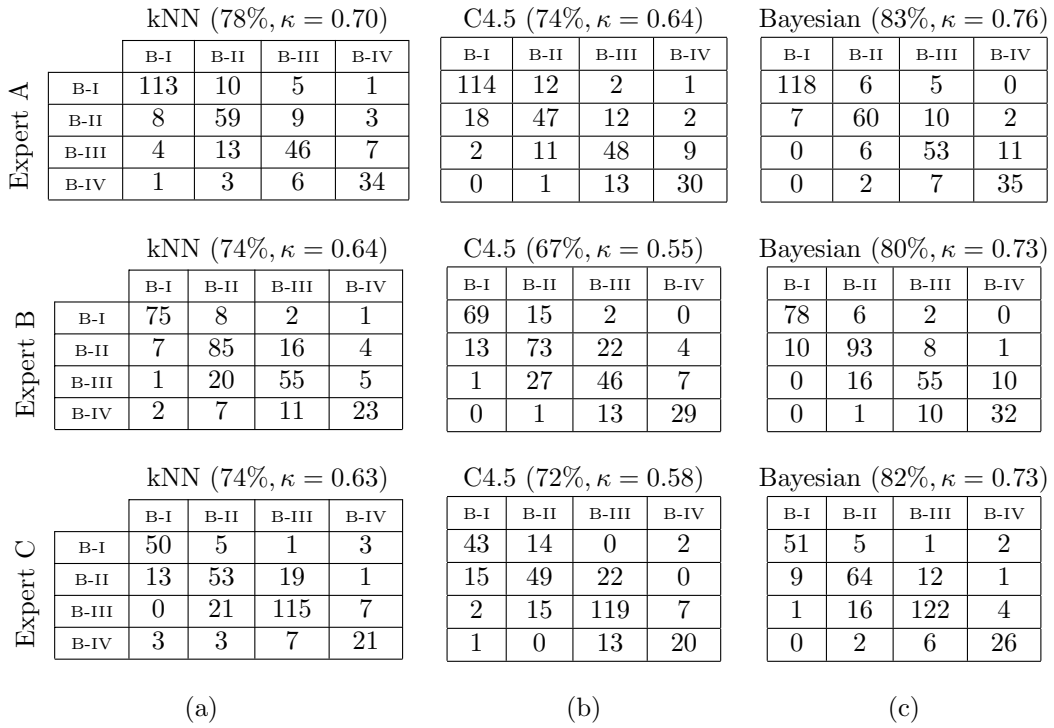
3.4.1.1 Results Based on Individual Manual Classification

Initial experiments consist of the evaluation of the proposed method using the individual expert classifications independently. We used a leave-one-woman-out methodology, i.e. the left and right mammograms of a woman are analyzed by a classifier trained using the mammograms of all other women in the database. The leave-one-woman-out methodology is used to avoid bias as the left and right mammograms of a woman are expected to have similar internal morphology [101]. The confusion matrices for the three classifiers: the SFS+kNN, C4.5, and Bayesian approaches are shown in Table 3.4, where each row corresponds to results based on the manual classification by an individual radiologist. In this work a value of $k = 7$ was used for kNN. Other odd values ranging from 5 to 15 were tested and gave similar results.

For expert A, we can see that the SFS+kNN correctly classifies about 78% of the mammograms, while the C4.5 decision tree achieves 74% of correct classification. kNN clearly outperforms C4.5 when classifying mammograms belonging to BIRADS II, while for the rest of BIRADS the performance is quite similar. On the other hand, C4.5 tends to classify the mammograms according to its own or its neighbouring BIRADS classification, while kNN shows a larger dispersion. The κ coefficient also reflects that kNN has better performances than C4.5, with values equal to 0.70 and 0.64, respectively. Note that both classifiers belong to the *Substantial* category according to the scale in Table 3.2.

The results obtained by the Bayesian classifier are shown in Table 3.4(c). This classifier shows an increase in the overall performance when compared to the individual classifiers, reaching 83% correct classification. This is an increase of 5% and 9% when compared to kNN and C4.5, respectively. When considering the individual BIRADS classes, the percentage of correct classification for BIRADS I is around 91%, whilst in the other cases, the percentages are 76% for BIRADS II, 76% for BIRADS III, and 80% for BIRADS IV. Note that using the Bayesian classifier, κ is increased to 0.76.

The results obtained for expert B are slightly decreased with respect to those



kNN (78%, $\kappa = 0.70$)

Expert A \	B-I	B-II	B-III	B-IV
B-I	113	10	5	1
B-II	8	59	9	3
B-III	4	13	46	7
B-IV	1	3	6	34

C4.5 (74%, $\kappa = 0.64$)

B-I	B-II	B-III	B-IV
114	12	2	1
18	47	12	2
2	11	48	9
0	1	13	30

Bayesian (83%, $\kappa = 0.76$)

B-I	B-II	B-III	B-IV
118	6	5	0
7	60	10	2
0	6	53	11
0	2	7	35

kNN (74%, $\kappa = 0.64$)

Expert B \	B-I	B-II	B-III	B-IV
B-I	75	8	2	1
B-II	7	85	16	4
B-III	1	20	55	5
B-IV	2	7	11	23

C4.5 (67%, $\kappa = 0.55$)

B-I	B-II	B-III	B-IV
69	15	2	0
13	73	22	4
1	27	46	7
0	1	13	29

Bayesian (80%, $\kappa = 0.73$)

B-I	B-II	B-III	B-IV
78	6	2	0
10	93	8	1
0	16	55	10
0	1	10	32

kNN (74%, $\kappa = 0.63$)

Expert C \	B-I	B-II	B-III	B-IV
B-I	50	5	1	3
B-II	13	53	19	1
B-III	0	21	115	7
B-IV	3	3	7	21

C4.5 (72%, $\kappa = 0.58$)

B-I	B-II	B-III	B-IV
43	14	0	2
15	49	22	0
2	15	119	7
1	0	13	20

Bayesian (82%, $\kappa = 0.73$)

B-I	B-II	B-III	B-IV
51	5	1	2
9	64	12	1
1	16	122	4
0	2	6	26

(a) (b) (c)

Table 3.4: Confusion matrices for MIAS classification according to BIRADS categories for individual mammographic experts classification. The results are based on a leave-one-woman-out methodology with 322 mammograms. (a) kNN classifier, (b) C4.5 decision tree, and (c) Bayesian classifier.

obtained for expert A. Specifically, 74% of the mammograms were correctly classified by using the SFS+kNN classifier, while the C4.5 results remained at 67%. The better results for the kNN classifier are independent of the BIRADS classes, except for the BIRADS IV class, in which C4.5 clearly outperforms kNN. The results obtained by the Bayes classifier shows an increase of the performance of 6% and 13% when compared to kNN and C4.5, respectively, obtaining an overall performance of 80%. When considering the individual BIRADS classes, the percentage of correct classification for BIRADS I is around 91%, whilst for the other cases, the percentages are 83% for BIRADS II, 68% for BIRADS III, and 74% for BIRADS IV. The κ value is equal to 0.73.

The last row of Table 3.4 shows the results obtained for Expert C. The performance of the classifiers is similar to that obtained by using the ground truth of Expert B. The kNN classifier obtained 74% correct classification, while C4.5 ob-

tained 72%. Using the Bayes classifier, 82% of the mammograms were correctly classified. In summary, 86% correct classification for BIRADS I, 74% for BIRADS II, 85% for BIRADS III, and 78% for BIRADS IV. The κ value is equal to 0.73.

In conclusion, the best classification rates are obtained using the Bayesian combination. For each individual expert 83%, 80%, and 82% correct classification are obtained, respectively.

In line with other publications [18, 148], we can reduce the four-class classification problem to the following two-class problem: {BIRADS I and II} vs {BIRADS III and IV}, or in words, low density (low risk) versus high density (high risk) classification, which from a mammographic risk assessment point of view might be more appropriate than the four-class division. Comparing to Expert A, the percentage of correct classification is about 92% for the three classifiers and low breast densities, while for dense breasts the percentage is 82%, 88%, and 93% for the kNN, C4.5 and the Bayesian combination, respectively. In contrast, for Expert B, the correct classification percentage for low density breasts is around 88% for the single classifiers and 94% for the combination, while for high density breasts it is reduced to 76% for each classifier, and 86% for their combination. On the other hand, using Expert C, the correct classification percentage for low density breasts is 83% for the single classifiers and 89% for the combination, while for high density breasts the kNN obtains 85%, and the other classifiers 89%.

For this two class approach, in summary, the results are 92%, 91% and 89% of correct classification for Experts A, B and C, respectively.

3.4.1.2 Results Based on Consensus Manual Classification

Table 3.5 shows results based on a leave-one-woman-out methodology for the classification of the whole MIAS database according to the consensus ground truth. The performance of the individual classifiers is 77% correct classification for kNN and 72% for C4.5. These are intermediate values between Expert A and both Expert B and C. However, the Bayesian combination of the classifiers results in improvement and 86% correct classification is achieved, which gives a better performance compared to those obtained by the individual experts without consensus. This result is

kNN (77%, $\kappa = 0.68$)

Consensus	B-I	B-II	B-III	B-IV
B-I	70	13	1	3
B-II	9	80	13	1
B-III	1	17	73	4
B-IV	3	2	8	24

(a)

C4.5 (72%, $\kappa = 0.61$)

B-I	B-II	B-III	B-IV
72	13	1	1
13	68	20	2
0	21	68	6
0	2	11	24

(b)

Bayesian (86%, $\kappa = 0.81$)

B-I	B-II	B-III	B-IV
79	1	3	4
3	86	6	8
0	2	85	8
0	6	4	27

(c)

Table 3.5: Confusion matrices for MIAS classification according to BIRADS categories using the consensus classification. The results are based on a leave-one-woman-out methodology with 322 mammograms. (a) kNN classifier, (b) C4.5 decision tree, and (c) Bayesian classifier.

confirmed by $\kappa = 0.81$, which belongs to the *Almost Perfect* category. Examining each class alone, BIRADS I reached 91% correct classification, BIRADS II 84%, BIRADS III 89%, and BIRADS IV 73%.

Using the low/high density division, low density mammograms are 89% correctly classified, while high density ones reach 94%, resulting in an overall two class classification equal to 91%.

3.4.2 DDSM Database

The developed methodology was also evaluated on a set of 831 mammograms taken from the Digital Database of Screening Mammographies (DDSM)[71], with the main objective to demonstrate the robustness of our proposal on a different and larger data set. Similarly to the MIAS database, DDSM provides for each mammogram additional information including the density of the breast. In contrast to MIAS, this information is already determined using the BIRADS categories.

The number of mammograms belonging to each category is: 106(13%), 336(40%), 255(31%), and 134(16%) for BIRADS I to IV, respectively. These proportion are consistent with the numbers reported by ongoing screening programs. As shown in the work of Lehman et al. [109], where a population of 46,340 women was studied, 13.6% were BIRADS I, 50.9% BIRADS II, 30.1% BIRADS III, and 5.5% BIRADS IV. Although these percentages vary with the age of the women, classes II and III tend to be larger than classes I and IV [35, 66, 202].

The DDSM database provides four mammograms (MLO left and right, CC left

kNN (70%, $\kappa = 0.56$)

Truth \	B-I	B-II	B-III	B-IV
B-I	54	40	12	0
B-II	44	266	25	1
B-III	9	60	177	9
B-IV	0	21	30	83

(a)

C4.5 (72%, $\kappa = 0.59$)

B-I	B-II	B-III	B-IV
51	30	25	0
22	279	35	0
16	59	178	2
8	14	25	87

(b)

Bayesian (77%, $\kappa = 0.67$)

B-I	B-II	B-III	B-IV
58	25	23	0
15	295	26	0
12	46	196	1
5	18	18	93

(c)

Table 3.6: Confusion matrices for DDSM classification according to BIRADS categories. The results are based on a leave-one-image-out methodology with 831 mammograms. (a) kNN classifier, (b) C4.5 decision tree, and (c) Bayesian classifier.

and right) for most women. To avoid bias we selected only the right MLO mammogram for each woman. This way, the leave-one-woman-out used for evaluating the system in the previous sections is now reduced to the typical leave-one-image-out evaluation methodology.

Using this evaluation strategy, Table 3.6 shows the results obtained with the classifiers. These results show a slightly reduced performance when compared to the MIAS database based results (see Tables 3.4 and 3.5). To be specific, the performance obtained by the classifiers is 70%, 72%, and 77% for kNN, C4.5, and Bayesian combination, respectively. Note that using this database, the performance using C4.5 is better than using kNN. This can be due to the use of more mammograms and a different distribution over the BIRADS classes in the training set. The κ value, equal to 0.67, indicates a *Substantial* correlation between the manual and the automatic Bayesian classification.

Examining each class alone, BIRADS I reached 55% correct classification, BIRADS II 88%, BIRADS III 77%, and BIRADS IV 69%. In contrast to the MIAS database, here BIRADS I shows the worst results, whilst BIRADS II shows the best. We believe that this result is due to the fact that in the DDSM database, mammograms belonging to BIRADS I have tissue very similar with those belonging to BIRADS II. Related to the classification of dense mammograms, the ones belonging to BIRADS III are better classified that the ones belonging to BIRADS IV. Moreover, only one mammogram not belonging to BIRADS IV is misclassified as this class.

Using the low/high density division, low density mammograms are 89% correctly

classified, while high density ones reach a 79%. It should be clear that compared to the MIAS consensus results, the performance is mainly reduced on the high density mammograms that has decreased, whilst a similar classification for the low density mammograms is obtained.

3.4.3 The Importance of the Segmentation Step

We include in this section a comparison between our strategy for breast density classification and the others found in the literature. In fact, the main difference among these approaches is the density segmentation, which can be divided in three general approaches: no density segmentation, segmentation according to the distance to the skin-line, and segmentation according to the internal tissue.

To quantitatively compute the improvement provided by our strategy, the same features and classifier as proposed are used. Below, the strategies are explained in more detail.

- **No Segmentation**. The first approach is the extraction of features of the global breast, without any kind of segmentation.

- **Segmentation According to the Distance to the Skin-Line**. This approach was first suggested by Karssemeijer [88]. The main idea is the assumption that a strong correlation exists between tissue thickness and distance to the skin line. To compute the distance to the skin line of the breast a distance transform is used. First, a binary object is formed by merging the breast tissue and the pectoral (or the equivalent segmented region). This object is eroded repeatedly using a circular structuring element. The number of erosions done for a particular pixel in order to remove itself is taken as the distance to the skin-line. Figure 3.3(c) shows the result of applying this algorithm in four different mammograms.

- **Segmentation According to the Breast Density**. Three different approaches have been implemented using such strategy. The first one is our evaluated proposal, which is based on the Fuzzy C-Means clustering (Figure 3.3(d)). The second approach is based on the work of Raba et al. [160],

in which the breast was divided using a fractal scheme. Figure 3.3(e) shows some results using this algorithm. Finally, we construct a novel approach based on a statistical analysis of the breast. Thus, patches of 25×25 pixels from a single mammogram are extracted and used as the ground-truth in order to segment the rest of mammograms. Some of the windows represent dense breast tissue while others represent non-dense tissue. Hence, using the fisherfaces approach [8], these windows are used to construct a model from each part of the mammogram, and subsequently, each subwindow of the mammogram is classified as one of those regions. Thus, we finally obtain a segmentation of the breast in two regions, which represents fatty and dense tissue, as is shown in Figure 3.3(f).

To quantitatively measure the improvement of our proposal we used in this experiment the MIAS database [183] with the annotations obtained from the consensus opinion (the set 322 mammograms divided as 87 BIRADS I, 103 BIRADS II, 95 BIRADS III, and 37 BIRADS IV). The same leave-one-woman-out procedure explained is used to evaluate each strategy.

The confusion matrix for the first strategy (no segmentation) is shown in Table 3.7(a). The overall performance of this approach is 67%, and detailed for each class, we obtained 77%, 74%, 55%, and 57%, from BIRADS I to BIRADS IV respectively. Note that mammograms with low density are better classified than mammograms with high density.

Table 3.7(b) shows the results obtained by the second approach, which is the segmentation of the breast in regions according the distance to the skin-line. Note that the performance is highly increased compared with the no-segmentation approach, resulting in 75% correct classification. The highests improvement are found in mammograms belonging to BIRADS I and BIRADS III, obtaining respectively 89% and 69% correct classification.

Finally, Table 3.8 shows the results obtained by using a segmentation of the breast according to the internal breast tissue. Here (a) shows the results obtained by the Fuzzy C-Means approach, (b) based on the Fractal approach, and (c) using the Statistical approach. Note that the overall performance for each algorithm is similar: 86%, 84%, and 85%, respectively, and all of them are clearly better than

No Seg. (67%, $\kappa = 0.54$)

Truth \	B-I	B-II	B-III	B-IV
B-I	67	19	1	0
B-II	11	76	16	0
B-III	7	22	52	14
B-IV	0	1	15	21

(a)

Skin-line (75%, $\kappa = 0.65$)

B-I	B-II	B-III	B-IV
77	9	1	0
11	77	14	1
0	25	66	4
1	5	9	22

(b)

Table 3.7: Confusion matrix for the classification of the mammograms of MIAS database (a) without segmentation of the breast and (b) segmenting according the distance to the skin-line.

FCM (86%, $\kappa = 0.81$)

Truth \	B-I	B-II	B-III	B-IV
B-I	79	1	3	4
B-II	3	86	6	8
B-III	0	2	85	8
B-IV	0	6	4	27

(a)

Fractal (84%, $\kappa = 0.77$)

B-I	B-II	B-III	B-IV
83	2	2	0
8	84	9	2
0	5	79	11
2	8	4	23

(b)

Statistical (85%, $\kappa = 0.79$)

B-I	B-II	B-III	B-IV
79	3	4	1
1	87	10	5
0	8	81	6
1	4	5	27

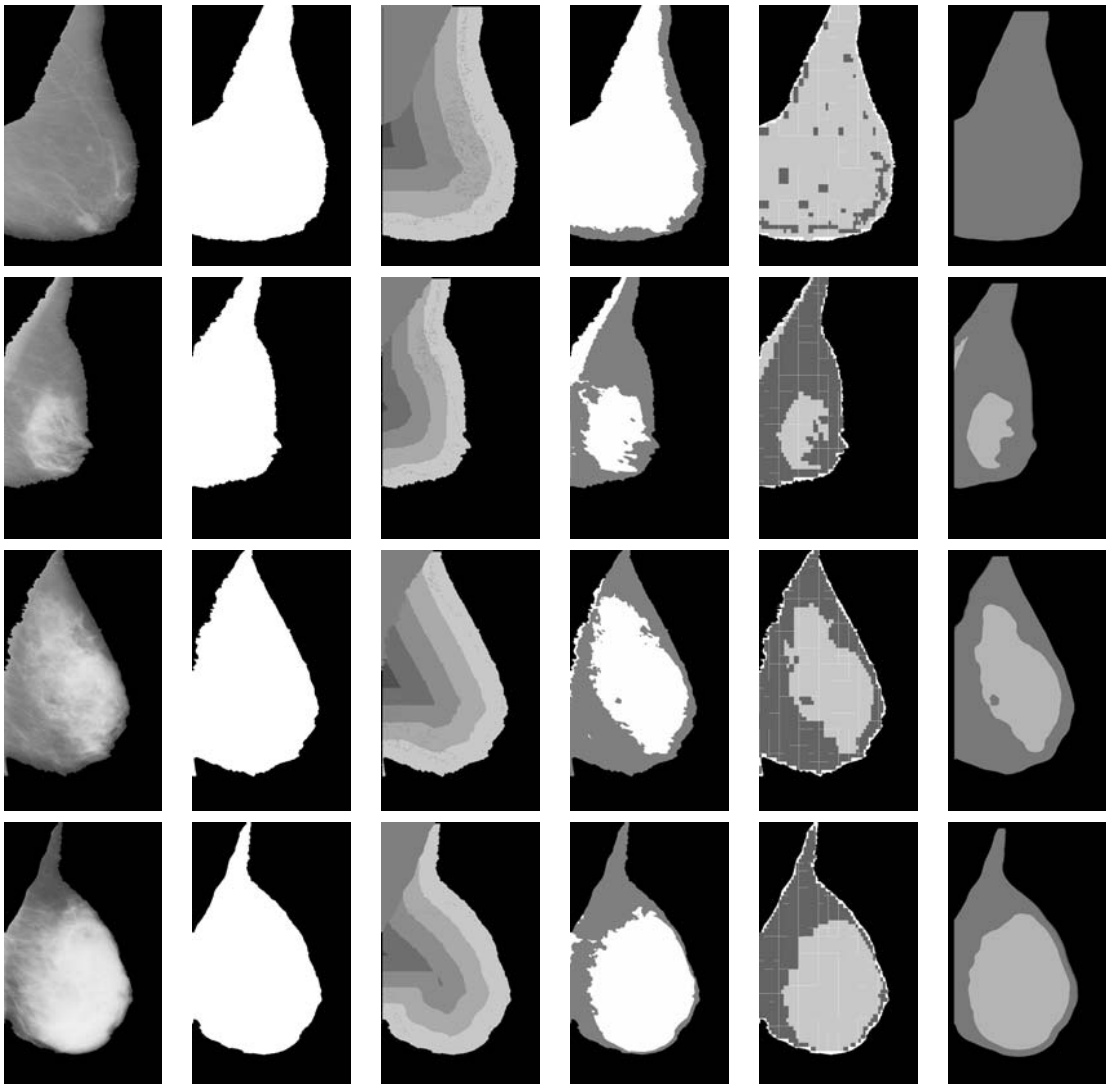
(c)

Table 3.8: Confusion matrices for MIAS mammogram classification by using the internal breast density as a segmentation strategy: (a) Fuzzy C-Means, (b) Fractal, and (c) Statistical approaches.

the results obtained by the other two approaches.

The results obtained by the Fuzzy C-Means and the Statistical approach are quite similar for all classes except BIRADS III. For BIRADS I both approaches obtained 91% correct classification, for BIRADS II the Statistical approach obtained 84% while the Fuzzy C-Means 83%, and for BIRADS IV both approaches obtained 73%. In contrast, for BIRADS III the performance of the Fuzzy C-Means is better, increasing the percentage of correct classification from 85% to 89%. On the other hand, the performance of the Fractal approach is slightly different. It obtains 95% correct classification for mammograms belonging to BIRADS I, while for the rest of classes this is reduced to 82%, 83%, and 62% from BIRADS II to BIRADS IV, respectively.

Figure 3.3 shows the segmentation of the breast according to the compared strategies. Except for BIRADS I, the three last columns (which corresponds to the segmentation algorithms that use breast tissue information) show similar results, and thus the classification results for these strategies are also similar. Note that for



(a) (b) (c) (d) (e) (f)

Figure 3.3: The reviewed strategies for dividing into regions a mammogram. The density of the mammograms shown in column (a) increases from the top row (BI-RADS I) to the bottom row (BIRADS IV). Segmentation using (b) a single breast area, (c) the distance between the pixels and the skin-line, (d) a Fuzzy C-Means clustering of pixels with similar appearance, (e) the fractalization of the image, and (f) the statistical approach.

BIRADS I the Fuzzy C-Means obtains a singular result, grouping in a cluster most of the pixels of the breast except those located near the skin-line, which form the second cluster. This is due to the fact that, for this set of mammograms, the breast is almost homogeneous and the algorithm only can distinguish between those pixels with different compressed tissue (the region is darker in those regions with less compressed tissue). As discussed in Section 3.3.1, the breast texture information is in the breast tissue cluster, while the small ribbon-like cluster does not provide significant information to the system.

Analyzing in more detail the segmentations of the mammograms belonging to the rest of BIRADS categories, one can conclude that the fractal approach provides a pixelated segmentation, while the statistical approach obtains larger and clearly separated regions. On the other hand, the Fuzzy C-Means performance is an intermediate solution and, thus, classification results are slightly improved compared to the other two.

The obtained results show that the segmentation step increase the performance of the classification, improving the results by, at least, 8%. Moreover, we have noticed that using the segmentation according to the breast tissue clearly outperforms the segmentation according to the distance to the skin-line. We have also noted that the strategy used to segment the internal breast tissue does not provide a major variation in the results, with the Fuzzy C-Means based results slightly better than the other ones.

3.5 Discussion

In this section, firstly the proposed approach is compared to other works that classify the breast according to BIRADS categories (see Table 3.1) and subsequently, conclusions are given.

3.5.1 Comparison with the Works which Classifies into BI-RADS Categories

Observing the literature, only the works of Bovis and Singh [18] and Petroudi et al. [148] have classified breast tissue according to BIRADS categories. Bovis and

	B-I(%)	B-II (%)	B-III (%)	B-IV (%)	$Total_4$ (%)	Low (%)	High (%)	$Total_2$ (%)
Bovis [18]	(20)	(21)	(25)	(33)	71			97
Petroudi [148]	91	64	70	78	76	91	94	
MIAS	91 (27)	84 (32)	89 (30)	73 (11)	86	89	94	91
DDSM	55 (13)	88 (40)	77 (31)	69 (16)	77	89	79	84

Table 3.9: Comparison with existing work, with classification according to BIRADS categories. Numbers indicate the overall percentage of correct classification for each class, while the numbers in brackets indicate the percentage of images belonging to each BIRADS class with respect to the whole database. $Total_4$ indicates the overall percent of correct classification for the four classes problem, while $Total_2$ for the two classes problem, detailed in the columns *Low* and *High*.

Singh reached 71% correctly classified mammograms, and Petroudi et al. achieved an overall correct classification of 76%. Table 3.9 summarizes in more detail the results they obtained, including the results of our developed approach. It can be seen that Petroudi et al. obtained similar results to our MIAS database based evaluation, but with significant lesser results on BIRADS II and III and, hence, on the overall classification (column $Total_4$). Moreover, the table shows that Bovis has lower four class results on a smaller DDSM dataset, but higher overall low/high classification (column $Total_2$). Note, however, that a direct comparison is difficult because both have used different datasets. Bovis and Singh used 377 DDSM MLO images (probably different from the ones used in our work), while Petroudi et al. used 132 local (non-publicly available) CC/MLO images. Moreover, it is likely that the distribution over the various BIRADS categories is different in each experiment, and in turn, this could influence the results in the sense that a dataset with a distribution skewed towards BIRADS classes I and IV can be expected to show better results than a dataset with a distribution with a higher proportion of II and III category images. In our experiments a similar behaviour could be seen in the results obtained using the MIAS database and Expert A, who in comparison with Experts B and C used a high percentage of BIRADS I classifications.

3.5.2 Conclusions

In this chapter we firstly reviewed qualitatively the different approaches on breast tissue quantification and classification. We noted that the actual protocol in radiol-

ogy does not tend to quantify the breast, but to qualitatively describe it. For this reason, we focused on breast tissue classification.

From the review of breast tissue classification approaches we noted that none of them used a segmentation between dense and fatty tissue to classify the breast, and we have proposed a new method that uses this segmentation strategy. To briefly summarize it, once the breast has been segmented from the background and pectoral muscle, a Fuzzy C-Means algorithm is used to segment different tissue types (fatty versus dense) in the mammograms. For each tissue region, morphological and texture features are extracted to characterize the breast tissue. Finally, using a Bayesian approach and obtaining the likelihood estimation by combining both kNN and C4.5 classifier results, the mammograms are classified according to BIRADS categories. It should be noted that to avoid bias we have adopted a leave-one-woman-out methodology.

Summarizing the results, we obtained for the MIAS database and individual experts 83%, 80%, and 82% correct classification, which increased to 86% when the classifiers are based on the consensus ground-truth. On the other hand, results based on the DDSM database (a set of 831 mammograms) showed a performance of 77% correct classification. The strength of the Bayesian classifier might be partially explained by the features that were mainly used by the individual classifiers. The SFS stage of the kNN classifier has a strong tendency to select texture features independently of the distance used for the co-occurrence matrices, whilst most of the selected features for the C4.5 classifiers are related to the statistics obtained using a distance equal to 9 for the co-occurrence matrices.

We exhaustively tested the method using MIAS and DDSM database, showing that our proposal outperforms current works on breast tissue classification using the BIRADS standard. Moreover, we compared the different strategies reviewed, showing also that our proposal obtains better results than the others.

Chapter 4

Mass Segmentation Using Shape and Size Lesion Information

Different approaches for mass segmentation have been proposed in recent years, although, as we have shown in Chapter 2, none of them obtain the best performances for all the studied cases. In this chapter, we develop a new algorithm for mammographic mass segmentation which takes shape and size information into account. The algorithm, which should be classified as model-based according to our survey, has been designed in two steps. Firstly, a set of real masses is used to obtain a mass prototype and its possible deformations. Secondly, a probabilistic template matching scheme is used to match the template to the masses present in a mammogram. The performance of the method, which is tested using two different databases and FROC and ROC analysis, demonstrates the validity of our approach.

4.1 Introduction

In Chapter 2 different proposals for mass detection were reviewed. We concluded that the pattern matching approach using mutual information was an adequate solution for finding small masses. This is a crucial issue in radiology, where successful prognosis (or life expectancy) is drastically increased when the cancers are detected in their early stages. However, this approach fails when looking for larger masses, which is likely due to the range of shapes present. On the other hand, the perfor-

mance of the classifier-based approaches do not highly depend on the size of the masses. This is probably due to the fact that these algorithms learn how to detect the masses based on pixel-based features, regardless of the global mass shape. If the training database includes enough representative cases, the algorithm should be able to detect them.

Furthermore, we have seen that few of the reviewed mass segmentation algorithms incorporate prior knowledge about the shape of the masses. Looking into Table 2.1 only some works classified as "Region" and "Model" strategies used shape information. In the "Region" approaches such information was mainly used as a stopping criteria of a region growing algorithm: when the segmentation reaches some particular shape, the algorithm stops the growing step. In contrast, in the "Model" approaches, this information is a fundamental issue. For instance, the works of Lai et al. [105] and Constantinidis et al. [38] were based on a template matching scheme, where region and shape information are equally important.

As also noticed in Chapter 2, the main problem of most of the mass detectors algorithms is the number of false positives, being large. This is particularly true for the template matching algorithm designed in Section 2.5.1.7. Thus, as our approach is likely to suffer from this drawback, we postpone the analysis of possible solutions for false positive reduction to Chapter 5. Moreover, we have seen in the survey of Chapter 2 that the breast tissue influences the algorithms' performance. The introduction of such information into our mass detection proposal will be investigated and incorporated in Chapter 6.

Therefore, our aim in this chapter is to develop a model-based algorithm able to find small and larger masses by means of shape and size analysis of real masses. Briefly, the algorithm follows a template matching scheme, but with two main differences with respect to the rest of the proposed algorithms. Firstly, contour and shape information coming from the analysis of roughly manually annotated masses is used, instead of using region information. Secondly, instead of using a similarity criterion, the algorithm is probabilistic based, following a Bayesian scheme [85]. Let us explain in more detail both differences.

Existing pattern matching approaches [38, 105] construct a rigid and "synthetic" pattern based on the following three facts: the brightness of the mass is higher than its surrounding tissue, the density of the mass is uniform, and the mass has a

circular shape. The result of such assumptions is a template similar to the one shown in Figure 2.5. In contrast, in our proposal, we will firstly find the most probable contours of a mass using real information, obtained from the analysis of the contours of a set of known masses. This step is based on the well-known eigenfaces algorithm [196], initially designed for the face recognition problem. This way, similar to the classifier-based approaches, our algorithm, initially learns the morphology of the masses from real cases. Note that the inherent assumption of such works is, as already commented, that the initial training database has sufficient variability to provide samples for all cases.

Once the template is constructed, it is searched in a mammogram. This is usually done using a similarity measure, such as normalized cross correlation [105] or mutual information (see Section 2.5.1.7). However, these approaches do not allow the template to vary according to the images. In contrast, with our proposal, the constructed template can be adapted to the edges of the image. Hence, instead of using traditional similarity measures, we follow a Bayesian template matching scheme.

The rest of this chapter is structured as follows. In next section, we briefly describe "modern" template matching techniques. Subsequently, to describe the construction of the template, we explain the eigenfaces approach, and why this algorithm is useful for our objective. Afterwards, the design of the template and the pattern matching algorithm are explained. The results, using FROC and ROC analysis and two different databases, are shown in Section 4.6. Finally, the chapter ends with discussion and conclusions.

4.2 A Brief Review on Deformable Template Models

When talking about deformable template models, people usually think of snakes. However, snakes are just a kind of specific template models, as shown in the survey of Jain and Dubes [84], where deformable template models are classified according to Figure 4.1. This section briefly describes the different categories explained in this survey.

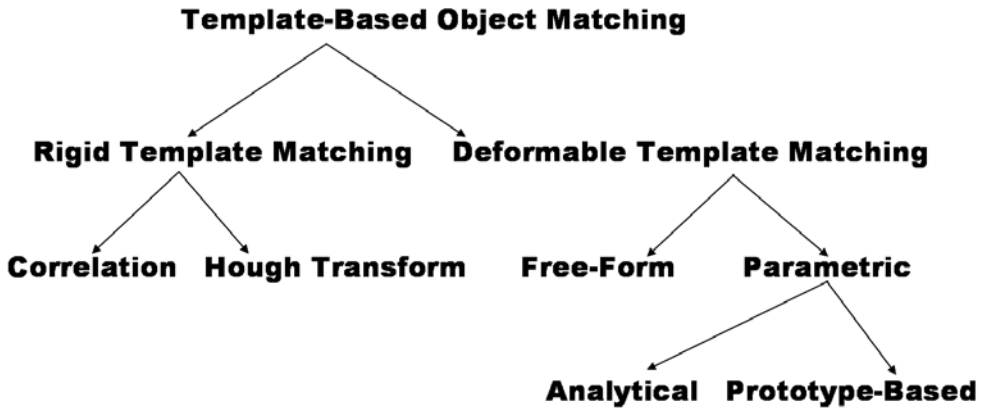


Figure 4.1: An overview of the template matching techniques. Extracted from the survey of Jain et al. [84].

There are two main trends on template models: those which deal with a rigid (fixed) template and the rest where this template varies. It should be noted that the former trend is composed by the early template matching approaches and, although they can be applied on some industrial applications, nowadays are less used. Despite this fact, the reviewed pattern matching approaches for mass segmentation belong to this class of algorithms.

In contrast a deformable model is active in the sense that it is able to adapt itself to fit the given data. Two different sub-trends can be found in this direction: free-form models and parametric models. The former models can represent any arbitrary shape as long as some regularization constraint (continuity, smoothness) is satisfied. The well known active contours approaches (snakes) [91] are classified in this category. Examples of snakes used in mass segmentation reviewed in Chapter 2 are the works of Kobatake et al. [95] and Sahiner et al. [168, 170]. Both approaches used such technique to refine a previous rough segmentation.

On the other hand, there are approaches that provide information related to the shape of the object to the system. Typically, these works firstly try to characterize the shape using a parametric formula, and secondly define a set of deformation modes which let the initial shape vary and adapt to the real images. As usually it is difficult to find a parametric formula to describe the shape, and therefore a prototype template is also used. The active shape models algorithm [39] is the most common

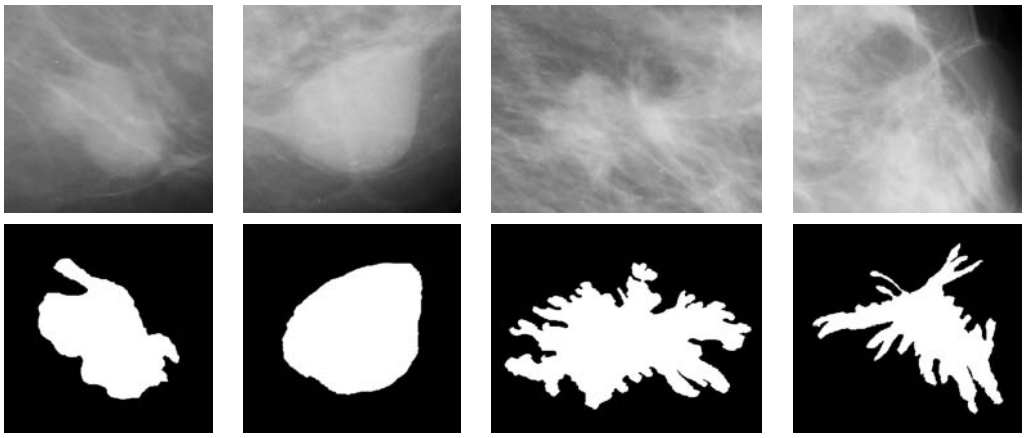


Figure 4.2: Four RoIs corresponding to manually detected masses.

algorithm of this category. In this algorithm, a database of manually segmented images is necessary to construct the mean shape and find their modes of deformation using PCA analysis. In a second step, a new image is segmented using its gradient description and finding which are the main deformed modes. However, an important drawback of such algorithm is the tricky manual segmentation, marking the same number of points at the same position. As is shown in Figure 4.2 it is even difficult to say that two masses have similar positions, due to the large variation found in masses.

Thus, according to Figure 4.1 the algorithm proposed in this work should be classified as a prototype parametric deformable template matching algorithm. By means of the application of the eigenfaces algorithm [196] over a set of real masses, a template and its deformation's modes are found. Subsequently, and using a Bayesian scheme, the prototype is searched in the images. In contrast to Active Shape Models, the initial database of our proposal can be easily obtained from the different public mammographic databases, as only a rough manual segmentation is needed. Concretely, only the centre and the size of the masses are necessary as a starting point (just the bounding box of the mass).

4.3 From Eigenfaces to Eigenmasses

In this section we firstly describe the original eigenfaces approach, and subsequently we will provide the details of how this approach is useful for our objective.

4.3.1 Eigenfaces

The original eigenfaces approach of Turk and Pentland [196] for face recognition is based on the use of the Karhunen-Loeve transform in order to find the vectors that best account for the distribution of face images (forming the face subspace) within the entire image space. The total scatter matrix (the covariance matrix) is calculated as:

$$S_t = \sum_{k=1}^{M} (x_k - \mu)(x_k - \mu)^t \tag{4.1}$$

where $\mu = \frac{1}{M} \sum_{k=1}^{M} (x_k)$ is the mean of all face samples and M the number of face images represented here by vector x_k. Using the Karhunen-Loeve transform it is possible to obtain the subspace which maximizes:

$$W_{pca} = arg \max_{W} |W^t S_t W| \tag{4.2}$$

where W is a unitary column vector. With such approach, the usefulness of the different eigenvectors to characterize the variation among the images is ranked by the value of the corresponding eigenvalue. Hence, it is possible to reduce the dimensionality of the problem in only a few set of eigenvectors, which are the so called eigenfaces. Thus, the eigenfaces span the face subspace of the original image space, and each face image can be transformed into this space by using them. The result of this transformation is a vector of weights describing the contribution of each eigenface in representing the corresponding input image.

Furthermore, a model of each face is constructed by doing the above transformation for each face in the database. Thus, when a new face has to be tested, it will be classified as belonging to the most similar class. In the original algorithm, this similarity is calculated using the Nearest Neighbour algorithm [196].

4.3.2 Eigenmasses and Eigenrois

We can establish a significant parallelism between face images of the same person and mammographic RoIs images. Two of the most common problems in face recognition are related to illumination and pose changes. Note that, in mammography, we can

also talk about variations in illumination and pose. Thus, changes in illumination are related to the acquisition parameters (number and energy of X-rays that go through the breast, the exposure time, the film sensitivity, etc) as well as to the internal density of the breast. On the other hand, changes in the pose can be explained twofold as changes in the global mammogram or in the RoI. Changes of pose in the mammogram are related to the different compression suffered by the breast when the mammogram is acquired. Thus, the shape of a mass, as well as the shape of other internal structures, can be different according to the degree of such compression. Looking at a RoI level, changes in pose can be seen as changes of size and shape of the masses.

Although this parallelism, the transition from face recognition to mass detection is far from trivial, due to the explained previous changes. Namely, the main drawbacks of applying the eigenfaces approach to the detection of masses are the variance of the grey-level range and the variable size of the RoIs. Note that the size of the RoIs depends on the size of the (possible) mass, and there is a huge range of mass sizes [101].

Grey-level and texture variation of RoIs are mainly related to the variation of the acquisition parameters (exposure time, X-ray energy) of mammograms obtained at different time intervals and also to the nature of the breast (breast density and thickness). Using a commonly used simplification, these parameters are considered to affect only to the range of the grey-level values of each RoI. Thus, a solution to take these variations into account can be easily computed by equalizing the images. In this sense, we assume a uniform distribution model. On the other hand, and in contrast with face recognition where a database of faces of the same size is available, the size of the RoIs is not always the same. In order to deal with RoIs of variable size, different proposals can be considered:

1. Cropping the RoI by a size determined by the largest mass in the database.

2. To crop according to the size of the mass, and resize it to obtain equal RoI size.

3. To cluster the database of RoIs in different groups according to their size. Thus, when a new RoI is detected, depending on their size, the corresponding cluster will be used.

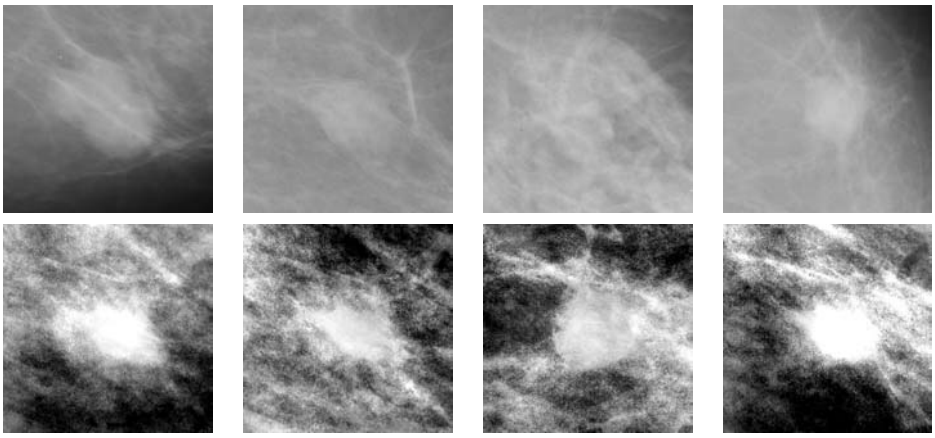


Figure 4.3: The upper row shows four RoIs corresponding to manually detected masses of similar size, while the lower row shows the corresponding obtained eigenmasses.

Experiments have shown that the best results are obtained using the third approach [141]. Note that the main drawback of this approach is the need of a classification of each RoI into the different RoI size clusters. Figure 4.3 shows originals and eigenmasses found using a cluster of RoIs. Note that the eigenmass images show a central mass with brighter grey-level than the surrounding tissue.

To avoid confusion, we shall distinguish between eigenrois and eigenmasses. If all RoIs of the training database contain a mass, the result of the algorithm will be called **eigenmasses**, while if the database is composed by RoIs with masses and RoIs with normal tissue, we will call the resulting images **eigenrois**. We will discuss the eigenrois in Chapter 5.

4.4 Probabilistic Mass Contour Template

The first objective of the proposal is the construction of a general mass template, which should take the shape variations into account. The main aim is that pixels with a boundary morphology which has a major representation in the database have a higher probability than the rest of the pixels. Hence, as the template is represented as an image, the pixel brightness will be associated with the probability of belonging to a contour. Thus, the designed template will have intensity 0 for those pixels which

certainly do not represent a contour, intensity 1 for the pixels which in all images of the database are on a contour (if any), and intermediate values for the rest of the pixels.

An initial solution for the construction of this template consists in considering only the boundaries of manually segmented masses. Note, however, that this solution prefixes a set of contours, and contours different to them will probably be refused while, in contrast, the probability to find two masses with similar shape is very low. Thus, in order to obtain a more general template, it is constructed by looking for the sub-space that these boundaries define. This is achieved by adapting the eigenfaces approach described in Section 4.3. Moreover, using this approach only a rough manual segmentation is needed, just including the centre and size of the mass.

With the obtained eigenmasses, it is possible to construct a probabilistic template per size (note from the previous section that the masses have been clustered according to their size and different templates can be created). For constructing these templates, the N eigenvectors containing 95% of variation explanation were used, considering more probable shapes those with the greatest eigenvalue. Therefore, an initial template is constructed as :

$$\psi^0(x,y) = \frac{1}{N}\sum_{k=1}^{N} w_k W_k(x,y) \tag{4.3}$$

where $\psi^0(x,y)$ is the template, $W_k(x,y)$ is the k-th eigenmass and w_k its normalized eigenvalue (the corresponding eigenvalue divided by their sum). The contour of the eigenmasses is found by extracting the gradient from $\psi^0(x,y)$:

$$\begin{aligned} \nabla\psi^0(x,y) &= \nabla\left\{\frac{1}{N}\sum_{k=1}^{N} w_k W_k(x,y)\right\} = \\ &= \frac{1}{N}\sum_{k=1}^{N} w_k \nabla W_k(x,y) \end{aligned} \tag{4.4}$$

This equation (image) represents the template as a weighted contours of the eigenmasses. In order to obtain a deformable template, it is necessary to specify the

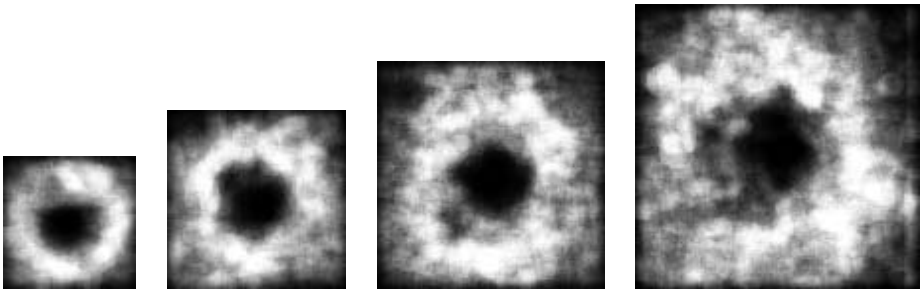


Figure 4.4: The probabilistic templates obtained by clustering the dataset in four clusters of various sizes. Lighter pixels represent a higher probability of a mass contour.

modes of deformation of such initial (rigid) template. Note that the object deformation in an image is an unknown parameter of the model which will be estimated during the template matching step.

Plausible shapes are those obtained from linear combinations of the eigenmass contours, and deformation will only affect the weight of the eigenvalues of each eigenmass. This is represented by a vector ξ of size N:

$$\nabla\psi^d(x,y) = \kappa \sum_{k=1}^{N} \xi_k w_k \nabla W_k(x,y) \tag{4.5}$$

where $\psi^d(x,y)$ is the deformed template and κ is just a normalization factor. With this definition, the vector ξ is all ones when no variations from the template occur, and results in larger difference to the original template as it increases/decreases its values. Hence, assuming a Gaussian distribution, the probability of finding a template with such deformation is:

$$Pr(\xi) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{1}{2\sigma^2} \sum_{k=1}^{N} (\xi_k - 1)^2\} \tag{4.6}$$

Note that with this definition a new parameter (σ) is included. Changes in the value of σ represent a more rigid (small σ) or a more flexible (large σ) template. Figure 4.4 shows the templates for four classes representing the range of mass sizes in the database.

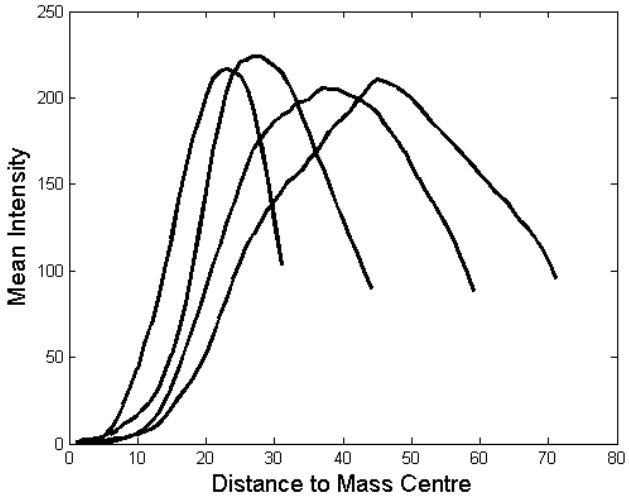


Figure 4.5: Intensity profile of the templates as a function of the distance to the centre (in pixels).

Moreover, Figure 4.5 shows the average intensity over a circle (y-axis) as a function of the radii to the mass centre (x-axis), detailed for each of the four sizes. The peak in each curve represents the radii with highest probability of being a contour of a mass. Note that larger templates are not only a translation of the smaller ones at different radii but also have a different profile, showing the need of having a different training set for each size.

4.5 Template Based Detection

Once a template per size range is constructed, the second step is to match it with the mass boundaries in a mammogram. The developed approach is inspired by the work of Jain et al. [85], where a Bayesian inference scheme was adopted. The three following subsections describe in more detail each of the prior, likelihood and posterior terms. In addition, other details of the design of the algorithm are given.

4.5.1 Prior Distribution

The prior distribution is used to bias the global transformations (changes in translation and scale) and local deformations that can be applied to a prototype template. In contrast to the work of Jain et al. [85], rotation is not taken into account as we assume that this is represented in the probabilistic template.

$\psi^{s,\xi,d}$ denotes a deformation of the original template ψ^0. This deformation is performed by locally deforming the template by a set of parameters ξ, scaling the local deformation by a factor of s, and translating the scaled version along the x and y directions by an amount $d = (dx, dy)$.

Assuming that translations and scale sizes have equal probability[1], and using Eq. 4.6 for the deformation probability, the prior distribution results in:

$$Pr(s, d, \xi) = K \exp\{-\frac{1}{2\sigma^2} \sum_{k=1}^{N} (\xi_k - 1)^2\} \tag{4.7}$$

where K is a normalization factor. Intuitively, a deformed template with a geometric shape similar to the prototype template is favoured, regardless of its size and location in the image.

4.5.2 Likelihood

The likelihood is a measurement of the similarity between the deformed template and the object(s) present in the image. The deformable template will be attracted and aligned to the salient edges in the input image via a directional edge potential field. For a pixel (x, y) in the input image its edge potential can be defined as:

$$\phi_Y(x, y) = -\exp(-\rho\sqrt{\delta_x^2 + \delta_y^2}) \tag{4.8}$$

where δ_x is the displacement to the nearest edge point in the horizontal direction, δ_y in the vertical direction, and ρ is a smoothing factor which controls the degree of smoothness of the potential field. This potential is modified by introducing a

[1]Note that it would be possible to use the database of segmented masses to find a distribution of size and location of the masses. However, it is hard to justify that a mass with a specific shape or location cannot be present in a mammogram

directional component relating the deformed template $\psi^{s,\xi,d}$ to the edges of the input image Y:

$$\Upsilon(\psi^{s,\xi,d}, Y) = \frac{1}{T} \sum_{x,y \in \psi^{s,\xi,d}} (1 + \phi_Y(x,y)|cos(\beta(x,y))|) \tag{4.9}$$

where the summation is over all the pixels on the deformed template, T is the number of pixels on the template, $\beta(x,y)$ is the angle between the tangent of the nearest edge and the tangent direction of the template at position (x,y), and the constant 1 is added so that $\Upsilon(\psi^{s,\xi,d}, Y)$ is positive and takes values between 0 and 1. This definition requires that the template boundary agrees with the image edges not only in position, but also in the tangent direction. Figure 4.6 shows three different mammograms and their respective potential images, where a lighter colour indicates a higher potential. The vertical and horizontal stripes comes from those points far away of either a vertical or a horizontal edge.

Using the above energy function, the probability density of the likelihood of observing the input image, given the deformations of the template is:

$$Pr(Y|s,d,\xi) = \alpha \exp\{-\Upsilon(\psi^{s,\xi,d}, Y)\} \tag{4.10}$$

where α is a normalizing constant to ensure that the above function integrates to 1. The maximum likelihood is achieved when $\Upsilon(\psi^{s,\xi,d}, Y) = 0$ i.e., when the deformed template $\psi^{s,\xi,d}$ exactly matches the edges in the input image Y.

4.5.3 Posterior Probability Density

Using Bayes rule, the posteriori probability density of the deformed template given the input image is:

$$Pr(s,d,\xi|Y) = \frac{Pr(s,d,\xi)\ Pr(Y|s,d,\xi)}{Pr(Y)} \tag{4.11}$$

where $Pr(Y)$ is the normalization factor assuring the sum of all probabilities is equal to 1. Using Eqs 4.7 and 4.10, the posterior results in:

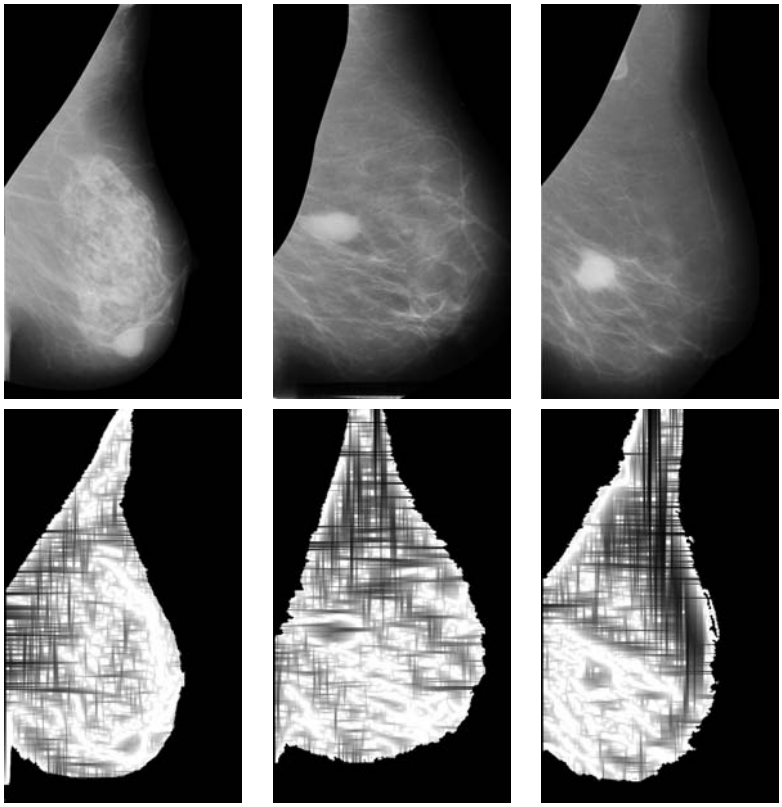


Figure 4.6: Three different mammograms containing clear masses and their potential images. Note that contours of the internal tissue are clearly defined.

$$Pr(s, d, \xi|Y) = K_1 \exp\{-\frac{1}{2\sigma^2}[\sum_{k=1}^{N}(\xi_k - 1)^2 + \Upsilon(\psi^{s,\xi,d}, Y)]\} \tag{4.12}$$

As the objective is to maximize this probability, we seek to minimize the following objective function with respect to s, ξ, d:

$$\Lambda(\psi^{s,\xi,d}, Y) = \sum_{k=1}^{N}(\xi_k - 1)^2 + \Upsilon(\psi^{s,\xi,d}, Y) \tag{4.13}$$

As in the work of Jain et al. [85], this function consists of two terms: a first term that measures the deviation of the deformed template from the prototype, and a second one which describes the fitness of the deformed template to the boundaries of the image.

4.5.4 Final Considerations

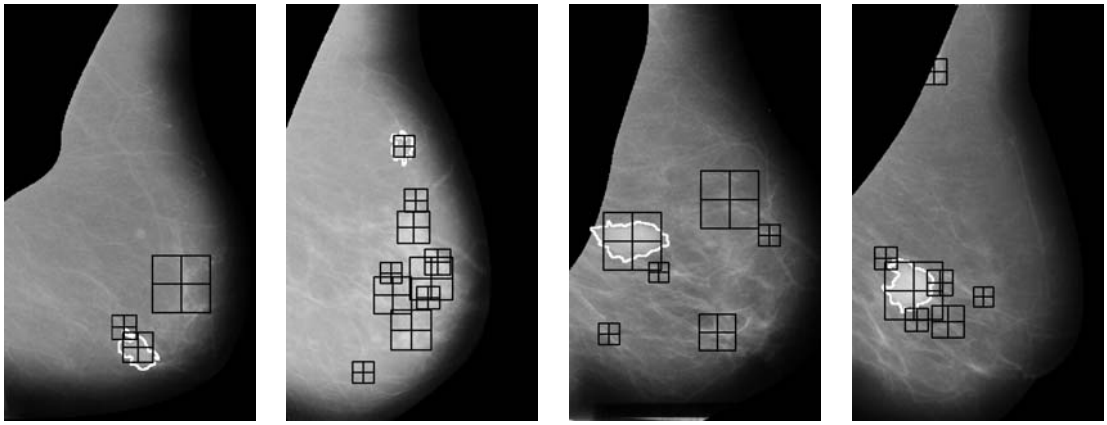
To reduce the complexity of the proposed algorithm, a multiresolution scheme was used. Thus, the initial search is done on a subsampled image and only the potential points found in this resolution (points where $\Lambda(\psi^{s,\xi,d}, Y)$ is greater than a threshold) are used in subsequent steps. These consist of incrementing the resolution of the image and, at these potential points (p_x, p_y), re-calculate the parameters of the template (size and ξ) by using a gradient descent algorithm. With this new template a refinement of these potential points is done. This loop is repeated until the original size of the image is reached or there are no potential points in the image (no mass in the mammogram).

When the algorithm has finished, a set of regions of the mammogram are marked as suspicious RoIs. Each RoI consists of a centre and a surrounding box indicating the size of the possible mass. Figure 4.7 shows the centre of the suspicious RoIs found in different mammograms. We can recover also the final shape of the suspicious RoI. However, as the training step is done using rough annotations, the final shape of the template provides poor information.

The performance of the algorithm is qualitatively shown in Figure 4.7. Each square centre corresponds to the centre of the found mass, while the square size represents the mass bounding square. Note that a large number of the detected RoIs actually correspond to normal tissue. Thus, a subsequent step will be necessary in order to reduce the number of false positives.

4.6 Results

The evaluation is done using a leave-one-out methodology and Receiver Operatic Characteristics (ROC) and Free Receiver Operating Characteristics (FROC) analysis [130] (see Appendix C.3 for details). In the leave-one-out methodology, each query mammogram is analyzed using a model created with the rest of mammograms/RoIs, and this procedure is repeated until all mammograms/RoIs have been used as a query image.



(a) (b) (c) (d)

Figure 4.7: Centre and size of the suspicious regions found in four mammograms of the MIAS database [183] with clear masses (marked in white).

4.6.1 MIAS Database

The performance of our approach is evaluated using a total of 120 mammograms extracted from the MIAS mammographic database [183]. Among them, 40 show confirmed masses (the ground-truth provided by an expert) while the rest were normal mammograms. We used four different groups for the 40 mass RoIs according to their size. Each group corresponds to the following intervals of mass sizes: $<$ 1.20 cm^2, $(1.20 - 1.80)$ cm^2, $(1.80 - 3.60)$ cm^2, > 3.60 cm^2. In each interval there were, respectively, 10, 8, 10 and 9 masses. Three masses were excluded from the modeling: one for being much larger than the rest, and the other two for being located at the border of the mammogram. In order to evaluate our proposal a direct comparison with algorithms *d1* and *d2*, described in Sections 2.5.1.7 and 2.5.1.8 respectively, is given.

As is shown in Figure 4.8, the proposed approach has a better performance compared to both *d1* and *d2* approaches. Note that algorithm *d1* has a tendency to produce a large number of false positives at high sensitivity rates. For instance, at a sensitivity of 0.8, *d1* has 9.69 false positive per image in mean, while *d2* 4.33, and our approach (*Eig*) 2.33.

Figure 4.9 shows the behaviour of the proposed algorithm according to the size of the lesions (note that we have reduced the scale on the x-axis for a better visu-

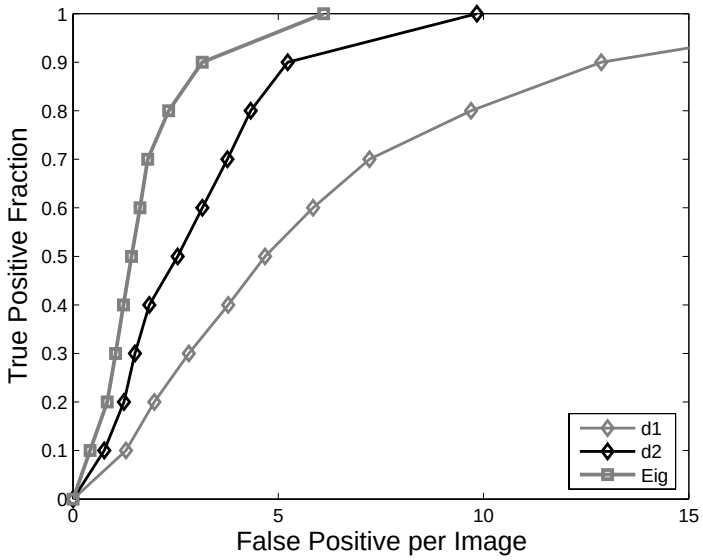


Figure 4.8: FROC analysis of the algorithm over the set of 120 mammograms. The proposed algorithm performs better than both others approaches.

alization). At lower sensitivities the algorithm is more or less independent of the size of the lesion. However, when sensitivity is around 0.7 becomes an important factor, being the performance of the algorithm inversely proportional to the size of the mass. This is due to the fact that the proposed algorithm only detects the centre and the bounding square of the mass: if the mass is larger than the bounding square the sensitivity decreases because there are some pixels not considered as a mass. The opposite can also happen, when the bounding square is larger than the mass the sensitivity decreases because there are some pixels being considered as a mass which are not really part of a mass. Note that both problems are more likely to appear in large masses with spicules than in small masses.

Once the mammograms containing masses are detected, ROC curves are obtained measuring the accuracy with which the masses have been detected. The overall performance of the developed approach over the 40 mammograms containing masses resulted in a $A_z = 89.3 \pm 5.9$, while using *d1*: $A_z = 84.1 \pm 7.9$ and using *d2*: $A_z = 88.1 \pm 8.4$. Thus, the proposed approach obtains a better performance compared to the original algorithms.

Table 4.1 shows the effect of the lesion size for the different algorithms in terms

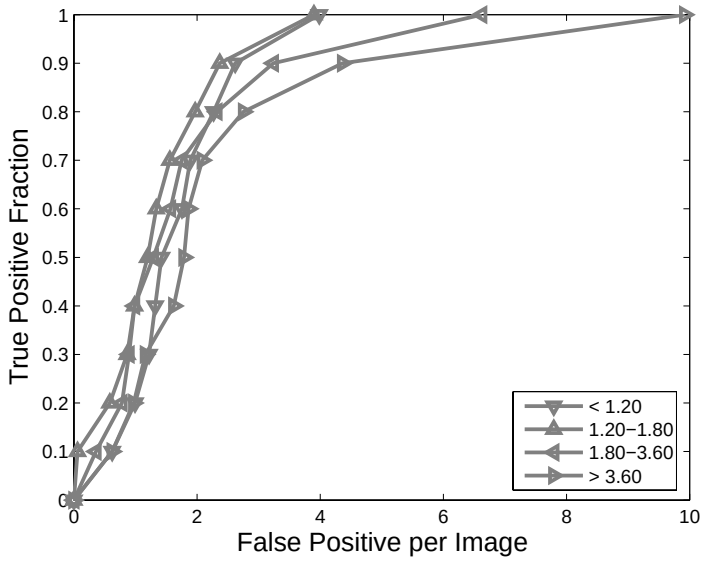


Figure 4.9: FROC analysis of the algorithm detailed for each lesion size.

		Lesion Size (in cm^2)			
		<1.20	**1.20-1.80**	**1.80-3.60**	**>3.60**
Alg.	***d1***	92.1 ± 5.5	85.8 ± 8.2	82.4 ± 7.3	79.1 ± 7.2
	d2	84.9 ± 8.8	86.7 ± 8.1	89.1 ± 9.6	92.1 ± 6.5
	Eig	91.3 ± 7.4	90.3 ± 3.3	89.6 ± 4.7	85.5 ± 5.5

Table 4.1: A_z mean and standard deviation obtained by the different algorithms using the MIAS database, detailed for each mass size (in cm^2).

of A_z mean and standard deviation. Note that the proposed algorithm has a similar performance for the three smallest sizes, while for the largest one the performance decreases. This is due to the shape variability of larger masses. For example, Figure 4.7(c) shows an elliptic mass correctly detected, but with a high number of pixels not being part of a mass classified inside the bounding square. Comparing the performance with algorithms *d1* and *d2*, our approach has a similar trend to *d1*, with a better performance for smaller masses than for larger masses. In contrast, algorithm *d2* tends to increase its performance proportionally to the mass size.

4.6.2 Málaga Database

The performance of our approach has also been evaluated using a set of 35 cases containing masses from the Málaga mammographic database (see Appendix B.4 for more information). Since each case has both CC and MLO views of the breast, we can study the performance of our method using mammograms of each view.

One of the main features of this database is that the ground-truth of the masses is provided by six experts. In the evaluation of our results only the pixels where all radiologists agree are considered as mass, while the rest of pixels are considered normal tissue (we will see in next section that the results are not highly dependent on this parameter). As all the database cases contain masses, we only evaluated our approach using ROC analysis.

4.6.2.1 MLO Views

We clustered the 35 MLO views into four different groups according to their mass size. Each group corresponds to the following intervals for mass sizes: $< 0.70\ cm^2$, $(0.70 - 1.20) cm^2$, $(1.20 - 2.00)\ cm^2$, $(2.00 - 3.80)\ cm^2$. In each interval there were, respectively, 8, 8, 7 and 12 masses. The intervals are different compared to the MIAS database because the masses in this database are smaller.

As already mentioned, the database ground-truth is provided by six experts. Thus, we can compute the performance of the algorithm using those regions where different number of radiologists agree. Considering a true mass those pixels where the six radiologists coincide we obtained $A_z = 90.7 \pm 5.8$. In contrast, if we consider a mass those pixels were at least five radiologists agree, A_z was 90.3 ± 6.2. And decreasing the number of agreement we obtained 89.9 ± 6.7, 89.6 ± 6.8, 89.2 ± 6.9, and 88.6 ± 7.0 for 3, 2, and 1, respectively. This shows an overall trend of performance decrease as the number of radiologist agreement also decreases. This is due to the fact that a different number of thin spicules appears when considering all radiologist annotations as ground-truth. In contrast, only the centre of the mass and clear spicules are taken into account when considering a mass those pixels where all radiologists coincide. In the rest of the evaluation with this database, only this case is analyzed.

The mean A_z for all mammograms for *d1* and *d2* algorithms was respectively

		Lesion Size (in cm^2)			
		<0.70	**0.70-1.20**	**1.20-2.00**	**>2.00**
Alg.	***d1***	97.4 ± 3.5	88.5 ± 10.5	88.4 ± 9.4	84.6 ± 10.4
	d2	91.7 ± 3.9	89.4 ± 5.1	90.5 ± 5.5	90.4 ± 3.3
	Eig	93.4 ± 4.8	91.8 ± 6.6	87.7 ± 5.8	88.0 ± 6.4

Table 4.2: Influence of the lesion size (in cm^2) for algorithms *d1*, *d2*, and the proposed approach, using the MLO images of Málaga database. The results show mean and standard deviation of the A_z values.

90.6 ± 8.8 and 90.6 ± 4.7, while our approach obtained 90.7 ± 5.8. Note that using this database the overall results obtained by the algorithm *d1* are in line with the obtained by the others algorithms. This is due to the fact that the masses in this database are smaller than in the MIAS database, where *d1* performs well for small size masses. Table 4.2 shows the performance of the algorithms depending on the size of the masses. Note that the same trend shown in the previous section for algorithms *d1* and *Eig* are still valid, and they perform better for smaller masses than for larger ones. In contrast, algorithm *d2* shows a similar behaviour for all sizes.

4.6.2.2 CC Views

We evaluated our algorithm using also the set of 35 CC mammograms obtained from Málaga database. In this experiment, we used as a training set a database of RoIs extracted from the CC views. Thus, the same leave-one-out methodology used before is applied in order to not bias the results.

The mean A_z obtained for all mammograms was 90.5 ± 9.5, 89.1 ± 5.9, and 91.9 ± 4.7 for algoritms *d1*, *d2*, and *Eig*, respectively. Moreover, Table 4.3 shows the performance of the algorithms depending on the mass size. Note that this time the proposed algorithm obtains similar performance independently on this factor, while the other algorithms vary its performance depending on the mass size.

Using this database we can also compare the performance of our proposal when dealing with MLO views and CC views. Note that the performance for CC images seems more size independent than in MLO images. Further, in this case the mean is slightly increased. However, we consider that there are not enough images to obtain reliable conclusions.

Alg.	Lesion Size (in cm^2)			
	<0.70	**0.70-1.20**	**1.20-2.00**	**>2.00**
d1	94.2 ± 9.3	93.1 ± 7.8	87.7 ± 6.8	87.8 ± 10.5
d2	87.4 ± 4.3	89.7 ± 6.0	88.2 ± 6.7	90.3 ± 5.7
Eig	91.7 ± 5.3	93.2 ± 5.8	92.9 ± 4.0	91.9 ± 4.7

Table 4.3: Influence of the lesion size (in cm^2) for algorithms *d1*, *d2*, and the proposed approach, using the CC images of Málaga database. The results show mean and standard deviation of the A_z values.

Alg.	Lesion Size (in cm^2)			
	<0.70	**0.70-1.20**	**1.20-2.00**	**>2.00**
Eig (CC)	91.7 ± 5.3	93.2 ± 5.8	92.9 ± 4.0	91.9 ± 4.7
Eig (MLO)	92.1 ± 2.5	93.4 ± 4.8	92.1 ± 2.6	90.4 ± 3.1

Table 4.4: Comparison of the performance when using as a training set RoIs from the same mammogaphic view (CC) or using RoIs from different views (MLO), detailed for lesion size (in cm^2). The results show mean and standard deviation of the A_z values.

4.6.2.3 CC Views with MLO Training

Finally, we also evaluated our algorithm using the set of 35 CC viewed mammograms but using as a training set the RoIs extracted from MLO mammograms. Thus, the leave-one-out methodology is now not necessary as the mass of the same mammogram represented by both views is morphologically different.

The overall performance when using such approach was $A_z = 91.8 \pm 3.4$, while when using RoIs of the same view was $Az = 91.9 \pm 4.7$. Table 4.4 shows the performance of the algorithms depending on the mass size. Note that the results are very similar for the overall mean as well as for each specific size. Thus, we can conclude that the performance of the algorithm is independent on the RoIs view of the training set.

4.7 Discussion

We have presented a new algorithm for mass detection based on the eigenfaces approach, which has been reported to be very useful for face detection and classification problems. The approach learns to detect masses using a database of RoIs

only containing masses, and a probabilistic template is created representing the most probable contours (shapes) of masses. This template forms the basis of an algorithm for looking for masses in a mammogram using a probabilistic scheme. The result of this algorithm is a set of RoIs containing suspicious regions.

The performance of our approach has been evaluated using a leave-one-out methodology and FROC and ROC analysis, and two different databases. In addition a comparison with similar approaches from the state of the art has been given, obtaining slightly improved results. Although, in general, the obtained results are considered promising, the number of false positive obtained at high sensitivity levels is still significant. Moreover, one of the characteristics of the algorithm is that it performs better when dealing with smaller masses than for the larger ones. In fact, this behaviour could be expected as the algorithm is template-based and, as we have shown in Chapter 2, these algorithms demonstrate that behaviour. Moreover we have seen that the algorithm is independent of the RoIs view.

Chapter 5

False Positive Reduction

A new statistical-based approach for the discrimination of normal RoIs and RoIs depicting true masses is presented in this chapter. The method is based on modeling the tissue variation of both kinds of RoIs by extracting the principal components of a set of already classified RoIs. Subsequently, the system projects each new RoI onto a feature space that spans the significant variations among the known RoIs. The performance of the method is tested in two ways. Firstly, using ROC analysis, the method discriminates between both kinds of RoIs using a leave-one-out methodology. Secondly, the false positive reduction is integrated in the algorithm developed in the previous chapter in order to demonstrate its validity.

5.1 Introduction

Almost all works trying to detect masses in mammography need a final step in order to reduce the number of false positives (regions being normal marked as suspicious by the algorithm). This is due to the complexity of the internal breast tissue, which induces the detection of regions which are not masses, but normal variations in tissue characteristics.

A set of different techniques for false positive reduction have been developed in recent years. These algorithms are based on the classification of the RoIs as normal tissue or as depicting an abnormality, which in our case are masses. Thus, all the algorithms are based on a typical classifier scheme: using a database of known cases

Classifier-Based					
Author	**Year**	**Features**	**Classifier**	**RoIs**	**Results**
Sahiner [169]	1996	Texture, Morphologic	LDA, NN	168/504	$A_z = 0.90$
Christoyianni [34]	2002	Grey-level, Texture, ICA	NN	119/119	88.23%
Qian [155]	2001	Texture, Shape	NN	200/600	$A_z = 0.86$
Tourassi [194]	2005	Grey-level	NN	681/984	$A_z = 0.84$

Template-Based					
Author	**Year**	**Features**	**Similarity**	**RoIs**	**Results**
Chang [29]	2001	Grey-level, shape	Likelihood function	300/300	$A_z = 0.83$
Tourassi [195]	2003	Grey-level	Mutual Information	809/656	$A_z = 0.87$

Table 5.1: Summary of the reviewed works on false positive reduction, with the features used, the classifier/similarity used (where LDA means linear discriminant analysis, NN neural network analysis, and ICA independent component analysis), the number of RoIs depicting masses vs the number of normal RoIs, and the results obtained. Note that for all works accuracy is given in terms of A_z (the area under the ROC curve) except for the work of Christoyianni et al. [34] which just gives the correct classification percentage.

the system learns how to differentiate between both kinds of RoIs. Subsequently, once the system has been trained, a new RoI can be classified.

Observing these algorithms, we can distinguish between two main strategies. The first one includes the set of algorithms which firstly extracts features from the RoIs, usually related to their texture, and subsequently trains the classifier. On the other hand, a second strategy handles this problem as a template matching algorithm. Each new image is compared to all the RoIs of the database and then it is classified as an image containing a mass or not. Table 5.1 summarizes some works belonging to both strategies.

Note that among all those works, one of the main differences are the ratio between the number of RoIs depicting masses and the total number of RoIs. This is an important issue because the number of wrong classified RoIs will increase as the number of normal RoIs increases. One should remember that the aim of this step is to reduce the number of false positives, which is usually higher than the number of true positives (as we have seen in Chapter 2).

Sahiner et al. [169] extracted a huge set of features, and subsequently used genetic algorithms to select the most discriminative ones. With this subset of features, a neural net (NN) and a linear classifier (LDA) are trained and used to classify a

new RoI. A similar strategy is used by Christoyianni et al. [34], who extracted grey-level, texture, and features related to independent component analysis (ICA), and use them to train a neural net. Note also, that they apply a principal component analysis (PCA) pre-processing step to reduce the complexity of the problem. On the other hand, Qian et al. [155] analyzed the implementation of an adaptive module to improve the performance of an automatic procedure which consists of training a Kalman-filter based neural net using features obtained from a wavelet decomposition.

As explained, the works of Chang et al. [29] and Tourassi et al. [195] are based on comparing a new RoI with all the RoIs in the database. The two most clear differences between them arise from the similarity measure and the database used. More specifically, the former developed a likelihood measure which depends on the grey-level and the shape of the RoIs. Both parameters were compared with the new RoI and the set of RoIs present in the database, which was only composed by RoIs depicting masses. From this comparison a likelihood measure was computed. On the other hand, the work of Tourassi et al. [195] consists of comparing all the RoIs of the database (including RoIs with and without masses) with the new one using a mutual information based similarity measure. Thus, the new RoI will be labeled as belonging to the closest class.

Note that with the last strategy, the similarity used for classifying has to be re-computed for each new element, as it measures the difference between the new RoI and all the RoIs in the database. On the other hand, a different drawback found in the first strategy is that a large set of features needs to be computed and only some of them will be finally selected. In contrast, we will show that our approach, which is again based on the eigenfaces algorithm, is more straightforward and efficient. Furthermore, we will show that using the recent developed 2DPCA decomposition [212] instead of using the typical PCA, the results are greatly improved.

5.2 PCA-Based False Positive Reduction

When the algorithm used to detect potential masses has finished, a set suspicious regions (centre of masses and lesion size) have been found, some of them being actually masses whilst the rest represent normal tissue. Thus, a posterior processing

step is needed to classify a RoI according to these two classes. For doing such task we adapted, once again, the eigenfaces approach.

The eigenfaces approach assumes that a database of already classified RoIs is available. This contains only two types of RoIs: RoIs containing masses and RoIs of normal tissue[1] (this is equivalent to a face recognition system containing two people). Different instances for each class are included in the database. Their intra class variability is mainly due to grey-level and texture differences and to the shape and size of the mass or other structures present in the RoI. Again, a parallelism with face detection can be established related to variations of pose and illumination, respectively.

According to the Karhunen-Loeve transform (Eq. 4.2) the usefulness of the different eigenvectors to characterize the variation among the images is ranked by the value of the corresponding eigenvalue. Note that in this case we should refer them as eigenrois. Thus, the eigenrois span the RoI subspace of the original image space, and each RoI can be transformed into this space by using them. The result of this transformation is a vector of weights describing the contribution of each eigenroi in representing the corresponding input image.

Therefore, a model for each RoI in the database is constructed by using Eq. 4.2. When a new RoI has to be tested, it will be classified as belonging to the most similar class. Although in the original algorithm the similarity was calculated using the k-Nearest Neighbour algorithm, here we used the already explained Bayesian combination of this algorithm with the C4.5 decision tree (see Section 3.3.3 in Chapter 3). With this algorithm we can obtain a degree of membership for each class and hence a ROC analysis can be done obtaining an A_z value.

5.3 2DPCA-Based False Positive Reduction

The 2DPCA approach [212] is a recent improvement of the typical eigenfaces approach. As the authors argue 2DPCA has important advantages over PCA in two main aspects: firstly, it is simpler and more straightforward to use for image feature

[1]Note that although in this work the database contains only two types of RoIs (masses or normal tissue) this can be extended to include other RoIs containing microcalcifications or architectural distortions.

extraction since 2DPCA is directly based on the image matrix, and secondly, it is easier to accurately evaluate the covariance matrix[2].

In the original eigenfaces approach, each image of size $m \times n$ is transformed into a vector of size $m \cdot n$, in contrast to the natural way to deal with two dimensional data, which would be treating it as a matrix. This is the motivation of 2DPCA [212].

The algorithm starts with a database of M training images. The *image covariance matrix* G_t is calculated by:

$$G_t = \frac{1}{M} \sum_{j=1}^{M} (A_j - A_\mu)^t (A_j - A_\mu) \tag{5.1}$$

where A_μ is the mean image of all training samples. Then, using the Karhunen-Loeve transform it is possible to obtain the corresponding face space, which is the subspace defined as:

$$\begin{cases} \{X_1, ...X_d\} = arg \max |X^t G_t X| \\ X_i^t X_j = 0, \quad i \neq j, \quad i, j = 1, ...d \\ X_i^t X_i = 1, \quad i = 1, ...d \end{cases} \tag{5.2}$$

where X is a unitary column vector. The first equation looks for the set of d unitary vectors where the total scatter of the projecting samples is maximized (the orthonormal eigenvectors of G_t corresponding to the first d largest eigenvalues). On the other hand, the other two equations are needed to ensure orthonormal constraints.

With the selected set of eigenvectors it is possible to construct a family of feature vectors for each image. Thus, for an image sample A, the projected feature vectors (the principal components) $Y_1, ..., Y_d$ are found by:

$$Y_k = AX^k, k = 1...d \tag{5.3}$$

It is important to note that while for PCA each principal component is a scalar,

[2]One could argue that the template in the detection step could be obtained using such algorithm. However, this was not used here for computational reasons. Note that using our approach, the deformations of the template are modeled using only a vector of coefficients. Instead, with the 2DPCA approach, a vector of vectors will be necessary as each principal component will be a vector.

for 2DPCA each principal component is a vector. It is this set of vectors for image that is used to construct the feature image (a matrix of size $m \times d$) referred to as $B = [Y_1, ...Y_d]$.

In a similar way to the eigenfaces approach, comparing images means to compare the constructed features. As the dimension of the feature space has increased in one dimension, now the comparison of images is done by comparing matrices:

$$d(B_i, B_j) = \sum_{k=1}^{d} ||Y_k^i - Y_k^j|| \tag{5.4}$$

where $||Y_k^i - Y_k^j||$ denotes the Euclidean distance between the two principal component (vectors) Y_k^i and Y_k^j. To obtain an A_z value we adopt the analogous probabilistic scheme described in Section 5.2.

5.4 Evaluation of the False Positive Approaches

In this section, the results using the PCA and 2DPCA approaches over a different set of manually selected RoIs are explained. The evaluation is done by using the MIAS [183] and the DDSM [71] databases and ROC analysis [130]. In contrast to the previous chapters where this analysis was performed counting pixel by pixel, we now proceed image by image, i.e. dealing with all the image as a RoI with mass or as a RoI with normal tissue (more details of this kind of evaluation are given in Appendix C.2.2). To avoid confusions, we will express the obtained A_z values using this image per image strategy in the interval $[0 - 1]$, while when using the pixel per pixel strategy the results will be given as a percentage.

In order to perform a more global evaluation of our results we propose to compute the A_z value for different ratios of number of RoIs depicting masses and number of RoIs depicting normal tissue (from ratio 1/1 to ratio 1/6). The idea of analyzing these different ratios is twofold: firstly, to evaluate the performance of our method on different levels of difficulty (a ratio 1/1 will obtain more optimistic results than 1/6), and secondly, to compare our proposal with existing methods (the ones presented in Table 5.1). It is important to notice that previous works only provide results for specific (and usually different) ratios. Hence, analyzing all these ratios will enable

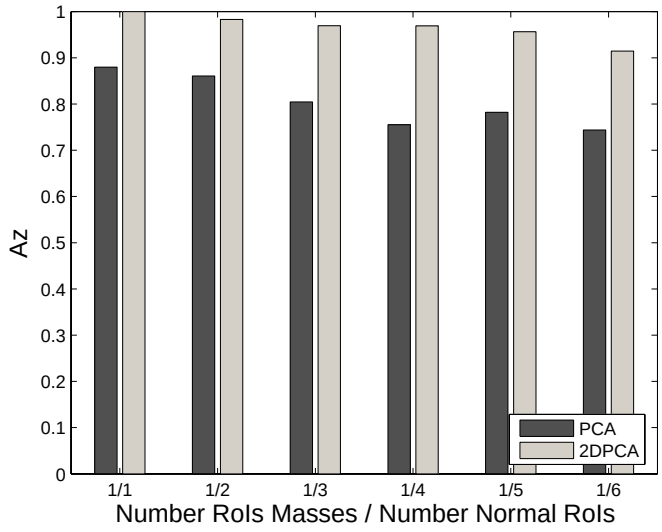


Figure 5.1: Performance of the system using the MIAS database.

the comparison with them.

5.4.1 MIAS Database

For the MIAS database four groups of RoIs were used according to their size. Each group corresponds to the following intervals for mass sizes: $< 1.20\ cm^2, (1.20 - 1.80)\ cm^2, (1.80 - 3.60)\ cm^2, > 3.60\ cm^2$. In each interval there were, respectively, 10, 8, 10 and 9 masses, while the rest of RoIs represent normal tissue.

Figure 5.1 shows the mean A_z obtained using the leave-one-out strategy according to increasing number of RoIs representing normal tissue (from the same number of RoIs to six times this number). Obviously the performance of both systems decreases as the number of normal tissue RoIs increases. For instance, for the PCA approach, the performance is reduced from $A_z = 0.88$ to $A_z = 0.74$, while the 2DPCA approach goes from $A_z = 1.00$ to $A_z = 0.96$. Note that the 2DPCA clearly outperforms the normal PCA.

Table 5.2 shows the A_z values for both approaches for the ratio of one RoI with mass and three normal RoIs in the database (1/3). The overall performance of the system at this relation is 0.80 for PCA and 0.97 for the 2DPCA. In the first row of the table, a detailed comparison of the performance for each group is shown. Clearly,

		Lesion Size (in cm^2)			
		<1.20	**1.20-1.80**	**1.80-3.60**	**>3.60**
Alg.	***PCA***	0.72	0.82	0.83	0.92
	2DPCA	0.92	0.98	1.00	1.00

Table 5.2: Detailed A_z results per size for the classification of masses for both approaches using the MIAS database. The results are detailed for each size group.

both approaches are more suitable for larger masses than for small ones. This is due to the fact that larger masses have a larger variation in grey-level contrast with respect to their surrounding tissue than small masses, which are usually more subtle.

Observing Figure 5.1 at different ratios we can also quantitatively compare our approach with the ones found in the literature. With the same 1/3 ratio that studied above, Sahiner et al. [169] and Qian et al. [155] obtained A_z values of 0.90 and 0.83 respectively. Both values are superior to the obtained by the PCA approach, but clearly inferior to the obtained by the 2DPCA approach. On the other hand, with the ratio of one RoI of each class our approaches outperforms the presented by the rest of surveyed works. Note however, that the total number of RoIs is too small in order to extract significant conclusions. We leave this discussion for Subsection 5.4.2.

Figure 5.2 shows the obtained mean kappa statistic (defined in Equation 3.8) when using the leave-one-out strategy at a determined threshold[3] for all the ratios of number of masses vs number of normal RoIs (from the ratio 1/1 to the ratio 1/6). Obviously the same behaviour found for the A_z values is now repeated. Thus, the performance of both systems when the number of normal tissue RoIs increases is reduced from $\kappa = 0.68$ to $\kappa = 0.37$ for the PCA approach, and from $\kappa = 1.00$ to $\kappa = 0.82$ for the 2DPCA. Note that the differences in performance are clearer analyzing the kappa statistic than the A_z value.

5.4.2 DDSM Database

The algorithm was also evaluated using a database of 1,792 RoIs extracted from the DDSM mammographic database [71]. From this set, 256 depicted a true mass,

[3]Remember that for each RoI we obtain a membership value to each class (mass or normal tissue). We obtain the ROC curve plotting this membership as a function of a threshold. A threshold of 0.5 ensures that all RoIs will be classified as being either a mass or normal tissue. Hence, this is the threshold value used for computing the kappa statistic.

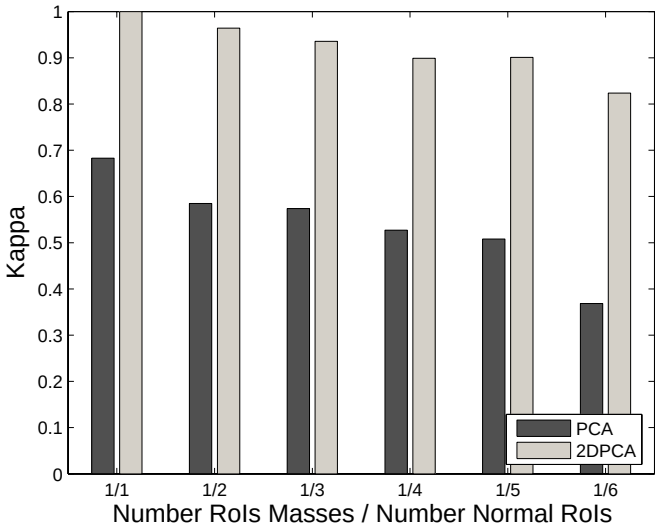


Figure 5.2: Performance of the system using the MIAS database using the Kappa statistic.

while the rest, 1,536, were normal, but suspicious tissue. According to the size of the lesion, we used six different groups of RoIs. Each group of RoIs corresponded to the following mass sizes intervals: $< 0.10\ cm^2$, $(0.10-0.60)\ cm^2$, $(0.60-1.20)\ cm^2$, $(1.20-1.90)\ cm^2$, $(1.90-2.70)\ cm^2$, $> 2.70\ cm^2$, and the number of masses en each interval was respectively, 28, 32, 37, 57, 69, and 33 masses.

Figure 5.3 shows the mean A_z value obtained using the leave-one-out strategy and varying the ratio between both kind of RoIs. Note that, again, the performance of both PCA and 2DPCA approaches decreases as the ratio of RoIs depicting masses decrease. For the PCA approach we obtained $A_z = 0.73$ for the ratio 1/1 and $A_z = 0.60$ for the ratio 1/6, while using the 2DPCA approach we obtained $A_z = 0.92$ and $A_z = 0.81$ respectively. Thus, the 2DPCA approach obtained better performances than the PCA.

The A_z values for the ratio 1/3 are detailed in the first row of Table 5.3. The overall performance of the system at this ratio is 0.65 for PCA and 0.86 for the 2DPCA. As in MIAS results, both approaches are more suitable for false positive reduction of larger masses than smaller ones. As already explained, this is due to the fact that larger masses have a larger variation in grey-level contrast with respect to their surrounding tissue than smaller masses, which are usually more subtle, even

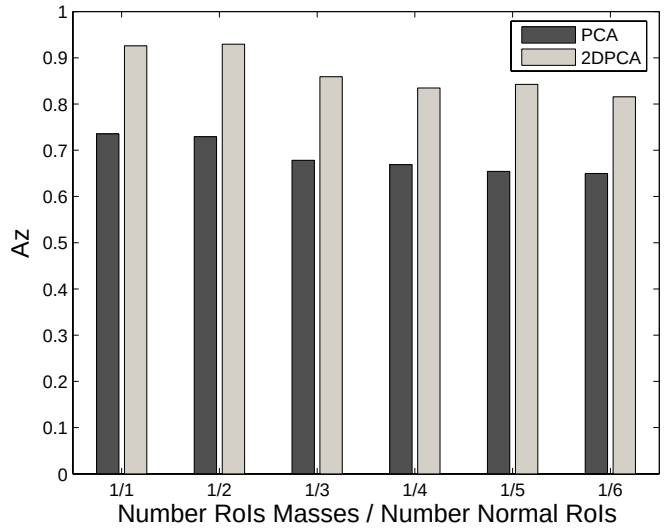


Figure 5.3: Performance of the system for the DDSM database.

		Lesion Size (in cm^2)					
		<0.10	**0.10-0.60**	**0.60-1.20**	**1.20-1.90**	**1.90-2.70**	**>2.70**
Alg.	***PCA***	0.53	0.70	0.70	0.68	0.72	0.83
	2DPCA	0.81	0.83	0.87	0.84	0.89	0.93

Table 5.3: A_z results for the classification of masses using RoIs extracted from the DDSM database, detailed per size (in cm^2).

for an expert.

Comparing the results between both MIAS and DDSM databases, it is obvious that the ones obtained using MIAS were better than the obtained using the DDSM database. This is mainly due to two different reasons: firstly, the fact that we can extract a more larger subset of RoIs using the DDSM than using the MIAS database, and secondly, the masses in MIAS database were larger than in the DDSM, and as we have explained, this is an important increasing factor of the performance of both algorithms.

Finally, Figure 5.4 shows the mean kappa statistic obtained using the leave-one-out strategy at a determined threshold (0.5). The same behaviour found for the A_z values is repeated. Thus, the performance of both approaches are reduced when increasing the number of normal tissue. Comparing with the results obtained using the MIAS database, accuracy is also reduced. For the 2DPCA approach, only when

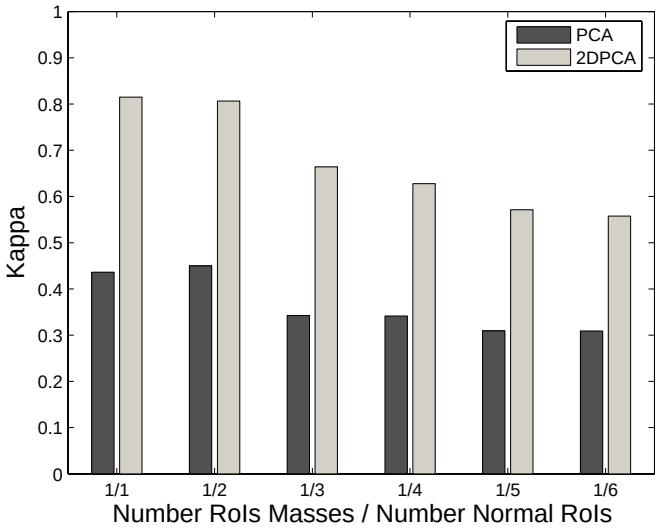


Figure 5.4: Mean Kappa values obtained by the system using the DDSM database.

there is one or two RoIs with normal tissue for each RoI with masses the agreement is *almost perfect*, while for ratio one-to-three the agreement is *substantial* and for the rest of cases it is in the high part of the *moderate* agreement.

Using this large dataset, we can compare the proposed PCA and 2DPCA-based algorithms with the ones surveyed at the beginning of this chapter. With the same ratio 1/3 Sahiner et al. [169] and Qian et al. [155] obtained A_z values of 0.90 and 0.83 respectively. While the performance of the PCA-based approach is inferior to the other algorithms, the 2DPCA-based approach clearly outperforms the results of Qian et al. In contrast, the mean value obtained using this approach is inferior to the one obtained by Sahiner et al.

Comparing with the other approaches where the authors use the ratio 1/1, the PCA approach still has inferior values. However, the 2DPCA approach outperforms the existing approaches.

As an illustration of the information provided by PCA analysis[4]. Figure 5.5 shows the nine images constructed by using the nine first eigenvectors of the third group of RoIs. Note that each image contributes with different information to the system. For instance, the first image (the first eigenvalue) represents the main

[4]Note that a similar result is obtained using 2DPCA analysis.

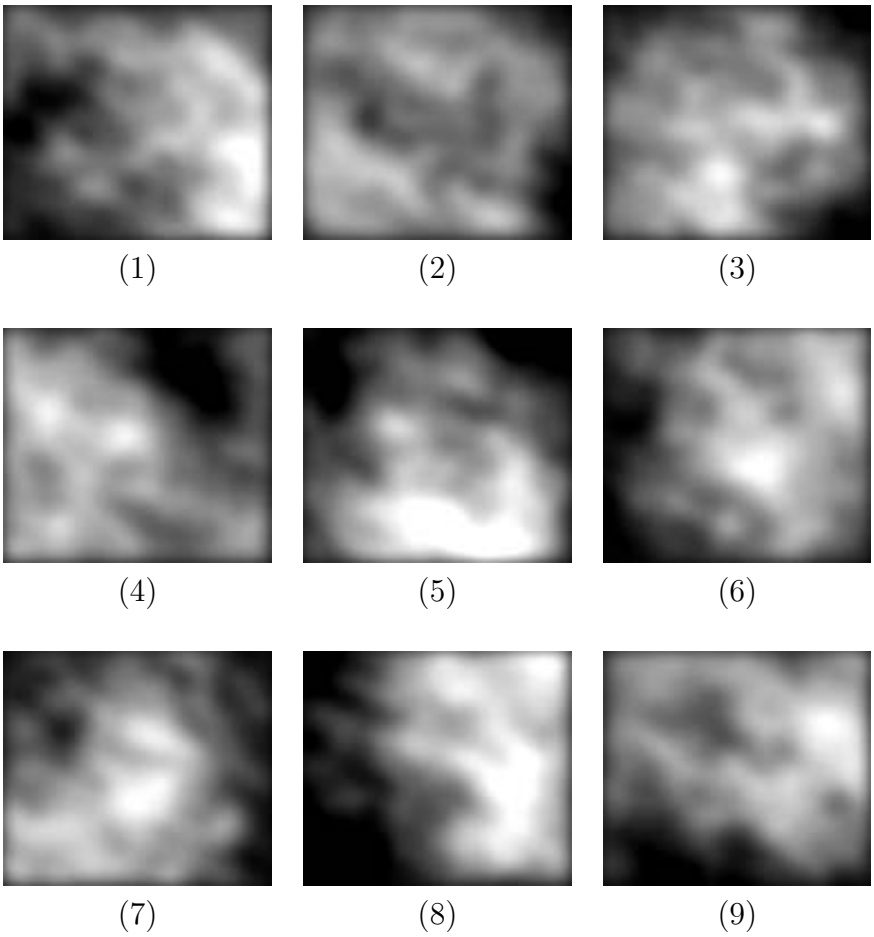


Figure 5.5: The first nine eigenimages found using the third group of RoIs, obtained using PCA analysis.

variation in the grey-level transition going from top-left to down-right. The second one represents the variation of the grey-level values from the outside and the inside of the image. Note also that this second eigenvector is related to the non-presence of masses, as well as eigenvectors 6 and 7 are related to their presence.

5.5 Combining the Bayesian Pattern Matching and the False Positive Reduction Step

In this section we evaluate the results of the combination of both developed algorithms: the Bayesian template matching algorithm developed in Chapter 4 and the false positive reduction approach developed in this chapter. Thus, for training the algorithms two different subsets are necessary: one to construct the template for finding the suspicious regions and another one to discard the suspicious regions actually being normal tissue. It is important to note that the first subset contains only RoIs depicting masses, while the second one contains different kind of RoIs. Moreover, we only use the 2DPCA approach for false positive reduction since we have demonstrated that outperforms PCA.

Similar to Section 4.6 the evaluation is done using a leave-one-out methodology and Receiver Operating Characteristics (ROC) analysis. In the leave-one-out methodology, each mammogram is analyzed using a model created with the rest of RoIs not belonging to this mammogram, and this procedure is repeated until all mammograms have been used as a query image. The evaluation is done here using the MIAS database.

5.5.1 MIAS Database

The performance of the system is evaluated using a total of 120 mammograms, 40 with confirmed masses (the ground-truth provided by an expert) and the rest being normal mammograms.

Both sets of RoIs (the one containing only masses and the other containing masses and normal tissue) were extracted from these mammograms. We used four different groups according to their size. For the first dataset, each group corresponds to the following intervals for mass sizes: $< 1.20\ cm^2, (1.20-1.80)\ cm^2, (1.80-3.60)\ cm^2, > 3.60\ cm^2$, and there were, respectively, 10, 8, 10 and 9 masses. For the second set of RoIs these groups were completed with 3 normal, but suspicious, RoIs images for each mass RoI. In the results, algorithms *d1*, *d2* and the algorithm without this false positive reduction step are also included for direct comparison.

Figure 5.6 shows the FROC curve for our proposal explained in the Chapter 4

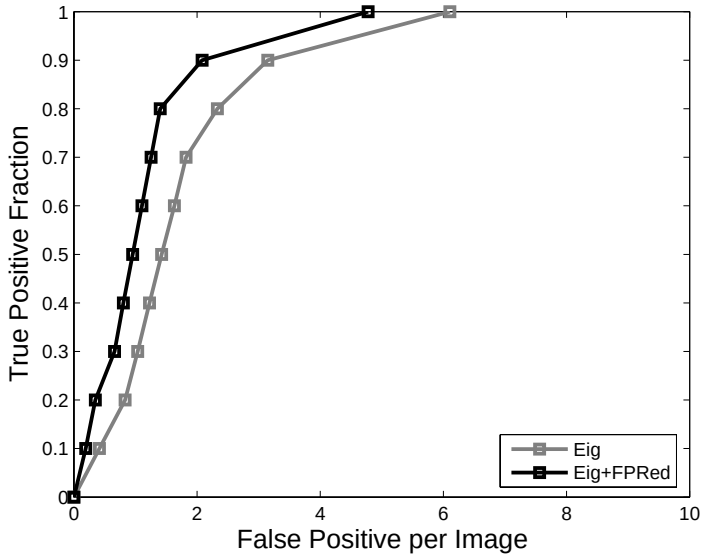


Figure 5.6: FROC analysis of the algorithm including the false positive reduction step (black line) over the set of 120 mammograms compared to the algorithm without false positive reduction (grey line). It is clear that the use of the false positive reduction clearly outperforms the proposed algorithm.

(grey line) and the same approach integrated with the proposed false positive reduction algorithm (black line). Note that the inclusion of this step clearly improves the performance of the algorithm: at the same sensitivity, the number of false positives per image is reduced. For instance, one false positive per image is reduced at a sensitivity of 0.87. Analyzing in the other direction, the inclusion of the false positive reduction algorithm allows to increase the sensitivity at a given false positive rate. For example, at one false positive per image the sensitivity increases from 0.27 to 0.58.

On the other hand, Figure 5.7 shows the FROC curve for the algorithms *d1*, *d2*, and the proposed system including the false positive reduction step (the black line with squares). The difference between the proposed algorithm and both approaches is now clearer than in Figure 4.8. For instance, at the same sensitivity analyzed in Section 4.6.1 (Sensitivity = 0.8) the mean number of false positive per image is now 1.40, which is 0.93 less compared to the algorithm without the false positive reduction step. This shows the benefits of including this algorithm.

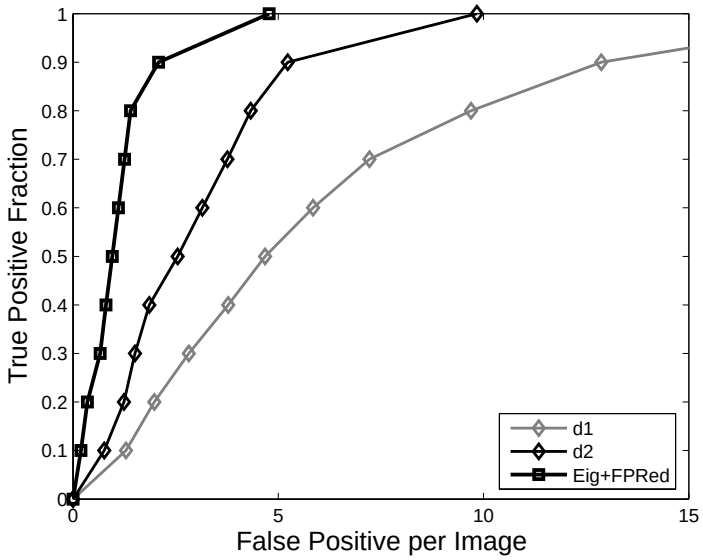


Figure 5.7: FROC analysis of the algorithm over the set of 120 mammograms compared to *d1* and *d2* algorithms. The inclusion of the false positive step improves the proposed probabilistic template matching.

We include again the performance of the algorithm detailed for each lesion size in Figure 5.8. Note that larger masses are still more difficult to be accurately detected. However, the inclusion of the false positive reduction step allows to detect them at almost 3.00 false positive per image less than without this step. Moreover, the performance of the three smaller sizes is now more similar than without using the false positive reduction step.

Once the mammograms containing masses are detected, ROC curves are constructed to measure the accuracy in which the masses are detected. The overall performance over the 40 mammograms containing masses resulted in A_z values of 89.3 ± 5.9 and 89.1 ± 4.1 without and with the false positive reduction step, while the results for the both compared approaches were $A_z = 84.1 \pm 7.9$ and $A_z = 88.1 \pm 8.4$ for algorithms *d1* and *d2* respectively. Note that the false positive reduction step introduces a penalization term in the accuracy with which the algorithm detects masses. This is due to the elimination of some RoI that were actually representing a true mass.

Table 5.4 shows the effect of the lesion size for the different algorithms in terms

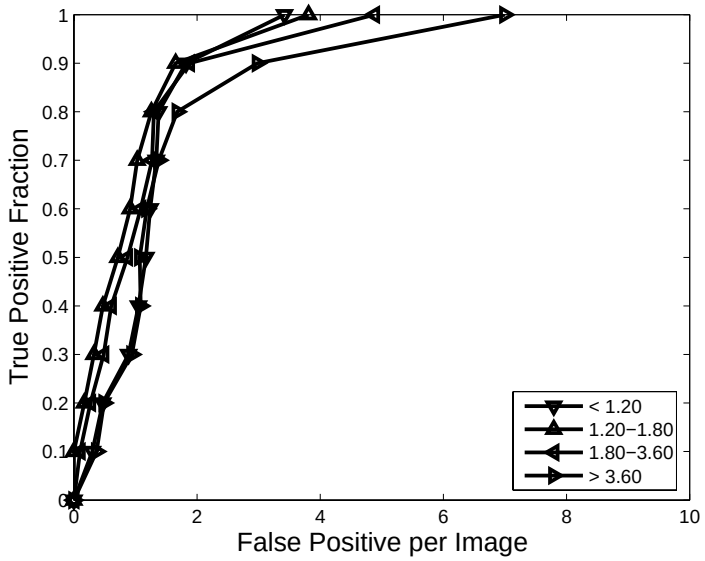


Figure 5.8: FROC analysis of the algorithm detailed for each lesion size.

		Lesion Size (in cm^2)			
		<1.20	**1.20-1.80**	**1.80-3.60**	**>3.60**
Alg.	***d1***	92.1 ± 5.5	85.8 ± 8.2	82.4 ± 7.3	79.1 ± 7.2
	d2	84.9 ± 8.8	86.7 ± 8.1	89.1 ± 9.6	92.1 ± 6.5
	Eig	91.3 ± 7.4	90.3 ± 3.3	89.6 ± 4.7	85.5 ± 5.5
	Eig* & *FPRed	89.9 ± 3.1	91.4 ± 2.1	88.5 ± 5.0	86.6 ± 3.2

Table 5.4: Influence of the lesion size (in cm^2) for the *d1*, *d2*, and the proposal without (*Eig*) and with (*Eig & FPRed)*) false positive reduction. The results show the mean and the standard deviation of A_z values.

of mean and standard deviation of A_z values. Note that the inclusion of the false positive reduction step in some cases slightly decreases the performance of the proposal. This is due to the above mentioned fact, where a mass which was correctly detected using the proposal, was then considered as normal tissue by the false positive reduction algorithm. When this is not the case, the obtained A_z is increased.

5.5.2 Training and Testing using Different Databases

We analyze in this section the performance of the algorithm when using different databases for learning the size and shape of the masses and for detecting them in

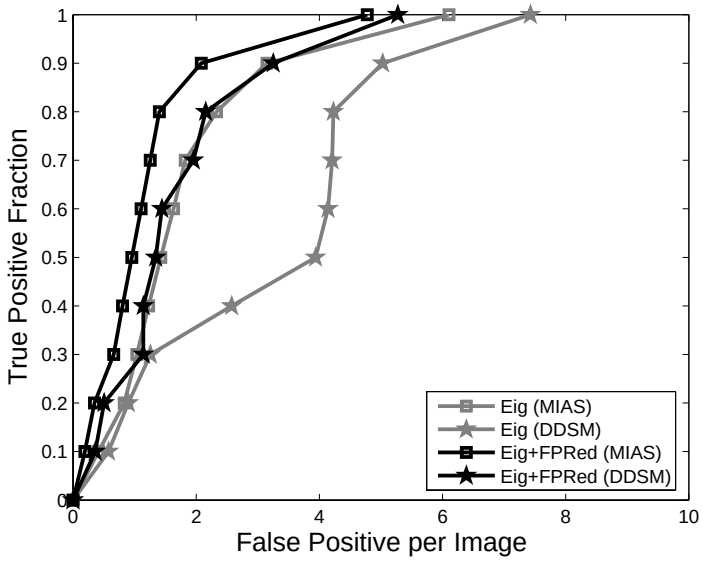


Figure 5.9: Influence of the training database. MIAS is referred to training and testing the systems using the same database (the MIAS one), while DDSM is referred to training and testing using different databases (the DDSM for training and the MIAS for testing).

the images. For such task, we used the MIAS database to test the system and the DDSM [72] one for training. Due to the large mass variability of DDSM database we used six different sizes to train the system: $< 0.10\ cm^2, (0.10-0.60)\ cm^2, (0.60-1.20)\ cm^2, (1.20-1.90)\ cm^2, (1.90-2.70)\ cm^2, > 2.70\ cm^2$, and the number of masses in each interval was respectively, 28, 32, 37, 57, 69, and 33 masses. Moreover, for the false positive reduction step, 3 normal RoIs for each mass RoI were included in each size-cluster.

Figure 5.9 shows the performance of the algorithm . The grey lines show the performance of the proposal without the false positive reduction step, while the black ones including it. The lines with squares are obtained when training and testing with the same database, while the lines with pentagrams when training and testing using different databases. We can see that the Bayesian pattern matching has more false positives per image when is trained with different database. This is basically due to the fact that we are now training with more sizes and, further, smaller patterns. Thus, there is a large set of small regions being normal tissue

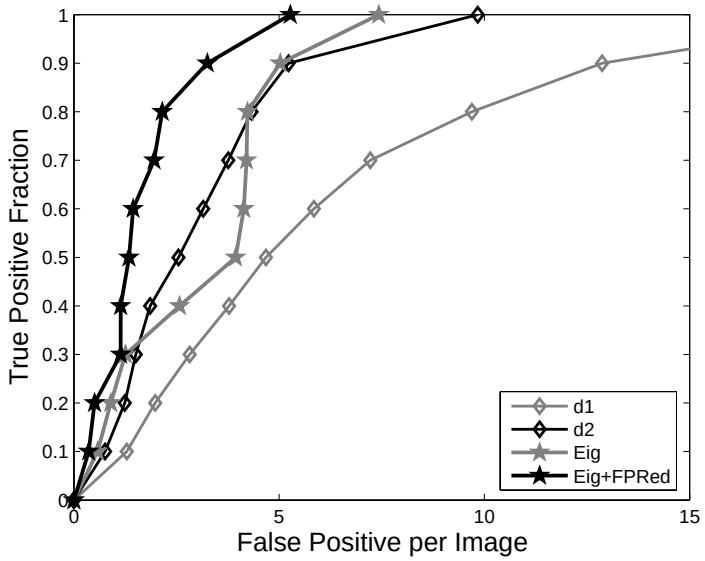


Figure 5.10: FROC analysis comparison using algorithms *d1*, *d2*, and the proposal without (*Eig*) and with (*Eig & FPRed)*) false positive reduction but using different databases for training and testing.

but detected as suspicious by the algorithm. However, the false positive reduction step allows to greatly reduce this number, although the performance is slightly worst compared to training and testing with the same database. For instance, at a sensitivity of 0.8 the number of false positives per image when training and testing using different databases was 4.22 without the false positive reduction algorithm and 2.15 when including it, while when training and testing using the same database the false positives were 2.33 and 1.40, respectively.

Figure 5.10 shows the comparison of the algorithm trained with DDSM when testing the set of 120 mammograms from MIAS database and algorithms *d1* and *d2*. Note that the performance of the proposal *Eig* is similar to algorithm *d2* at sensitivities around $0.8 - 0.9$. In contrast, is clearly better at higher sensitivities and worst at intermediate sensitivities. Note that when including the false positive reduction step the performance is clearly better.

On the other hand, using ROC analysis for the set of 40 mammograms containing masses, we found that mean A_z without false positive reduction was 86.6 ± 6.9, while including it was 86.2 ± 7.3. This results are slightly worse compared to the algorithm

Alg.	Lesion Size (in cm^2)			
	<1.20	**1.20-1.80**	**1.80-3.60**	**>3.60**
Eig (MIAS)	91.3 ± 7.4	90.3 ± 3.3	89.6 ± 4.7	85.5 ± 5.5
Eig* & *FPRed (MIAS)	89.9 ± 3.1	91.4 ± 2.1	88.5 ± 5.0	86.6 ± 3.2
Eig (DDSM)	80.6 ± 9.8	91.1 ± 3.8	86.2 ± 4.5	87.5 ± 3.2
Eig* & *FPRed (DDSM)	81.4 ± 9.4	89.4 ± 3.9	86.0 ± 5.5	88.4 ± 3.5

Table 5.5: Influence of the lesion size (in cm^2) for the proposed algorithms when training and testing using the same (MIAS) or different (DDSM) databases. The results show the mean and the standard deviation A_z values.

trained and tested using the same database ($A_z = 89.3 \pm 5.9$ and $A_z = 89.1 \pm 4.1$, respectively) and also compared to algorithm *d2* ($A_z = 88.1 \pm 8.4$). However, note that this algorithm is still trained and tested using the same database. On the other hand, both proposals outperforms algorithm *d1* ($A_z = 84.1 \pm 7.9$).

Table 5.5 shows the mean A_z values detailed per mass size when the training and testing was done using the same database or using different databases. Note that the main performance drop is for the smaller masses, where the mean A_z is reduced around 10 units. This is basically due to the number of false positives detected by the template matching algorithm at small template sizes. The false positive reduction step allows to decrease the number of false positives, although in the cases where this algorithm increases the number of false negatives (classifying a true mass as normal tissue) the mean A_z of the system is drastically reduced. For the other sizes, the performance is similar when the training and testing was done using the same database or using different databases, and also in some cases, the performance is better when using different databases.

5.5.3 Computational Cost

In this section we roughly explain the computational cost of the proposed algorithm. The cost of the global algorithm can be divided into two different blocks: one for training the system and the other one to match the templates.

For training the system the algorithm firstly computes the eigenfaces and their contours for creating the templates and their deformations. On the other hand, it also computes the 2DPCA approach for the false positive reduction model. Both process are relatively fast, taking around one minute in conventional computers

(Windows-based P-IV, and programming in Matlab).

On the other hand, the matching process is slower than the training step. The matching first reduces the mammograms using a multilevel approach, obtaining a number of possible RoIs. Subsequently, these RoIs are dealt with using bigger resolution. Thus, the amount of time needed in this step highly depends on the number of these suspicious regions. For instance, the mean of all 120 cases takes about five minutes per image.

5.6 Discussion

We have presented a new strategy which is a generic, simple and cost-effective method for false positive reduction. The strategy consists on training a classifier with RoIs representing masses and normal tissue, and using a statistical-based approach to classify a new query RoI as belonging to one of the training set.

We have evaluated two different algorithms: one based on PCA and the other based on 2DPCA. The performance of the system has been evaluated using a leave-one-out methodology and ROC analysis calculated at different ratios of RoIs with masses and RoIs depicting normal tissue. The obtained results demonstrate that, for false positive reduction, the 2DPCA approach outperforms the traditional PCA.

Moreover, we have integrated the approach into the algorithm developed in the previous chapter. The performance of this integration has been evaluated using FROC and ROC analysis, obtaining promising results. Moreover, we have tested the system using the same and different databases for training and testing. We noted that when using different databases the number of false positives as well as the performance of the system decreased, particularly in the smaller masses. This is probably due to the fact we trained the system with a larger number of smaller masses compared to the ones found in the MIAS database, which are slightly larger in size.

Chapter 6

Automatic Mass Segmentation using Breast Density Information

This chapter explains the result of including the breast tissue information into the algorithm for mass detection developed in Chapter 4 and the false positive reduction developed in Chapter 5. The breast density information is computed using the algorithm presented in Chapter 3. We will prove the effectiveness of introducing the breast density information into the proposed mass segmentation scheme.

6.1 Introduction

In the conclusions of Chapter 2 we showed that the sensitivity of most of the mass detection algorithms decreases as the density of the breast increases. In order to take advantage of such information, we introduce a breast density classification step with the goal of increasing the overall breast mass detection. This step consists in a first classification of the database of RoIs according to the breast density parameter. Therefore, we can divide our database of images, not only according to the shape but also based on their breast density.

Note that this training step needs a huge database of mammograms with annotated masses. However, the MIAS database does not provide this set. For this reason, in this chapter we will train the system using the DDSM database, while the MIAS will only be used for evaluation. Moreover, in order to evaluate the improve-

ment of using this information, we will repeat our experiments of mass detection for both situations: the first one in which the original database of RoIs is directly used and the second one in which the database of RoIs is previously classified according to the breast density.

The rest of this chapter is structured as follows. Section 6.2 describes in more detail this pre-classification step. Afterwards, Section 6.3 and Section 6.4 explain the obtained results for the mass detection and the false positive reduction algorithms respectively. In Section 6.5 the results obtained using a new digital database are described. This will prove the effectiveness of our method also on this kind of images. The chapter ends with a discussion.

6.2 Including Breast Density Information in our Mass Detection Approach

The sensitivity of mass detection algorithms decreases as the density of the breast increases. In order to deal with this problem an initial pre-processing of the mammogram can be done estimating the density of the breast. This pre-processing may be performed manually or, as shown in Chapter 3 automatically. In both cases, this parameter is known before the detection algorithm begins.

Once the breast density is known, the question is how to introduce this information into the algorithm. Obviously, there is not an unique solution, and actually this will be algorithm-dependent. A way forward could be fine-tuning the parameters of the algorithm depending on the tissue type.

In our case, however, this information is easier to introduce. We cluster both RoIs databases (the first one containing only masses for template creation and the second one containing masses and normal tissue for false positive reduction) according to the breast density parameter. In order to avoid the need of manual intervention, this tissue classification step is automatically done using the automatic classification approach analyzed in Chapter 3. In Section 6.5 we compare the obtained results using this algorithm and the ones obtained using the experts manual classification. Note that this cannot be done with the MIAS database as there is not a significant difference between both classification (only one of the forty mammograms with

masses was classified in different BIRADS categories using the manual annotations and the automatic algorithm).

6.3 Results Obtained Including Breast Tissue Information

The performance of our approach is evaluated using a total of 120 mammograms extracted from the MIAS mammographic database [183]. Among them, 40 show confirmed masses (the ground-truth provided by an expert) while the rest are normal mammograms. It is important to note that the MIAS database has been used for evaluation as we have accurate manual annotations for the lesions. However, the number of cases found in the MIAS is small for building the statistical models for detection and false positive reduction steps using breast density information. The DDSM database [72], on the other hand, presents less accurate annotations but a larger number of cases. For this reason, the DDSM database has been used for statistical training (remember that the algorithm does not need an accurate set of manual annotations) and the MIAS for testing the mass detection accuracy. Hence, two databases of RoIs have been extracted from the DDSM database containing both masses and non-masses. We are clustering the DDSM database in 4 BIRADS classes, and each class is clustered according to the size. Again, we used six different sizes according to the lesion size: $< 0.10\ cm^2$, $(0.10 - 0.60)\ cm^2$, $(0.60 - 1.20)\ cm^2$, $(1.20 - 1.90)\ cm^2$, $(1.90 - 2.70)\ cm^2$, $> 2.70\ cm^2$, and the number of masses in each interval was respectively, 28, 32, 37, 57, 69, and 33. Moreover, for the false positive reduction learning step, 3 normal RoIs for each mass RoI were included in each size-cluster.

The evaluation is again done using Free Receiver Operating Characteristics (FROC) and Receiver Operating Characteristics (ROC) analysis. Remember that FROC analysis quantifies the performance of the algorithms to distinguish between mammograms with and without masses, while a ROC curve indicates the accuracy in which the masses are detected.

In Figure 6.1 the performance of the presented algorithms is evaluated. The grey line with pentagrams shows the proposed template matching performance, obtaining a high number of false positives per image (regions marked as suspicious but

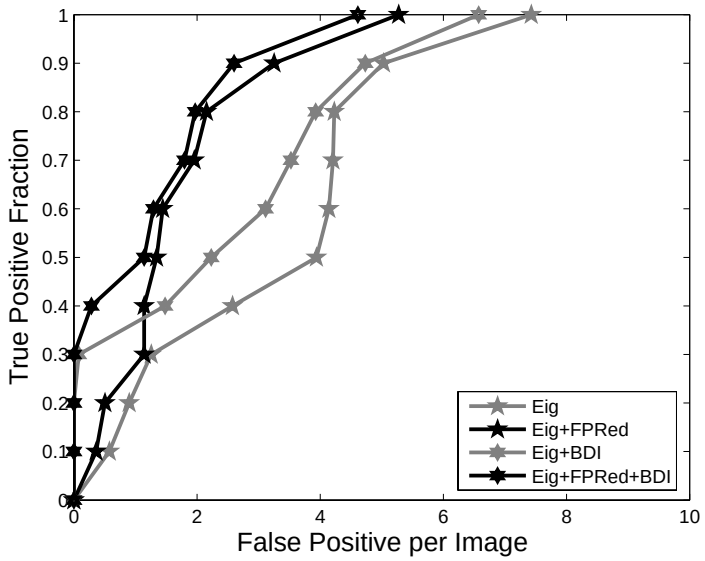


Figure 6.1: FROC analysis of the proposed algorithms over the set of 120 mammograms. The grey lines show the results obtained using the template matching algorithm, while the black ones show the proposed algorithm with false positive reduction. Lines with pentagrams are the result of the algorithm without the breast density information, while lines with hexagrams are the obtained ones including this information.

being normal tissue). This number is clearly reduced by the false positive reduction algorithm, the black line with pentagrams. The lines with hexagrams are obtained when including the breast density information. Note that including the breast density information the performance for both approaches is improved. For instance, at a sensitivity of 0.8 the performance for the algorithm without false positive reduction increases from 4.22 false positive per image to 3.92, while when including the false positive reduction step goes from 2.15 to 1.96.

A comparison between our approach and the algorithms *d1* and *d2* is also provided in Figure 6.2. Our approach (the black line with hexagrams) outperforms both algorithms, obtaining an intermediate performance between the results shown in Figure 5.7 and in Figure 5.10. One should recall that those figures were related to results using the same database for both training and testing (Figure 5.7) and to results using different databases (Figure 5.10). For instance, at a sensitivity of 0.8 now the mean number of false positive per image is 1.96, which is an intermediate

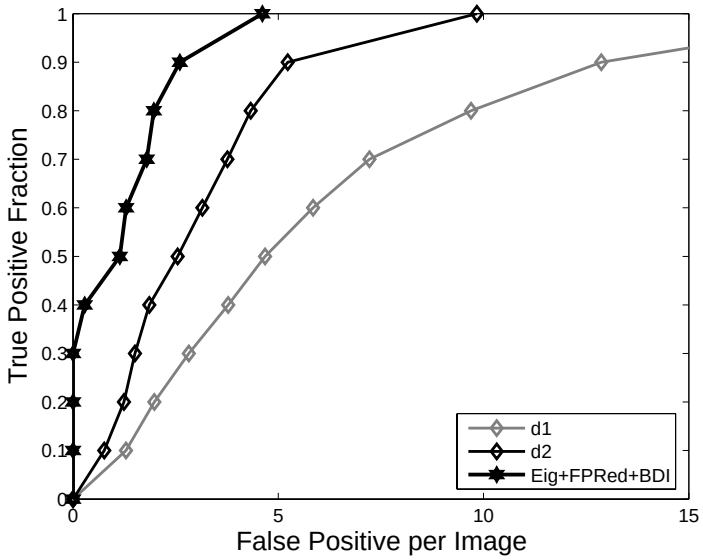


Figure 6.2: FROC analysis of the algorithm over the set of 120 mammograms. The two lines with rhombus show the results obtained using *d1* and *d2*, while the line with hexagrams shows the proposed algorithm.

value between 1.40 when training and testing using the same database and 2.15 when using different databases.

Once the mammograms containing masses are detected, a ROC curve and the corresponding A_z analysis is performed. The overall performance over the 40 mammograms containing masses resulted in a A_z value of 86.2 ± 7.3 and 88.0 ± 6.4 without and with considering breast density information, respectively. Thus, introducing this information has two effects: firstly, A_z mean is increased, and secondly, the deviation is reduced, showing that this information is also beneficial in those cases where the algorithm has a lower accuracy. Comparing with algorithms *d1* ($A_z = 84.1 \pm 7.9$) and *d2* ($A_z = 88.1 \pm 8.4$) the proposal is clearly better than *d1* and is similar to *d2*, despite the drawback of being trained and tested using different databases.

Table 6.1 shows the effect of the lesion size for the different algorithms in terms of mean and standard deviation of the A_z values. Our proposal has similar performances for each size except for the range $(1.20 - 1.80)$, which obtains the best results. The inclusion of the breast density outperforms the results in all classes

Alg.	Lesion Size (in cm^2)			
	<1.20	**1.20-1.80**	**1.80-3.60**	**>3.60**
d1	92.1 ± 5.5	85.8 ± 8.2	82.4 ± 7.3	79.1 ± 7.2
d2	84.9 ± 8.8	86.7 ± 8.1	89.1 ± 9.6	92.1 ± 6.5
Eig & ***FPRed***	81.4 ± 9.4	89.4 ± 3.9	86.0 ± 5.5	88.4 ± 3.5
Eig & ***FPRed*** & ***BDI***	86.6 ± 9.3	91.1 ± 4.3	87.2 ± 6.1	87.6 ± 3.3

Table 6.1: Influence of the lesion size (in cm^2) for algorithms *d1*, *d2* and the proposed algorithm with and without including breast density information. The results show the mean and the standard deviation A_z values.

except for the > 3.60 group. This can be explained by the fact that the system has more information about the shapes of larger masses when there is more instances in the training database. Note also that the problem of the smallest size described in Section 5.5.2 is now partially solved, obtaining also better results than algorithm *d2*.

We include in Table 6.2 a comparison of the performance of the algorithm according to the breast density. We detail in the table both BIRADS categories (using the consensus opinion of three different radiologists) and the fatty/glandular/dense annotations found in the MIAS database. Note that, independently of the classification criteria used, the performance of the algorithm is mainly independent of this factor. For instance, for BIRADS categories all classes have similar behaviour except BIRADS II where the algorithm performs slightly better. For the three-class annotations the dense class performs slightly worse than the other two. This lower performance in dense class is not that clear using BIRADS categories because the mammograms where the algorithm performs slightly worse are distributed between both BIRADS III and IV.

In Chapter 2 we also concluded that not only the lesion size and the breast tissue but also the shape of the mass affect the performance of the algorithms. In that sense, Table 6.3 shows the performance of the algorithm according to the mass shape: circular or spiculated. Note that the algorithm performs slightly better for circular masses that for spiculated ones, although this difference is not significant.

	Breast Tissue		
	Fatty	**Glandular**	**Dense**
Eig & ***FPRed*** & ***BDI***	88.3 ± 8.7	88.5 ± 7.1	85.8 ± 7.3

	BIRADS I	**BIRADS II**	**BIRADS III**	**BIRADS IV**
Eig & ***FPRed*** & ***BDI***	87.2 ± 6.7	89.3 ± 4.2	86.7 ± 4.5	87.3 ± 7.2

Table 6.2: Influence of the breast tissue in the performance of the proposed algorithm. The top table using the annotations for the breast tissue found in MIAS database while the bottom table using the BIRADS categories. The results show the mean and the standard deviation A_z values.

	Lesion Shape	
	Circular	**Spiculated**
Eig & ***FPRed*** & ***BDI***	88.2 ± 9.3	87.4 ± 8.7

Table 6.3: Influence of the lesion shape for the proposed algorithm. The results show the mean and the standard deviation A_z values.

6.4 False Positive Reduction Step with Breast Density Information

In this section we exhaustively evaluate the effect of introducing the breast density into the false positive reduction step. As already explained, this is done by clustering the DDSM RoIs database according to this parameter. The results are obtained by using the same leave-one-out method explained in Section 5.4, where 256 RoIs depicted a true mass and the rest 1536 were normal, but suspicious tissue.

Figure 6.3 shows the mean A_z value obtained using the leave-one-out strategy and varying the ratio between both kinds of RoIs. Note that, as expected, the performance of both PCA and 2DPCA approaches decreases as the ratio of RoIs depicting masses decrease. For the PCA approach we obtained $A_z = 0.81$ for the ratio 1/1 and $A_z = 0.71$ for the ratio 1/6, while using the 2DPCA approach we obtained $A_z = 0.96$ and $A_z = 0.85$ respectively. Again, the 2DPCA approach obtained better performances than the PCA.

The mean for each cluster size at ratio 1/3 is shown in Table 6.4. The overall performance of the system is up to 0.91. Moreover, a similar trend to the one mentioned in 5.4.2 is observed, with a better classification for larger masses. Thus,

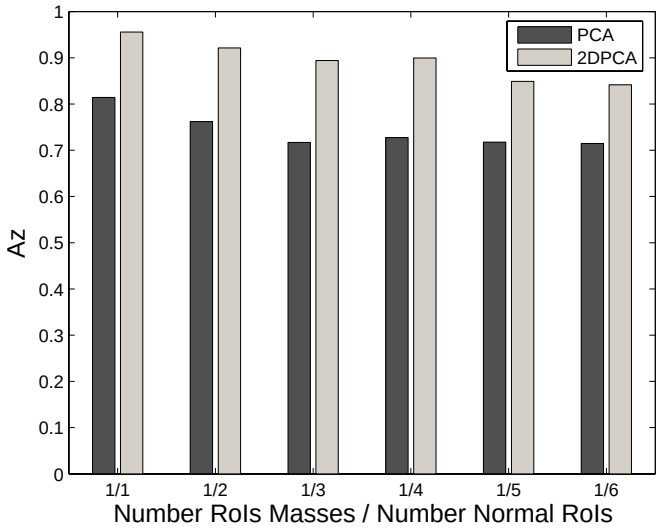


Figure 6.3: Mean A_z value of the system for the DDSM database at different RoIs ratio.

		Lesion Size (in cm^2)					
		<0.10	**0.10-0.60**	**0.60-1.20**	**1.20-1.90**	**1.90-2.70**	**>2.70**
Alg.	***PCA***	0.70	0.71	0.71	0.72	0.77	0.89
	2DPCA	0.88	0.93	0.91	0.92	0.89	0.92

Table 6.4: A_z results for the classification of masses taking the breast density into account at ratio 1/3.

comparing the performance of both results we show that considering the breast tissue obtain an improvement of 0.05 in A_z value.

Figure 6.4 shows the mean kappa value obtained using the leave-one-out strategy at different ratios and threshold 0.5. The same behaviour mentioned for A_z values is repeated, where the performance of both approaches are reduced when increasing the number of normal samples. On the other hand, Figure 6.5 shows a comparison for the 2DPCA approach with and without taking breast density information into account. Note that considering such parameter the results clearly improves. At ratios 1/1, 1/2, 1/3, and 1/4 the agreement can be considered as *almost perfect*, while for ratios 1/5 and 1/6 the agreement is *substantial*. In contrast, without using such information, the agreement for the two latter ratios was only *moderate*, and only the agreement in ratios 1/1 and 1/2 could be considered as *almost perfect*.

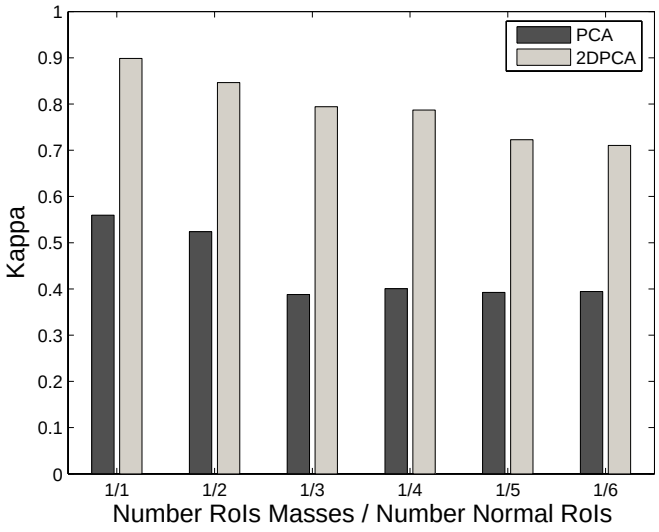


Figure 6.4: Performance of the system for the DDSM database.

		Breast Tissue			
		BIRADS I	**BIRADS II**	**BIRADS III**	**BIRADS IV**
Alg.	***PCA***	0.83	0.91	0.80	0.75
	2DPCA	0.94	0.93	0.94	0.88

Table 6.5: A_z results for each BIRADS category.

Finally, in Table 6.5 a comparison of the performance of the methods with respect to each BIRADS category is shown. Note that for the PCA-based method, mammograms with lower BIRADS were better classified that mammograms with higher BIRADS. This result seems plausible because it is equivalent to the performance of a human expert, as it is well known that experts radiologists have more difficulties to find masses in dense mammograms than in fatty ones. In contrast, the performance of the 2DPCA-based method is more independent of the breast tissue, although for BIRADS IV its performance decreased.

6.4.1 Comparison of the Method with Existing Approaches

We include in Table 6.6 a comparison of the presented 2DPCA false positive reduction approach incorporating breast density information with the performance of the false positive reduction schemes mentioned in Chapter 5, and already compared in

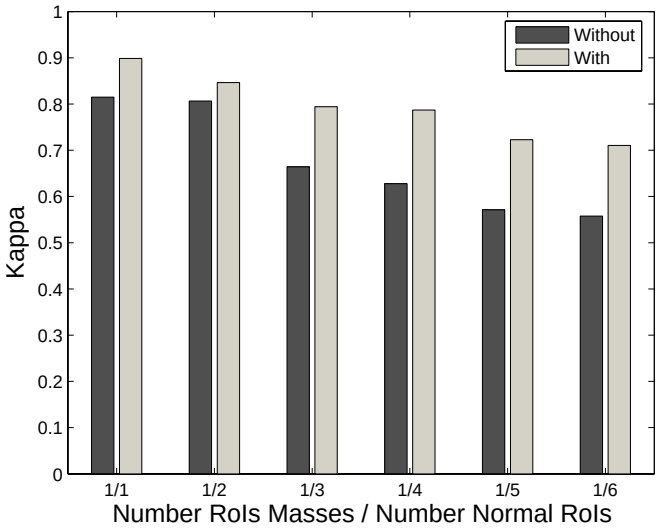


Figure 6.5: Comparison of the false positive reduction algorithm without and with using breast density information.

Section 5.4.2. For instance, without the breast density information, only the work of Sahiner et al. [169] at ratio 1/3 obtained better performance that the 2DPCA approach. Note that using the 2DPCA approach with specific density learning, we now obtain better performances, even using more images. We want to clarify that, however, the methods do not use the same databases and therefore our aim is only to provide a general view of the performance of our approach with respect to those strategies.

6.5 Testing the Approach Using a Full-Field Digital Mammographic Database

In this section, we test our approach using the same training set but using a test set from the Trueta digital database. This is composed by a set of 89 MLO and 87 CC views mammograms containing, at least, one mass.

The evaluation is done using ROC analysis, and the set of 256 DDSM RoIs depicting masses for the templates construction and the rest of RoIs for the false positive reduction model. In order to calculate how the breast density misclassifi-

	RoIs	**Ratio**	A_z
Chang [29]	600	1/1	0.83
Tourassi [195]	1465	$\cong$ 1/1	0.89
2DPCA	1024	1/1	0.96
Sahiner [169]	672	1/3	0.90
Qian01 [155]	800	1/3	0.86
2DPCA	1024	1/3	0.91

Table 6.6: Approaches dealing with mammographic mass false positive reduction, detailing the number of RoIs and the ratio (number of RoIs with masses / number of normal tissue RoIs) used. Further, we include the results obtained with the proposed approach (*2DPCA*) at the same ratio.

Bayesian (84%, $\kappa = 0.74$)

Truth \	B-I	B-II	B-III	B-IV
B-I	42	0	2	0
B-II	5	12	2	1
B-III	0	2	18	0
B-IV	0	1	1	0

(a)

Bayesian (74%, $\kappa = 0.59$)

Truth \	B-I	B-II	B-III	B-IV
B-I	36	4	3	2
B-II	6	11	2	0
B-III	2	1	17	0
B-IV	1	0	1	0

(b)

Table 6.7: Confusion matrix for breast density estimation. (a) MLO views and (b) CC views.

cation affects the performance of the system we will repeat our experiment twice: firstly, considering the breast density as annotated in the database, and secondly, classifying the breasts using the algorithm proposed in Chapter 3.

Table 6.7 shows the confusion matrices for both classifications and MLO and CC views. The algorithm clearly obtained better performance for MLO mammograms than for CC ones. The kappa value for the former is 0.74, which according to Table 3.2 is in the high part of the *substantial* agreement. In contrast, the kappa value for CC views is 0.59 which is on the border between *moderate* and *substantial*. Looking at class level, note that mammograms belonging to BIRADS I are almost all classified correctly for MLO mammograms, while for CC views the performance is reduced. Moreover, the two mammograms belonging to BIRADS IV are misclassified in both confusion matrices.

On the other hand, Table 6.8 shows the obtained results when training the proposed segmentation algorithms using the RoIs clustered according to both annotations: the manual and the automatic. Note that, in general, both results are less satisfactory compared with the ones obtained using the MIAS database (see Ta-

		Breast Tissue Estimation	
		Trueta's Annotations	**Automatic Estimation**
View	**CC**	78.0 ± 14.7	79.3 ± 13.6
	MLO	81.7 ± 13.3	83.6 ± 13.1

Table 6.8: Comparison of the algorithms performance (in terms of A_z) using manual or automatic breast density estimation.

ble 6.1). The main reason for this is due to the false positive reduction algorithm, which is still trained using digitized RoIs in contrast to using digital ones. It is the same effect we noticed when comparing the results obtained with the MIAS database but even more pronounced.

Comparing the results according to the annotations origin, note that the results obtained using the automatic estimation outperforms in almost 2% higher than the ones obtained using the manual annotations. This shows that the automatic method is able to capture the mammogram appearance with more objectivity than a human expert, although the mammogram will probably be misclassified according to the experts opinion.

6.6 Discussion

In this chapter we have shown the benefits of considering the internal breast density of the mammograms. The comparison between the results obtained using this information clearly outperforms the ones obtained without using it.

Moreover, the robustness of our mass detection approach has been reported in the sense that the algorithm is able to find masses in a mammogram independently of the database used for training the system. However results show that the false positive approach is more sensitive to the database used. We consider that this is due to the nature of both algorithms: the template matching is based on the adjustment of the contours of the template and the image, while the false reduction approach is mostly based on region information (grey-level values) which are highly dependent on the database used.

Chapter 7

Conclusions

In this last chapter we present the summary of the thesis and the extracted conclusions. Moreover, we describe the future directions of our work. Besides, a list of the publications directly related to this thesis is given.

7.1 Summary of the Thesis

The aim of this thesis has been the development of a reliable tool to help radiologists to detect breast cancer in mammographic images. We began studying and analyzing the proposals found in literature. From this study, we concluded that none of the proposals had optimal performance for all cases. Moreover, we showed that the shape and size of the masses and the breast density are parameters which affect the performance of those methods.

From these conclusions, we developed a new algorithm which takes these three parameters into account. The shape and size of the masses are learned creating a set of templates which subsequently are matched in the mammogram. This matching step is performed using a Bayesian approach. Moreover, as this step provides a large number of false positives (suspicious regions being normal tissue) we developed a novel false positive reduction algorithm based on the use of the recently developed 2DPCA approach.

To integrate breast density information into the algorithm we first studied the existing proposals. We found that there were not many approaches which classify

the breast according to BIRADS categories, which is the standard currently used by radiologists to classify breasts internal tissue. Thus, we developed a new method based on grouping the pixels according to their appearance (fatty or dense). Afterwards, texture features were extracted from each cluster and used to classify the breast into the BIRADS categories.

Once the mammogram is classified according to its density class, the proposed mass detection algorithm is used to detect the masses, but now only trained using RoIs belonging to the same density class. Results obtained using the DDSM database to train the system, and the MIAS and the Trueta databases for testing, demonstrate the feasibility of our proposal.

Moreover, the fact that we train the system using a different database shows the robustness of the approach. We also conclude that the false positive reduction approach is not as robust as the template matching algorithm, because the performance of the 2DPCA approach looses effectiveness when training and testing with different databases. In contrast, the proposed template matching does not highly depend on this aspect.

As a final overview of the developed work, Table 7.1 shows a comparison of the performance of our approach based on the set of 120 mammograms extracted from the MIAS database, 40 with confirmed masses and 80 being normal mammograms. In the table, the training database is specified (MIAS or DDSM) as well as the used approach: the Bayesian pattern matching Eig, Eig with false positive reduction $FPRed$, and including the breast density information BDI. Mean A_z and the number of false positive per image at a given sensitivity (0.8) are included in the table.

The best results in both A_z and false positives per image is obtained by using the same database for training and testing the system. In concrete, the best accuracy is obtained without using the false positive reduction step, while the smallest false positive number is obtained when including it. Training using a different database implies to both reduce the mean A_z and increase the number of false positives. This is due to the different nature of the databases, for example that the training database has more smaller masses than the testing one, and this implies a high number of false positives at smaller sizes. On the other hand, the introduction of the breast density information results in an increase of the accuracy in which masses are detected as

Table Summary

Proposals	Training	A_z	FPI(Sensitivity=0.8)
Eig	MIAS	89.3 ± 5.9	2.33
Eig & ***FPRed***	MIAS	89.1 ± 4.1	1.40
Eig	DDSM	86.6 ± 6.9	4.22
Eig & ***BDI***	DDSM	88.2 ± 5.4	3.92
Eig & ***FPRed***	DDSM	86.2 ± 7.3	2.15
Eig & ***FPRed*** & ***BDI***	DDSM	88.0 ± 6.4	1.96

Table 7.1: Table summary of the proposals testing the 120 mammograms of MIAS database, 40 with masses and 80 being normals.

well as a decrease in the number of false positives per image.

7.1.1 Contributions

Thus, we consider that the main contributions of this Thesis are:

- An extensive survey of mass segmentation algorithms, which are classified in both the strategy and the features used to segment the mammograms. Moreover, from the quantitative comparison of eight of these methods we prove that the breast density and the size and shape of the masses are three parameters which significantly influence the performance of the algorithms.

- A survey on breast classification methods. We review the main methods found to classify the breasts according to their internal density, highlighting their strategy, features, and the classification used. Moreover, from the quantitative comparison of the strategies, we found that grouping according to the pixel appearance outperforms current strategies.

- A new algorithm for breast density classification based on tissue segmentation and a posterior ensemble classification algorithms, which is exhaustively evaluated using MIAS and DDSM database.

- A new proposal on mass detection, which takes the three parameters above mentioned into account: the breast density which is known a priori, and the shape and size of the masses which are known during the matching step using

the proposed Bayesian algorithm. The performance is tested using MIAS and Trueta databases while DDSM is used for training.

- A new false positive reduction algorithm based on the 2DPCA approach. We have shown that this algorithm performs better when training and testing with the same database.
- Finally, as a result of the close relationship with the Radiologic Department of Hospital Josep Trueta of Girona, a new full-field mammographic database has been compiled and made available.

7.2 Further Work

The design of both a mass detection algorithm and a breast tissue classifier needs to consider an important number of issues. In addition to the various solutions which have been adopted and described in this thesis, different ideas and approaches have been described, analyzed, and discussed throughout this work. Nevertheless, an important number of ideas remain also undeveloped and need to be further analyzed and investigated in depth. For instance, in the design of the probabilistic template matching algorithm we used a given potential function, which can be modified in order to obtain a better performance. We have also assumed that the probability of finding a lesion is independent on the position and, in contrast, it is commonly accepted by the medical community that lesion distribution is not homogeneous. Thus, a prior probability map could be used ot introduce this information into our proposal.

Hence, future directions are organized in three different blocks: further work related to increase the reliability of the proposal, future research lines departing from this thesis, and technological further work.

7.2.1 Increasing the Reliability of the Proposal

This kind of further work is directed to increase the performance of the algorithm. In this sense, a set of different directions are possible: to improve the false positive reduction approach, to reduce the computational time, to change the initial breast

profile segmentation algorithm, or also to use more than one mammographic view in order to increase the performance of the full algorithm.

As we have shown, the performance of the false positive reduction algorithm is database dependent. Possible further work is directed to reduce this dependency, which could be done applying the standard mammogram form (also known as the h_{int} approach) proposed by Highnam and Brady [77]. This approach relies on a detailed knowledge of the mammographic systems and the imaging parameters, and as such might be less appropriate for mammograms where this information is not available (see also Blot and Zwiggelaar [16] and Highnam et al. [79] for a detailed discussion). In these later cases, only the study of common features in both databases will provide information for the correct normalization of both databases.

The reduction of the computational cost is necessary if we aim at obtaining an online tool, otherwise this is less important because we can execute the algorithm as a batch process which can be finished before the experts read the images. We are almost sure that translating the algorithm to a more efficient programming environment like C++ (instead of using Matlab) will improve the overall computational cost. On the other hand, a more difficult and also more interesting problem, is the fact that if we want to add new representative cases in the training database all models need to be recomputed. Thus, additional further work will focus on the development of an incremental training step.

The third mentioned further direction of this group is to apply the new proposal of breast profile because the one used in this work removes some pixels near the skin-line (see Appendix A for a detailed discussion). However, the use of this algorithm will not represent a great change to the obtained results of the mass detection algorithm, although the performance of the breast tissue classifier algorithm could be greatly modified, because we are sure that those pixels represent a new cluster for the Fuzzy C-Means algorithm. Thus, probably, we would have to use three classes instead of the two used in this thesis, which will represent: fatty tissue, dense tissue, and pixels near the skin-line. If this is the case when segmenting the images, only information coming from dense and fatty classes will be useful, and thus, the rest of the classifier algorithm (features used and classifier combination) will still be useful.

Finally, we would like to improve the performance of the algorithm using information coming from the analysis of other mammographic views, either from the same

breast or from the complementary one. In this sense, while the CC-MLO registration needs a major study, we can used different already developed algorithms [120, 121] to obtain the MLO-MLO registration.

7.2.2 Future Research Lines Departing from this Thesis

The second group of further work focuses on the development of new research directions stemming from this thesis. We think that these can be classified again in four different categories: looking for masses in the pectoral muscle, looking for other types of lesions, diagnosis of the found masses, and, finally, breast density quantification.

We have focused this work on looking for masses only in the breast regions. However there are some marginal cases where masses also appear in the pectoral muscle. We actually think that our approach would be useful also for finding this kind of masses, although we have not tested this. This assumption relies on the fact that the pectoral muscle appearance is more uniform than the breast tissue. A major problem would be those situations were the mass is located in the pectoral-breast boundary, as its neighbourhood will have very different grey-level values.

Another research direction will be the application of the developed algorithm to the detection of other types of mammographic lesions, like micro-calcifications or spiculated lesions. At a first glance, both the template approach and the false positive reduction methods could be easily applied for both kinds of diseases. However, more investigation has to be done in order to validate those assumptions.

The third aspect is the diagnosis of the found masses, which clearly is distinct from both explained further works above, where we only talk about detection. Once the masses have been found they can be diagnosed. For such a task, the shape and margin of the lesion have to be accurately studied. We have shown that our algorithm extracts a "rough" information of the mass, including the centre of masses and the lesion size. The traditional way to capture the shape of the lesion starting from the obtained bounding box would be the use of a snake, like the ones independently proposed by Kobatake et al. [95] and Sahiner et al. [168, 170]. However, we think that our template algorithm might be capable of adapting correctly to the mass if it is initialized using an accurate manual segmentation.

Finally, the fourth future topic is the breast density quantification using not only X-ray information but also using Magnetic Resonance Imaging (MRI). The key point here is to use the segmentation strategy not only in mammographic images but also in MRI.

7.2.3 Technological Further Work

The last further direction is the implementation of the proposed algorithms into a clinical practice such as the Hospital Josep Trueta, where we have already implemented the HRIMAC project. Thus, the initial step will be the integration of the algorithms in this framework.

Moreover, after the implementation of the proposals, we will have a great and large database to clinically test them. In fact, we expect that using a huge RoIs training database the algorithm will increase its performance, because with a greater training set the learning process will have more representative instances, and any new case will more likely have similar appearance to some of the templates.

7.3 Related Publications

A list of publications of the author for the PhD candidacy is given below, ordered according to their topic.

Publications related to the mass detection and false positive reduction proposals:

- [**IbPRIA 2007**] A. Oliver, X. Lladó, J. Martí, R. Martí, and J. Freixenet. False Positive Reduction in Breast Mass Detection using Two-Dimensional PCA. Iberian Conference on Pattern Recognition and Image Analysis, to appear. Girona, Spain. June 2007.

- [**MIUA 2006**] A. Oliver, J. Freixenet, R. Martí, E.R.E. Denton, and R. Zwiggelaar. Mammographic mass eigendetection. Medical Image Understanding and Analysis, pp 71-75. Manchester, UK. July 2006.

- [**ICPR 2006**] A. Oliver, J. Martí, R. Martí, A. Bosch, and J. Freixenet. A new approach to the classification of mammographic masses and normal breast

tissue. International Conference on Pattern Recognition, vol IV pp 707-710. Hong Kong. August 2006.

Publications related to the breast tissue characterization:

- [**TITB**] A. Oliver, J. Freixenet, R. Martí, J. Pont, E. Pérez, E. R. E. Denton, and R. Zwiggelaar. A novel breast tissue density classification methodology. IEEE Transactions on Information Technology in Biomedicine. Submitted in second revision.

- [**MICCAI 2006**] A. Oliver, J. Freixenet, R. Martí, and R. Zwiggelaar. A comparison of breast tissue classification techniques. International Conference on Medical Image Computing and Computer Assisted Intervention, vol II pp 872-879. Copenhagen, Denmark. October 2006.

- [**CVPR 2006**] A. Bosch, X. Muñoz, A. Oliver, and J. Martí. Modeling and Classifying Breast Tissue Density in Mammograms. IEEE Conference on Computer Vision and Pattern Recognition, vol II pp 1552-1558. New York. June 2006.

- [**ICIP 2005**] A. Oliver, J. Freixenet, and R. Zwiggelaar. Automatic Classification of Breast Density. IEEE International Conference on Image Processing, vol II pp 1258-1261. Genova, Italy. September 2005.

- [**IbPRIA 2005**] A. Oliver, J. Freixenet, A. Bosch, D. Raba, and R. Zwiggelaar. Automatic Classification of Breast Tissue. Iberian Conference on Pattern Recognition and Image Analysis, LNCS 3523, pp 431-438. Estoril, Portugal. June 2005.

- [**CARS 2005**] A. Oliver, J. Martí, J. Freixenet, J. Pont, and R. Zwiggelaar. Automatic classification of breast density according BIRADS categories using a clustering approach. Computed Aided Radiology and Surgery. Berlin, Germany. June 2005.

Publications related to the survey of mass detection algorithms:

- [**TMI**] A. Oliver, J. Freixenet, J. Martí, E. R. E. Denton, and R. Zwiggelaar. A Review of Automatic Mass Segmentation in Mammographic Images. IEEE Transactions on Medical Imaging. Submitted.

- [**CCIA 2005**] A. Oliver, J. Freixenet, M. Peracaula, and J. Martí. Mass Segmentation using a Pattern Matching Approach with a Mutual Information based Metric. Congrés Català d'Intel·ligència Artificial, pp 123-130. L'Alguer, Italy. October 2005.

- [**SERAM 2004**] A. Oliver, J. Freixenet, J. Martí, and R. Zwiggelaar. Estudio comparativo de cuatro métodos para la detección automàtica de masas en mamografías digitales. XXVII Congreso de la Sociedad Española de Radiología. Bilbao, Spain. May 2004.

- [**IbPRIA 2003**] J. Martí, J. Freixenet, X. Muñoz, and A. Oliver. Active Region Segmentation of Mammographic Masses based on Texture, Contour and Shape Features. Iberian Conference on Pattern Recognition and Image Analysis, LNCS 2652, pp 478-485. Mallorca, Spain. June 2003.

Other publications related to mammographic works:

- [**IbPRIA 2007**] R. Martí, A. Oliver, D. Raba and J. Freixenet. Breast Skin-Line Segmentation Using Contour Growing. Iberian Conference on Pattern Recognition and Image Analysis, to appear. Girona, Spain. 2007.

- [**IWDM 2006**] I. Muhimmah, A. Oliver, E.R.E. Denton, J. Pont, E. Pérez, and R. Zwiggelaar. Comparison between Wolfe, Boyd, BI-RADS and Tabár based Mammographic Risk Assessment. International Workshop on Digital Mammography, LNCS 4046, pp 407-415. Manchester, UK. June 2006.

- [**IWDM 2006**] R. Martí, D. Raba, A. Oliver, and R. Zwiggelaar. Mammographic Registration: Proposal and Evaluation of a New Approach. International Workshop on Digital Mammography, LNCS 4046, pp 213-220. Manchester, UK. June 2006.

- [**IbPRIA 2005**] D. Raba, A. Oliver, J. Martí, M. Peracaula, and J. Espunya. Breast Segmentation with Pectoral Muscle Suppression on Digital Mammograms. Iberian Conference on Pattern Recognition and Image Analysis, LNCS 3523, pp 471-478. Estoril, Portugal. June 2005.

- [**CARS 2005**] J. Freixenet, D. Raba, A. Oliver, and J. Espunya. Breast profile segmentation based on the region growing approach. Computed Aided Radiology and Surgery. Berlin, Germany. June 2005.

- [**IWDM 2004**] J. Martí, J. Freixenet, M. Peracaula, A. Oliver, D. Raba, J. Espunya, J. Pont, and R. Martí. Automatic segmentation of micro-calcifications based on the fusion of different algorithms over CC and MLO views. International Workshop on Digital Mammography. Chapel Hill, North Carolina. June 2004.

Other works related to various computer vision non-medical aspects:

- [**ICPR 2006**] A. Bosch, X. Muñoz, A. Oliver, and R. Martí. Object and scene classification: what does a supervised approach provide us? International Conference on Pattern Recognition, vol II pp 872-879. Hong Kong. August 2006.

- [**AQTR 2006**] A. Oliver, X. Muñoz, J. Batlle, Ll. Pacheco, and J. Freixenet. Improving clustering algorithms for image segmentation using contour and region information. IEEE International Conference on Automation, Quality & Testing, Robotics, pp 315-320. Cluj-Napoca, Romania. May 2006.

- [**CCIA 2005**] A. Bosch, X. Muñoz, J. Martí, and A. Oliver. Classifying Natural Objects on Outdoor Scenes. Congrés Català d'Intel·ligència Artificial, pp 115-122. L'Alguer, Italy. October 2005.

- [**BMVC 2003**] X. Lladó, A. Oliver, M. Petrou, J. Freixenet, and J. Martí. Simultaneous Surface Texture Classification and Illumination Tilt Angle Prediction. British Machine Vision Conference, pp 789-798. Norwich, UK. September 2003.

Appendix A

Breast Profile Segmentation

As already mentioned in the different chapters, the initial step of the algorithms developed in this work had been the segmentation of the breast profile. In this appendix we describe the used proposal for segmenting the breast from the background and the pectoral muscle, as well as a new approach to accurately segment the breast profile.

A.1 Introduction

In this work all we have centred all efforts on the segmentation of mammographic masses. However, a not easy initial step is breast localization. We can distinguish between works locating the boundary between the background and the breast and works looking for the boundary of the pectoral muscle and the breast (in MLO mammograms).

Nowadays, with the introduction of digital imaging, the background is homogeneous and is really easy to find the breast skin-line. However, for digitized mammograms, this is not an easy task because there is an important amount of non-homogeneous noise in the background (and usually some annotations). Different approaches to this task can be found in literature [28, 51, 73, 107, 129, 142, 171, 210, 213].

On the other hand, a still more complicated problem is to correctly localize the boundary between the breast and the pectoral muscle. In fact, there are some images where is easy to find it because the grey-level intensity of the pectoral muscle

is highly greater compared to the grey-level of the histogram. However, this is not always the case, and in some images only expert radiologists are able to find in. Not too many automatic approaches are found in the literature [51, 52, 89, 104]. See the work of Raba et al. [161] for a more extensive survey of both segmentation types.

In what follows we present two new approaches for breast localization. The first one is a rough approach to segment the breast: only those pixels visually perceptible are segmented, ignoring the pixels of the skin-line boundary. Subsequently, using a region-growing approach, the pectoral muscle is also removed. On the other hand, the second approach performs a better estimation of the skin-line boundary of the breast. It uses scale-space concepts in order to enhance the real boundary.

A.2 A Fast Breast Segmentation Algorithm with Pectoral Muscle Suppression

The algorithm is explained in six points:

- Construction of the intensities histogram. The histogram of a complete mammographic image has the behaviour shown in Figure A.1:
  - In the left (lower intensities values) there is a large peak corresponding to the background pixels.
  - In the middle (grey values) there are the pixels corresponding to the breast itself.
  - In the right (brightness pixels) there is another peak corresponding to the pectoral muscle and annotations.
- A threshold is used to extract the image from the background. The value of this threshold is determined using the minimum value between the first two most important peaks, which are the peaks of the background and the breast tissue.
- A Connected Component Labeling algorithm [40] is used in order to recover the largest region, which will be both the breast and pectoral muscle. In general, this algorithm is useful to find not connected objects in images.

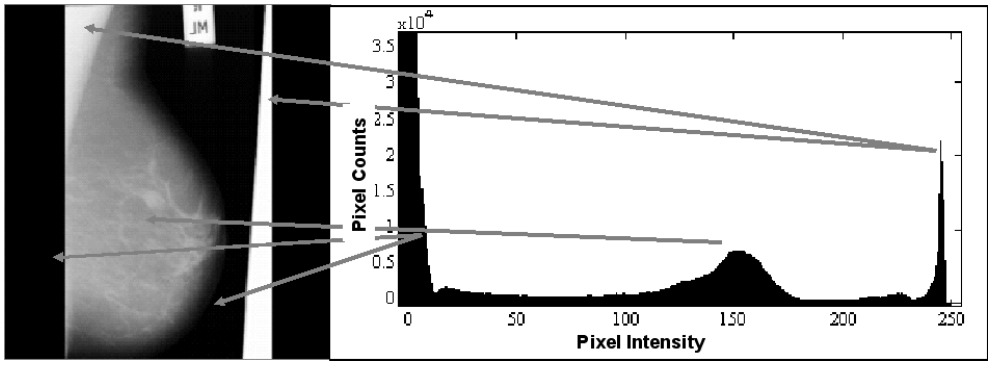


Figure A.1: Typical histogram of a mammogram. Clearly, there are three different zones: background in lowest intensities, breast tissue in medium intensities, and annotations and pectoral muscle in the highest intensities.

- For segmenting the breast from the pectoral muscle a new histogram of this biggest region is used. This histogram contains two zones: the pectoral muscle and the breast tissue.

- A region growing algorithm is used to extract the pectoral muscle region from the breast. The seed of this region growing is placed inside the pectoral with value between the brightness maximum and the minimum between the two zones of the histogram. An automatic control is used in order to adjust the intensity condition of the region growing that permits to identify a pixel belonging to the region or not.

- The last step is the use of morphological operations in order to smooth the boundary of the breast.

Figure A.2 shows a typical mammogram segmented using the above described approach. Its histogram is shown in Figure A.1, and a threshold between the two first major peaks is automatically selected in order to binarize the image, obtaining the Figure A.2(b). The result of applying this threshold is a collection of different regions, being the biggest the union of the breast and the pectoral muscle. This biggest region can be extracted using a CCL algorithm (Figure A.2(c)). In the last image, the breast has been extracted from the pectoral muscle using the region growing algorithm above described.

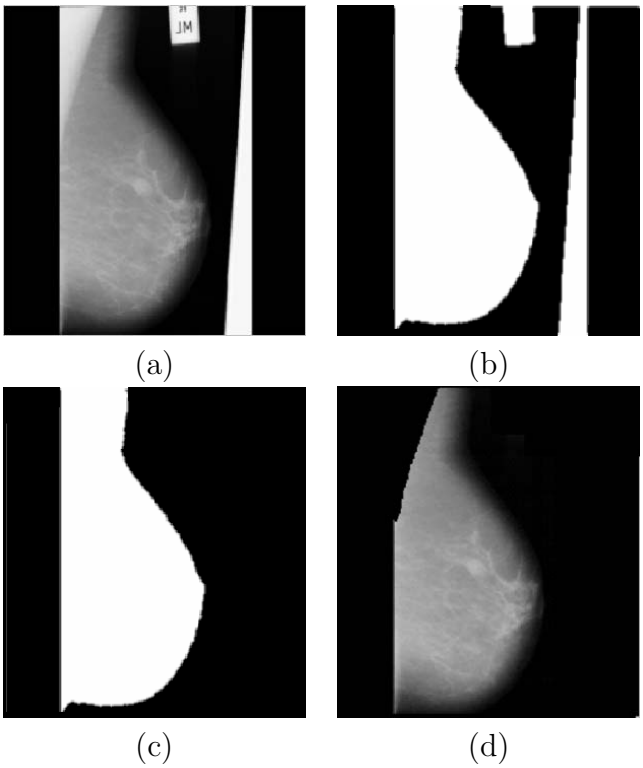


Figure A.2: Sequence of the breast profile segmentation. (a) is the original image, while (b) is the result of thresholding the image. In (c) the CCL algorithm has been applied in order to detect the biggest region, and finally (d) is the segmented image without background and pectoral muscle.

As is shown in Figure A.3, this segmentation results in a minor loss of skin-line pixels in the breast area, but those pixels are deemed not to be relevant for mass segmentation or breast density estimation, as the lost grey-levels are darker than the rest of the pixels of the breast.

A.3 A Contour-Based Approach to Breast Skin-Line Segementation

The idea behind the proposed method is based on finding the skin-line by using a contour growing technique. The growing process is stated following similar concepts of attraction and regularization found in active contours. The method starts by

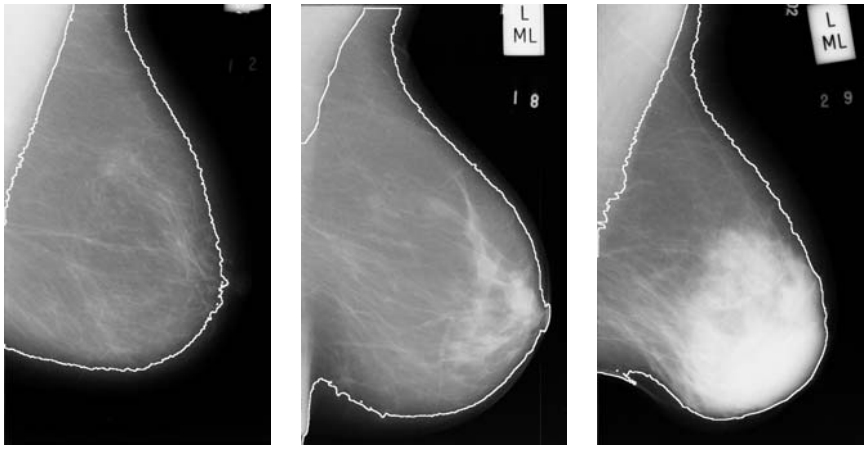


Figure A.3: Three different examples of the breast profile segmentation using the fast segmentatio algorithm.

computing an scale space representation of the image in order to perform edge detection using different scales. Subsequently, an initial seed point lying in the skin-line contour is located based on a robustly estimation process. Using this seed point, a contour growing process starts based on enlarging and adapting a contour using different criteria. Basically, and following the simile of active contours or snakes we adapt the concept of attraction forces (which make the contour enter a region) and what we refer to regularization forces which penalizes rapid curvature and position changes. The final segmentation result is obtained by fitting a cubic spline along the contour points in order to obtain a smooth and continuous result.

A.3.1 Skin-Line Detection in Scale Space

The scale-space representation [113] describes an image as its decomposition at different scales. This is achieved by the convolution of the image with a Gaussian smoothing function at various scales (given by the σ value of the Gaussian function). This representation has been used in conjunction with edge detection in order to automatically extract edges at their optimum scale. If a small scale is used, the edge localization is accurate but results are sensitive to noise. On the other hand, edges at larger scales have a better tolerance to noise but poor edge localization. The motivation of using scale space edge detection is given by the nature of the breast

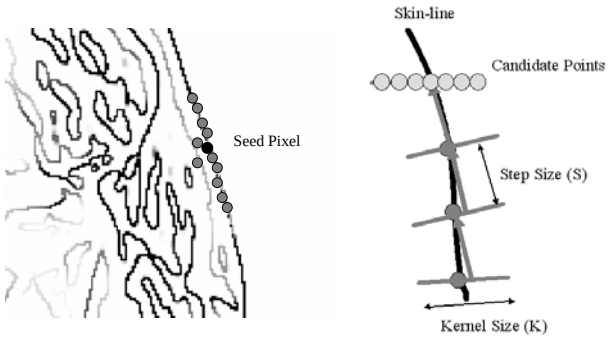
skin-line: a low contrast edge often affected by noise. It is our assertion that using a robust edge detection methodology would lead to a better skin-line estimation. Various approaches to automatic scale selection have been proposed [113]. A simple and common approach is to select as the optimum scale the one which obtains a maximum response from scale invariant descriptors. This is in general given by normalized derivatives, for instance Lindeberg [113] defines

$$L_{norm} = \sigma^{\frac{\alpha}{2}}(L_x^2 + L_y^2) \tag{A.1}$$

as an edge strength measure for scale σ. L_x and L_y are the convolution of the image function with a first derivative Gaussian function in x and y, respectively. Here α is a parameter used as an additional degree of freedom for edge and ridge detection. A typical value of 1 is generally used in the definition of normalized derivatives for edge detection. Edge points are obtained detecting zero-crossing points of the second derivative in the scale-space representation. The final edge strength of a zero crossing will be given by the maximum normalized strength measure along the different scales. This maximum scale is regarded as the edge scale at that particular point.

A.3.2 Seed Point

The first step of the method focuses on finding the starting point (or seed point) from which the contour will start growing. Special care has to be taken on estimating this point which directly affects the accuracy of the segmentation. As stated before, mammographic image segmentation presents difficulties mainly due to the low contrast in the skin-line and to the non-homogeneous background. From our experience this lower contrast is less severe for points close the nipple. Therefore a seed point can be easily detected in points at this area. An initial guess of a seed point is obtained as the first local maxima of the gradient in the scale space representation along the x axis at half the height of the image. Obviously, this first estimation lacks of robustness if this first local maxima does not correspond to the skin-line. That could be the case if the point lies inside the breast area (due to a low contrast of the skin-line) or in the background (due to noise, label and other image artifacts). A more robust approach is adopted based on analyzing the position of various seed



(a) (b)

Figure A.4: Contour growing scheme: (a) Initial seed point and (b) contour growing process.

points at close the same position (at a small range in the y coordinate). The final seed point is obtained using a least median error estimation. Edge direction will also provide an important information in the contour growing process. Therefore the estimation of the initial angle it is also important. In this case a similar least median error estimation is adopted for the angle measure. Figure A.4(a) shows an example of seed detection.

A.3.3 Contour Growing

Once the seed point has been obtained a contour growing process starts based on the combination of different criteria. For each point, a set of candidate growing points are obtained situated in a normal line along the gradient direction. A measure of affinity or cost (C_i) is computed for each point (i) and the value with the minimum cost is taken as the next growing point. This iterative process is illustrated in the Figure A.4(b).

As one may note from the figure the growing scheme incorporates several parameters which need to be defined. These include a kernel size K, normal to the previous point, and a growing step S. Those values have been empirically determined (typical values are $K = 51$ and $S = 20$) and kept constant trough all the experiments.

Also from the experiments we noted that a more robust approach should be used

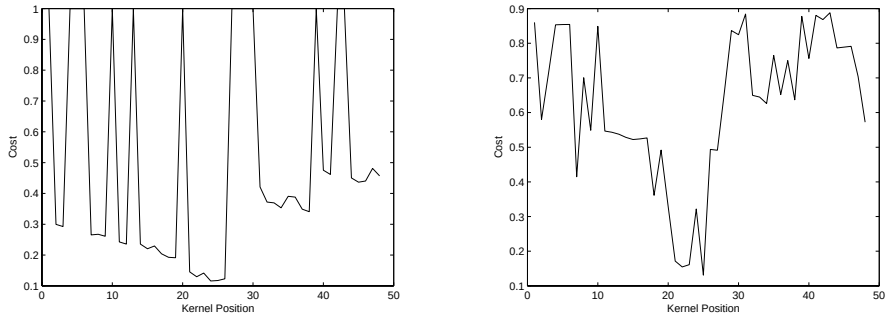


Figure A.5: Cost functions for robust candidate selection: (a) cost without shifting and (b) with cost shifting.

for the process of selecting the next candidate point as it was often affected by noise and outliers. Instead of evaluating only a set of normal points at a given distance and obtain the candidate with a minimum cost over C_i, several sets of points on the normal are evaluated at different positions close to the desired position of the candidate point. A set of cost functions C_i^k is then obtained for each set of normal points (where i means the actual pixel and k the set of normal points). Using this approach the candidate point will be the one with the minimum cost over all the different sets C_i^k. One should note that in the different cost functions, the same (or nearly the same) point can lie in a shifted position. In order to make those cost functions comparable the cost functions are iteratively right and left shifted. The global minimum cost for each point is obtained as the minimum using those shifted functions and the original cost. Figure A.5 shows the minimum cost function of candidate points with and without cost shifting. Note that transportation effects have been minimized when costs are shifted allowing a better estimation of the minimum cost.

Candidate points are obtained from the zero crossing points along the normalized gradient using the scale space representation described earlier. The cost of choosing a candidate point i is given by the following weighted function C_i which includes gradient, intensity, contour curvature and position information. Using this information, the contour tends to grow finding areas of increasing intensity keeping minimal position and direction changes.

$$C_i = \alpha G_i + \beta D_i + (1 - \alpha - \beta) A_i \tag{A.2}$$

where G_i refers to an attraction factor (i.e. intensity or gradient), while the other two respond to regularization terms penalizing position differences (D_i) and direction changes (A_i). The factors α and β are scalar constants which will weight the importance of each term. As in many other approaches using weighted cost functions, it is important to obtain a good estimation of those factors in order to achieve a satisfactory segmentation. The selection of those factors will be later discussed in the paper (see the results section). Different attraction factors can be stated based on the represented information and how it is computed. Here two commonly used attraction factors are evaluated based on gradient and intensity information.

$$G_i = 1 - \exp(-1/f_i) \tag{A.3}$$

where f_i is the gradient or the intensity image function (depending on the factor used) at the given point. Gradient is obtained from the gradient of the zero crossing pixels while intensity information is given by the median intensity value in a local small window (we used $5x5$ pixels).

The segmented breast skin-line should be continuous without having abrupt changes. This obviously corresponds to the continuous nature of the breast. A way to ensure this continuity is to impose some regularization conditions to the contour growing process. This continuity assumption might not hold in all cases (for instance, when the nipple appears in the skin-line) but in this case the attraction factors described earlier will be able to adapt the contour to those changes. The first regularization factor D_i biases the cost to points closer to the centre of the kernel of size K. This means that between two similar points the factor will select as a better point the one with a closer distance to the kernel centre. This factor is independent of the image contents and is given by,

$$D_i = \exp\{\frac{-1}{|(i-1) - (K-1)/2|/((K-1)/2)}\} \tag{A.4}$$

The last regularization term is defined computing the curvature change in a local neighbourhood. Local curvature values (directional change) at each pixel are

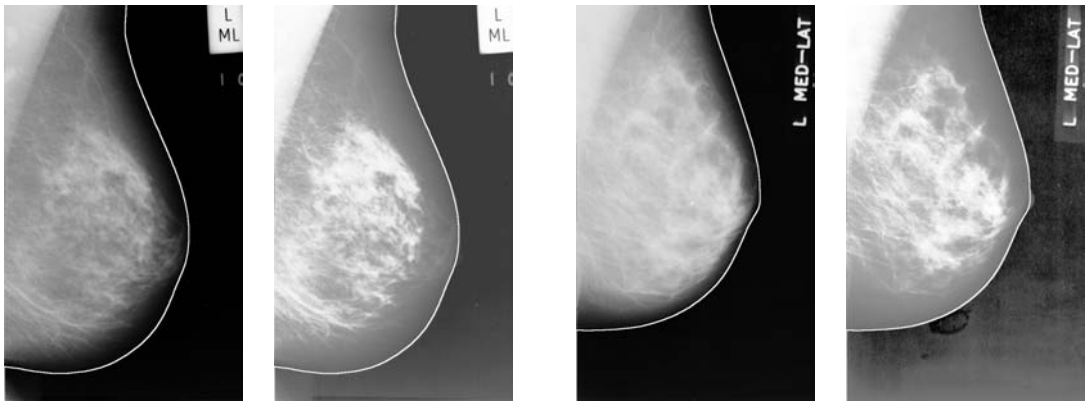


Figure A.6: Two examples of the breast-skin line segmentation. The first images show the segmentation over the original mammogram while the second ones show the segmentation over the enhanced mammograms. Note that the contour accurately follows the breast profile.

obtained with a similar approach as used in the work of Deschênes and Ziou [43]. Directional change between two pixels i and j is defined by the scalar product of their normal vectors. Hence, at a given pixel i the directional change is obtained by computing the scalar product between i and its neighbouring pixels,

$$A_i = \frac{1}{N} \sum_{j=1}^{N} \exp(-d_{ij}^2)(1 - \cos(\phi_i - \phi_j)) \tag{A.5}$$

where ϕ_i is the angle of the normal at a pixel i. N is the number of points in a local neighbourhood and d_{ij} is the Euclidean distance between points i and j. The distance factor is used here to weight the curvature of each point j, in order to incorporate a bias to points closer to i.

Figure A.6 shows an example of the performance of this algorithm. Note that the algorithm seems to segment far to the breast. However, if we equalize the image, the segmentation is accurately adapted to the real mammogram border. The main drawback of the algorithm is that in some cases there is a poor estimation of the initial seed point, due to the large amount of noise in the background an to the non-uniform breast intensity distribution. In these cases, the algorithm does not obtain what could be considered an acceptable segmentation.

Appendix B

A Brief Description of the Used Mammographic Databases

In this appendix we provide a brief description of the four mammographic databases used in this work: two widely known public databases of digitized mammograms and two private databases more, one of digitized mammograms and the other one of full-field obtained digital mammograms.

B.1 Introduction

The main aim of this work was the design of a new algorithm capable to detect masses. Obviously, its comparison with already published works was also necessary to evaluate the performance and quantitatively measure the improvements (if any). For this reason, the use of public mammographic databases was necessary. As shown in this work, MIAS and DDSM were used.

However, as explained in the initial chapter, full-field digital mammography is gaining importance compared to the typical film-based mammography, which is the technique in which both MIAS and DDSM databases were acquired. In order to compare the performance of our algorithms with mammograms obtaining using this newer technology, a digital database was constructed using the mammograms acquired on the Hospital Dr Josep Trueta of Girona.

In what follows we describe in more detail each of these three databases, along

with the Málaga database, which is another used database.

B.2 MIAS

The Mammographic Image Analysis Society Digital Mammogram Database [183] is composed by a set of 322 MLO digitized mammograms corresponding to left and right breast of 161 women.

B.2.1 Database Characteristics

The films were extracted from the UK National Breast Screening Programme, and digitised to 50 micron pixel edge with a Joyce-Loebl scanner, a device with linear response in the optical density range $0-3.2$. Each pixel was described as a 8-bit word. The database also includes "ground-truth" on the locations of any abnormalities may be present.

The mammograms are named as "mdbXXXBS", where:

- "XXX" represent the number of the image ranging from 001 to 322.
- "B" is the side of breast: "l" or "r" (left or right).
- "S" is the size of the image, which can be "s" for small images (1600×4320 pixels), "m" for medium size images (2048×4320 pixels), "l" for large images (2600×4320 pixels), and "x" for extra large images (4000×5200 pixels).

Each image is stored in raw format: each number of the stored file corresponds to the grey-level value (from 0 to 255) of their corresponding pixel in the image. Thus, reading the size of the image from its name, is straightforward to read correctly the image.

The annotations include kind of abnormalities, and the coordinates of their centre of masses and approximate radius of the circle enclosing them. Table B.1 summarizes the mammograms present in this database. In it, we can see the number of mammograms containing calcifications, masses, other kind of abnormalities, and the number of normal mammograms, all distributed according to the kind of breast tissue (fatty, glandular, or dense).

Abnormality	Fatty	Glandular	Dense	Total
Calc.	6	9	10	25
Masses	17	15	10	42
Others	17	15	16	48
Normal	66	65	76	207
Total	106	104	112	322

Table B.1: Summary of the MIAS database.

Masses	Fatty	Glandular	Dense	Total
Circumscribed	12	8	3	23
Spiculated	5	7	7	19
Total	17	15	10	42

Table B.2: MIAS mammograms containing masses.

B.2.2 Used Mammograms Containing Masses

According to the annotations of the database, 42 of the 322 mammograms contains masses. Table B.2 shows their distribution according to the tissue of the database (columns) and the shape of the lesion (rows).

The set of masses belonging to this database and used in this work (summarized in Table B.2), however, has only 40 of these 42 mammograms. This is due to our "ground-truth" annotations, where these 40 mammograms were accurately manually segmented. Moreover, one of the masses classified as circumscribed by the database was classified as spiculated in the text, as the manual segmentation reflected this behaviour.

B.3 DDSM

The Digital Database for Screening Mammography [71] of the University of South Florida is a huge database of digitized mammograms available online.

The database is divided into 43 volumes, and each volume is divided in a number of studies. The grouping factor is the study final diagnosis: volumes with normal cases, volumes with cases containing benign abnormalities, and volumes containing cases with cancerous abnormalities. In total, there are 2620 cases, and each case corresponds to the MLO and CC views of both woman breasts, along with some

Digitizer	Grey-levels (bits)	Sampling Rate(microns)
DBA M2100 ImageClear	12	42
Howtek 960	12	43.5
Lumisys 200 Laser	12	50
Howtek MultiRad850	12	43.5

Table B.3: DDSM digitization characteristics.

associated patient information (age, breast density, rating and keyword description for abnormalities) and image information (scanner, spatial resolution, ...). Moreover, images containing suspicious areas have associated "ground truth" information about the locations and types of suspicious regions.

B.3.1 Database Characteristics

In contrast with MIAS database where a single scanner was used to digitize the mammograms, in DDSM four different scanners were used. Thus, depending to which of them was used, the mammograms have different characteristics. Table B.3 shows some of these parameters, like the number of bits of the image and the sampling rate.

Each case of the database is named as M_XXXX_1, where "M" represents the used scanner (from "A" to "D"), XXXX is a number of four digits. The four images of the case are stored in a Lossless JPEG format, and the own database provides software to transform the images in a typical raw format.

The annotations stored in each case includes information related to the image (date of study, kind of film), related to the patient (patient's age, density of the breast in BIRADS), related to each image of the case (size), and an overlay file if necessary. In such file, the number of abnormalities and their description (kind of abnormalities, assessment, boundary) are explained.

B.3.2 Used Mammograms Containing Masses

A subset of 283 mammograms containing masses were selected from DDSM database and used for testing the algorithms. Table B.4 shows their distribution according to the breast tissue and using the own annotations of DDSM database. The proportion

	B-I	**B-II**	**B-III**	**B-IV**	**Total**
Masses	38	126	87	32	283

Table B.4: Summary of the DDSM mammograms with masses used in this work.

of mammograms of each class follows the typical distribution, where classes II and III tend to be larger that classes I and IV [35, 66, 202].

B.4 Málaga Database

We obtained from the University of Málaga a collection of 35 mammographic studies of a single breast. Each study corresponds to both CC and MLO views of the breast, and each case has, at least, one mass, which is described in an overlay file.

B.4.1 Database Characteristics

The mammograms are stored in a raw format, where each pixel is represented by two bytes (the grey level of each pixel is represented using 12 bits). The size of each mammogram is variable, but it can be known from the overlay file.

Each case is stored using the following nomenclature: XXXXAB.raw, where XXXX is a number representing the case, A indicates if the mammograms provides from a right breast ("R") or from a left one ("L"). The B can be "C" or "O", which respectively imply that the mammogram provides a CC or a MLO view.

The overlays are stored in a tif file, which the same name of the case, but with a constant suffix (_lb). Thus, looking for the size of these files, is possible to know the size of the raw files.

The most interesting fact of this database is that the overlay files are composed by the opinion of six different radiologists. Thus, each overlay file has only seven grey-levels (0 means normal pixels, and from 1 to 6 there is the opinion of the different radiologists), where 1 means that only one radiologist has marked the pixel as belonging to a mass, 2 that two radiologists marked the pixel as mass, and so on. Thus, linearly equalizing the images, the centre of the masses are the most brighter regions of the mass, and then, gradually, grey-level values decrease surrounding them.

B.4.2 Used Mammograms Containing Masses

As already explained, all cases of the database contains masses. Thus, all the MLO and CC views (35 in total) have been used.

B.5 Trueta Database

The Radiologic Department of Hospital Universitary Dr. Josep Trueta is provided with a Siemens Mammomat Novation digital mammographic unit. This mammograph stores the images in a computerized way in the hospital PACS system, thus resulting in a big digital mammographic database. The database is stored in 89 cases, where each case is composed by the two views of the same mammograms (18 cases) or the complete case (71 cases), consisting on both MLO and CC views of both patient' breasts.

B.5.1 Image Characteristics

Each image of the database has 70 micron pixel edge and is stored using 12-bit word. Two different image sizes are possible: 2560×3328 or 3328×4096 pixels, and one size or the other is used according to the patient breast size.

The images are stored according to the DICOM standard [135]. Although different formats of storing the images are possible using this standard, in this case, they are stored in raw format: each number of the stored file corresponds to the grey-level value (from 0 to 4096) of their corresponding pixel in the image.

Each case has its own annotations, including the breast density classified on BIRADS categories, the number of lesions, and if present a lesion, the location of its centre of masses and the minimum radius to evolve the lesion. Moreover, in the cases where the lesion is present, another image is stored in the DICOM which is a 1-bit-depth-image where the bit is 0 for pixels outside the lesion and 1 for pixels inside or on the boundary of the lesion.

		B-I	B-II	B-III	B-IV	Total
Masses	**MLO**	46	21	20	2	89
	CC	45	20	20	2	87
	Total	91	41	40	4	176

Table B.5: Summary of the Trueta mammograms with masses used in this work.

B.5.2 Used Mammograms Containing Masses

A subset of 89 MLO and 87 CC views mammograms containing masses were extracted from this database. Table B.5 shows their distribution according to the BIRADS breast density categories.

Appendix C

Evaluation Methodologies

We present here a brief description of the evaluation's procedures used in the work. Clearly, we can distinguish between two main trends: methods to evaluate the performance of a classifier (used in the evaluation of breast tissue classification and in the evaluation of the false positive reduction) and methods to evaluate the performance of the detection (the final goal of the Thesis).

C.1 Introduction

We show here an explanation of the evaluation procedures done in the Thesis. Firstly, we explain how we evaluated the classifications tasks performed in the work. Note that we did two different works related with classification: classification of the breasts according to their internal density in the four classes used in the BIRADS standard, and the classification of the RoIs when evaluating the false positive reduction method. In the former, we evaluated the classification using confusion matrices, while in the latter we used ROC analysis. Both techniques are analyzed in Section C.2.

On the other hand, we have also results obtained from the evaluation of the detection algorithm. In contrast with the above results, these are obtained from automatically segmented images, and the results are obtained by comparing these images with the corresponding manually annotated image. In this task, ROC and FROC analysis are performed, which are described in Section C.3.

Truth \ Automatic	A	B	C	Total
A	45	9	6	60
B	4	19	7	30
C	1	2	7	10
Total	50	30	20	100

(a)

Automatic	A	B	C	Total
A	30	18	12	60
B	15	9	6	30
C	5	3	2	10
Total	50	30	20	100

(b)

Table C.1: (a) Example of confusion matrix, and (b) the same data classified as would have been expected by mere chance, given the observed marginal totals.

C.2 Evaluation of Classifiers

C.2.1 Confusion Matrices

A confusion matrix is a visualization tool commonly used in supervised machine learning. It contains information about actual and predicted classifications by a classification system. Usually, each column of the matrix represents the instances of the predicted class, while each row represents the instances of the actual class.

The confusion matrix is useful to evaluate the performance of a classifier, showing the number per class of well classified and mislabeled instances. Moreover, it is easy to see if the automatic system is confusing two or more classes (mislabeling one class as another).

An example of confusion matrix is shown in Table C.1(a). Is easy to see, for instance, that 45 instances of Class A are correctly predicted as Class A and, in contrast, 9 and 6 instances of Class A are misclassified as Class B and Class C, respectively. The same can be done for the other classes. The numerical evaluation of a confusion matrix is commonly computed as the percentage of correctly classified instances. For instance, in Table C.1(a) an agreement of 67% is achieved.

Another measure which can be extracted from a confusion matrix is the kappa (κ) coefficient [36, 53], which is a popular measure to estimate agreement in categorical data. The motivation of this measure is to extract from the correctly classified percentage the actual percentage expected by chance. Thus, this coefficient is calculated as:

$$\kappa = \frac{P(D) - P(E)}{1 - P(E)} \tag{C.1}$$

κ	Agreement
< 0	Poor
$[0, 0.20]$	Slight
$[0.21, 0.40]$	Fair
$[0.41, 0.60]$	Moderate
$[0.61, 0.80]$	Substantial
$[0.81, 1.00]$	Almost Perfect

Table C.2: Common interpretation of the various κ values [106].

where $P(D)$ is the percentage of correct classified instances (the sum of diagonal terms divided by the sum of total instances) and $P(E)$ is the expected proportion by chance (the sum of the multiplication of the marginal probabilities per class divided by the sum of total instances). A κ coefficient equal to one means a statistically perfect model whereas a value equal to zero is the chance value. Table C.2 shows a commonly used interpretation of the various κ values [106].

Thus, following the example, Table C.1(b) shows the classification of the instances as expected by mere chance, given the observed marginal totals. Using the above formula $\kappa = 0.51$, which looking at Table C.2 is in the moderate agreement. However, what means kappa? Looking at the example, note that of the correctly classified 71 instances of (a) (the sum of the diagonal values), 41 of them were in fact expected by chance, thus showing that the classifier agrees in 30 more cases. Similarly, the number of mislabeled instances expected by chance is $100 - 41 = 59$. The coefficient κ is simply this ratio (30/59), which can be translate as "of all the 59 items that would have mislabeled by chance, a total of 30 are in fact correctly classified".

C.2.2 ROC Analysis

ROC analysis proceeds from the analysis of a special case of confusion matrix when there are only two classes: the instances can only be *positive* or *negative*. Table C.3 shows graphically a general confusion matrix for this special case. The entries in the confusion matrix have the following meaning:

- a is the number of correct predictions that an instance is positive.
- b is the number of incorrect predictions that an instance is negative (and

		Automatic	
		Positive	**Negative**
Truth	**Positive**	a	b
	Negative	c	d

Table C.3: Example of confusion matrix with only two classes.

actually is positive).

- c is the number of incorrect of predictions that an instance is positive (and actually is negative).
- d is the number of correct predictions that an instance is negative.

For this 2x2 confusion matrix a set of parameters [46] are typically extracted in order to evaluate the result:

- Accuracy: is the proportion of the total number of positive predictions. It is determined as:

$$Accuracy = \frac{a+d}{a+b+c+d} \tag{C.2}$$

- True positive rate (also known as recall or sensitivity): is the proportion of positive cases that were correctly identified, as calculated using the equation:

$$TPR = \frac{a}{a+b} \tag{C.3}$$

- True negative rate (or specificity): is the proportion of negative cases that were correctly identified:

$$TNR = \frac{d}{c+d} \tag{C.4}$$

- False positive rate: the proportion of negatives cases that were incorrectly classified as positive:

$$FPR = \frac{c}{c+d} \tag{C.5}$$

- False negative rate: the proportion of positives cases that were incorrectly classified as negative:

$$FNR = \frac{b}{a+b} \tag{C.6}$$

- Precision: is the proportion of the predicted positive cases that were correct, as calculated using the equation:

$$Precision = \frac{a}{a+b} \tag{C.7}$$

A ROC graph is a plot with the false positive rate on the X-axis and the sensitivity (the true positive rate) on the Y-axis. Thus, each axis ranges from 0 to 1. The point $(x = 0, y = 1)$ is the perfect classifier: it classifies all positive cases and negative cases correctly. The point $(x = 0, y = 0)$ represents a classifier that predicts all cases to be negative, while the point $(x = 1, y = 1)$ corresponds to a classifier that predicts every case to be positive. Point $(x = 1, y = 0)$ is the classifier that is incorrect for all classifications. When no useful discrimination is achieved the true positive rate is always equal to the false positive rate, obtaining thus a point in the diagonal line from point $(x = 0, y = 0)$ to point $(x = 1, y = 1)$.

However, a ROC graph has more information that a single confusion matrix. In many cases, a classifier has a parameter that can be adjusted to increase true positive rate at the cost of an increased false positive rate. Therefore, each parameter setting provides a point on the graph, and varying the parameter a curve is achieved.

Figure C.1 shows an example of a ROC graph with two ROC curves labeled $C1$ and $C2$, and the probability obtained by chance. Curve $C2$ obtains better performance than curve $C1$, as it goes closer to the point $(x = 0, y = 1)$, the perfect classifier. A measure commonly derived form a ROC curve is the area under the curve [20], which is an indication for the overall sensitivity and specificity of the observer, commonly called Az. As closest to the upper-left-hand corner of the graph, the area increases until a maximum area of 1.

C.3 Detection Evaluation

The evaluation of segmentation algorithms is still an open topic [134]. However, in mammography, a common meaning of segmentation is detection, and evaluation of the detection is possible if a database with ground-truth is available. In this case, a database with ground-truth is referred to a database where all cases have been

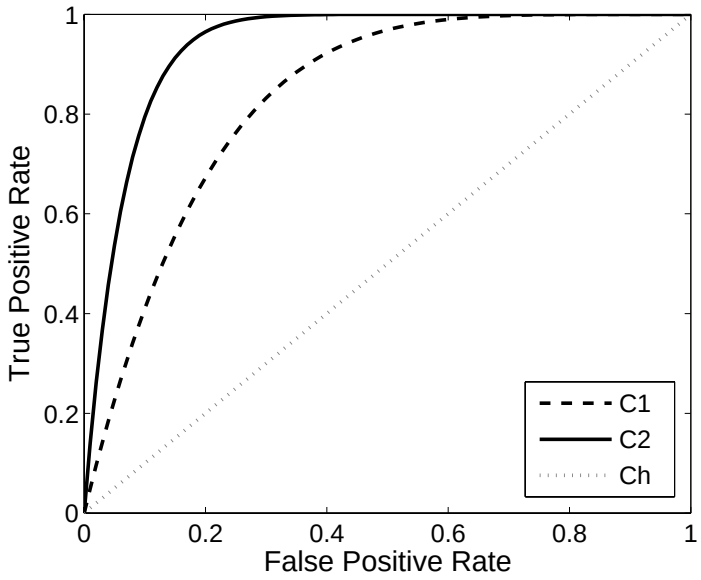


Figure C.1: Two ROC curves and the diagonal line marking the chance classifier.

examined by at least one expert and, if any abnormality has been found, it has been marked and evaluated.

C.3.1 ROC Analysis

ROC analysis for the evaluation of mass detection algorithms is closely related to ROC analysis for classification evaluation. Thus, each pixel of the image is treated as an instance of the classification process, and thus it can be a pixel belonging to a mass or a pixel not belonging to the mass. Therefore, this pixel is compared to the same pixel on the image obtained from the manually segmented image, resulting then in a well-classified pixel or a bad-classified pixel. A pixel is well-classified if in both images its state is the same: mass (in this case, it is a true positive) or not mass (true negative). A pixel is bad-classified if it has different state in both images: if the pixel is classified as not mass by the radiologist and classified as mass in the CAD system (false positive) or the inverse case (false negative). Table C.4 summarizes this notation. Note that is closely related to Table C.3.

		Definition	
		Ground-Truth	**Segmentation**
Notation	True Positive (TP)	Mass	Mass
	True Negative (TN)	No Mass	No Mass
	False Positive (FP)	No Mass	Mass
	False Negative (FN)	Mass	No Mass

Table C.4: Description of the notation used in CAD diagnosis. TP and TN represents the pixels classified correctly by the CAD systems. FP and FN are the bad classified pixels

C.3.2 FROC Analysis

In the past section, ROC analysis has been analysed. Instead of a pixel-based evaluation, Free Response Operating Characteristic (FROC) is based on a region-based analysis [27, 130]. The FROC paradigm is, nowadays, being increasingly used in the assessment of medical imaging systems, particularly in the evaluation and comparison of CAD algorithms [17, 87].

FROC analysis is similar to ROC analysis, except that the false positive rate on the X-axis is replaced by the number of false positives per image. Thus, FROC seeks location information from the observer (the algorithm), rewarding it when the reported disease is marked in the appropriate location and penalizing it when it is not. Note that this task is more relevant to the clinical practice of radiology, where it is not only important to identify disease, but also to offer further guidance regarding other characteristics (such as location) of the disease.

Before FROC data can be analyzed, a definition of a detected region is needed. Although there are different opinions in the literature [48, 87, 144], in our work we use a typical approach which expects a 50% overlap between the annotated and detected regions to indicate a true positive.

Bibliography

[1] American Cancer Society. Breast cancer: facts and figures. 2003-04. *ACS*, 2003.

[2] American College of Radiology. *Illustrated Breast Imaging Reporting and Data System BIRADS*. American College of Radiology, 3rd edition, 1998.

[3] I. Anttinen, M. Pamilo, M. Soiva, and M. Roiha. Double reading of mammography screening films: one radiologist or two? *Clinical Radiology*, 48:414–421, 1993.

[4] Y. Attikiouzel and R. Chandrasekhar. DSP in mammography. In *Proc. International Conference on Digital Signal Processing*, pages 29–34, 2002.

[5] D. Avrin, R. Morin, D. Piraino, A. Rowberg, N. Detorie, M. Zuley, J. A. Seibert, and E. D. Pisano. Storage, transmission, and retrieval of digital mammography, including recommendations on image compression. *Journal of the American College of Radiology*, 3:609–614, 2006.

[6] L. W. Basset and R. H. Gold. *Breast Cancer Detection: Mammograms and Other Methods in Breast Imaging*. Grune & Stratton, New York, 1987.

[7] A. H. Baydush, D. M. Catarious, C. K. Abbey, and C. E. Floyd. Computer aided detection of masses in mammography using subregion Hotelling observers. *Medical Physics*, 30(7):1781–1787, 2003.

[8] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[9] R. Bellotti, F. De Carlo, S. Tangaro, G. Gargano, G. Maggipinto, M. Castellano, R. Massafra, D. Cascio, F. Faucio, R. Magro, G. Raso, A. Lauria, G. Forni, S. Bagnasco, P. Cerello, E. Zanon, S. C. Cheran, E. López Torres, U. Bottigli, G. L. Masala, P. Oliva, A. Retico, E. Fantacci, R. Cataldo, I. De Mitri, and G. De Nunzio. A completely automated CAD system for mass detection in large mammographic database. *Medical Physics*, 33(8):3066–3075, 2006.

[10] J. Benois and D. Barba. Image segmentation by region-contour cooperation for image coding. In *Proc. IAPR International Conference on Pattern Recognition*, volume C, pages 331–334, 1992.

[11] E. A. Berns, R. E. Hendrick, M. Solari, L. Barke, D. Reddy, J. Wolfman, L. Segal, P. De Leon, S. Benjamin, and L. Willis. Digital and screen-film mammography: Comparison of image acquisition and interpretation times. *American Journal of Roentgenology*, 187(1):38–41, 2006.

[12] J. C. Bezdek. *Pattern Recognition With Fuzzy Objective Function Algorithms*. Plenum Press, New York, 1981.

[13] R. E. Bird, T. W. Wallace, and B. C. Yankaskas. Analysis of cancers missed at screening mammography. *Radiology*, 184:613–617, 1992.

[14] R. L. Birdwell, D. M. Ikeda, K. D. O'Shaughnessy, and E. A. Sickles. Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection. *Radiology*, 219:192–202, 2001.

[15] L. Blot and R. Zwiggelaar. Background texture extraction for the classification of mammographic parenchymal patterns. In *Proc. Medical Image Understanding and Analysis*, pages 145–148, 2001.

[16] L. Blot and R. Zwiggelaar. A volumetric approach to risk assessment in mammography: a feasibility study. *Physics in Medicine and Biology*, 50:695–708, 2005.

[17] H. Bornefalk. Estimation and comparison of CAD system performance in clinical settings. *Academic Radiology*, 12:687–694, 2005.

[18] K. Bovis and S. Singh. Classification of mammographic breast density using a combined classifier paradigm. In *Proc. Medical Image Understanding and Analysis*, pages 177–180, 2002.

[19] N. F. Boyd, J. W. Byng, R. A. Jong, E. K. Fishell, L. E. Little, A. B. Miller, G. A. Lockwood, D. L. Tritchler, and M. J. Yaffe. Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian national breast screening study. *J. Natl Cancer Inst.*, 87:670–675, 1995.

[20] A. P. Bradley. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30:1145–1159, 1997.

[21] D. Brzakovic, X. M. Luo, and P. Brzakovic. An approach to automated detection of tumors in mammograms. *IEEE Transactions on Medical Imaging*, 9(3):233–241, 1990.

[22] S. Buseman, J. Mouchawar, N. Calonge, and T. Byers. Mammography screening matters for young women with breast carcinoma. *Cancer*, 97(2):352–358, 2003.

[23] J. W. Byng, N. F. Boyd, E. Fishell, R. A. Jong, and M. J. Yaffe. Automated analysis of mammographic densities. *Physics in Medicine and Biology*, 41:909–923, 1996.

[24] C. B. Caldwell, S. J. Stapleton, D. W. Holdsworth, R. A. Jong, W. J. Weiser, G. Cooke, and M. J. Yaffe. Characterization of mammographic parenchymal pattern by fractal dimension. *Physics in Medicine and Biology*, 35:235–247, 1990.

[25] R. Campanini, D. Dongiovanni, E. Iampieri, N. Lanconelli, M. Masotti, G. Palermo, A. Riccardi, and M. Roffilli. A novel featureless approach to mass detection in digital mammograms based on support vector machines. *Physics in Medicine and Biology*, 49:961–975, 2004.

[26] D. M. Catarious, A. H. Baydush, and C. E. Floyd. Incorporation of an iterative, linear segmentation routine into a mammographic mass CAD system. *Medical Physics*, 31(6):1512–1520, 2004.

[27] D. P. Chakraborty, H. J. Yoona, and C. Mello-Thoms. Localization accuracy of radiologists in free-response studies: Inferring perceptual FROC curves from mark-rating data. *Academic Radiology*, 14:4–18, 2007.

[28] R. Chandrasekhar and Y. Attikiouzel. Gross segmentation of mammograms using a polynomial model. In *Proc. International Conference IEEE Engineering in Medicine and Biology Society*, volume 3, pages 1056–1058, 1996.

[29] Y. H. Chang, L. A. Hardesty, C. M. Hakim, T. S. Chang, B. Zheng, W. F. Good, and D. Gur. Knowledge-based computer-aided detection of masses on digitized mammograms: A preliminary assessment. *Medical Physics*, 28(4):455–461, 2001.

[30] C. H. Chen and G. G. Lee. On digital mammogram segmentation and microcalcification detection using multiresolution wavelet analysis. *Graphical Models and Image Processing*, 59(5):349–364, 1997.

[31] P. C. Chen and T. Pavlidis. Segmentation by texture using a co-occurrence matrix and a split-and-merge algorithm. *Computer Graphics and Image Processing*, 10:172–182, 1979.

[32] H. D. Cheng and M. Cui. Mass lesion detection with a fuzzy neural network. *Pattern Recognition*, 37:1189–1200, 2004.

[33] I. Christoyianni, E. Constantinou, and E. Dermatas. Automatic detection of abnormal tissue in bilateral mammograms using neural networks. In *Methods and Applications of Artificial Intelligence, Helenic Conference on AI*, pages 267–275, 2004.

[34] I. Christoyianni, A. Koutras, E. Dermatas, and G. Kokkinakis. Computer aided of breast cancer in digitized mammograms. *Computerized Medical Imaging and Graphics*, 26:309–319, 2002.

[35] S. Ciatto, N. Houssami, A. Apruzzese, E. Bassetti, B. Brancato, F. Carozzi, S. Catarzi, M. P. Lamberini, G. Marcelli, R. Pellizzoni, B. Pesce, G. Risso, F. Russo, and A. Scorsolini. Categorizing breast mammographic density: intra- and interobserver reproducibility of BI-RADS density categories. *Breast*, 14(4):269–275, 2005.

[36] J. Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:27–46, 1960.

[37] A. S. Constantinidis, M. C. Fairhust, F. Deravi, M. Hanson, C. P. Wells, and C. Chapman-Jones. Evaluating classification strategies for detection of circumscribed masses in digital mammograms. In *Proc. International Conference on Image Processing and Its Application*, pages 435–439, 1999.

[38] A. S. Constantinidis, M. C. Fairhust, and A. F. R. Rahman. Detection of circumscribed masses in digital mammograms using behaviour-knowledge space method. *Electronic Letters*, 36(4):302–303, 2000.

[39] T. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[40] E. R. Davies. *Machine Vision*. Academic Press, London, UK, 2nd edition, 1997.

[41] H. J. De Koning, J. Fracheboud, R. Boer, A. L. Verbeek, H. J. Collette, J. H. C. L. Hendriks, B. M. van Ineveld, A. E. de Bruyn, and P. J. van der Maas. Nation-wide breast cancer screening in the Netherlands: support for breast cancer mortality reduction. National evaluation team for breast cancer screening. *International Journal of Cancer*, 60(6):777–780, 1995.

[42] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum-likelihood from incomplete data via EM algorithm. *Journal of the Royal Statistical Society: Series B*, pages 1–38, 1977.

[43] J. Deschênes and D. Ziou. Detection of line junctions and line terminations using curvilinear features. *Pattern Recognition Letters*, 21(6–7):45–54, 2000.

[44] Digital Imaging and COmmunications in Medicine. http://medical.nema.org. *accessed 01/01/07*.

[45] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley & Sons, New York, 2 edition, 2001.

[46] Editorial. Glossary of terms. *Machine Learning*, 30(2–3):271–274, 1998.

[47] Editorial. Looking back on the millennium in medicine. *New England Journal of Medicine*, 342(1):42–49, 2000.

[48] D. C. Edwards, M. A. Kupinski, C. E. Metz, and R. M. Nishikawa. Maximum likelihood fitting of FROC curves under an initial-detection-and-candidate-analysis model. *Medical Physics*, 29:2861–2870, 2002.

[49] J. Esteve, A. Kricker, J. Ferlay, and D. Parkin. Facts and figures of cancer in the European Community. Technical report, International Agency for Research on Cancer, Lyon, France, 1993.

[50] Eurostat. Health statistics atlas on mortality in the European Union. *Official Journal of the European Union*, 2002.

[51] R. J. Ferrari, R. M. Rangayyan, and J. E. L. Desautels. Segmentation of mammograms: Identification of the skin boundary and the pectoral muscle. In *Proc. International Workshop on Digital Mammography*, pages 573–579, 2000.

[52] R. J. Ferrari, R. M. Rangayyan, J. E. L. Desautels, R. A. Borges, and A. F. Frère. Automatic identification of the pectoral muscle in mammograms. *IEEE Transactions on Medical Imaging*, 23(2):232–245, 2004.

[53] J. L. Fleiss. *Statistical Methods for Rates and Proportions*. John Wiley & Sons, New York, 1981.

[54] T. W. Freer and M. J. Ulissey. Screening mammography with computer-aided detection: Prospective study of 12860 patients in a community breast center. *Radiology*, 220:781–786, 2001.

[55] J. Freixenet, X. Muñoz, J. Martí, and X. Lladó. Colour texture segmentation by region-boundary cooperation. In *Proc. European Conference on Computer Vision*, volume II, pages 250–261, 2004.

[56] J. Freixenet, X. Muñoz, D. Raba, J. Martí, and X. Cufí. Yet another survey on image segmentation: Region and boundary information integration. In *Proc. European Conference on Computer Vision*, volume III, pages 408–422, 2002.

[57] J. Freixenet, P. Planiol, J. Martí, R. García, J. Batlle, and R. Bassaganyas. Cluster-shape characterization of microcalcifications. In *IEE Proceedings of Medical Applications of Signal Processing*, pages 5/1–5/5, 2002.

[58] K. S. Fu and J. K. Mui. A survey on image segmentation. *Pattern Recognition*, 13:3–16, 1981.

[59] GE Healthcare. http://www.gehealthcare.com. *accessed 01/01/07.*

[60] F. Georgsson. Differential analysis of bilateral mammograms. *International Journal of Pattern Recognition and Artificial Intelligence*, 17(7):1207–1226, 2003.

[61] M. L. Giger, F. F. Yin, K. Doi, Y. Wu, C. J. Vyborny, R. A. Schmidt, and Z. Huo. Computerized detection and characterization of mass lesions in digital mammography. In *Proc. IEEE International Conference on Systems, Man and Cybernetics*, pages 1370–1372, 1992.

[62] F. Gil, I. Méndez, A. Sirgo, G. Llort, I. Blanco, and H. Cortés-Funes. Perception of breast cancer risk and surveillance behaviours of women with family history of breast cancer: a brief report on a spanish cohort. *Psycho-Oncology*, 12:821–827, 2003.

[63] Y. C. Gong, M. Brady, and S. Petroudi. Texture based mammogram classification and segmentation. *Lecture Notes in Computer Science*, 4046:616–625, 2006.

[64] M. Goto, A. Morikawa, H. Fujita, T. Hara, and T. Endo. Detection of spicules on mammograms based on a multistage pendulum filter. In *Proc. International Workshop on Digital Mammography*, pages 135–138, 1998.

[65] D. Guliato, R. M. Rangayyan, W. A. Carnielli, J. A. Zuffo, and J. E. L. Desautels. Segmentation of breast tumors in mammograms by fuzzy region growing. In *Proc. International Conference IEEE Engineering in Medicine and Biology Society*, volume 20, pages 1002–1005, 1998.

[66] F. M. Hall. Mammographic density categories. *American Journal of Roentgenology*, 178:242–242, 2002.

[67] F. M. Hall, J. M. Storella, D. Z. Siverstond, and G. Wyshak. Nonpalpable breast lesions: Recommendations for biopsy based on suspicion of carcinoma at mammography. *Radiology*, 167:353–358, 1988.

[68] R. M. Haralick, K. S. Shanmugan, and I. Dunstein. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3(6):610–621, 1973.

[69] A. E. Hassanien, J. M. Ali, and H. Nobuhara. Detection of spiculated masses in mammograms based on fuzzy image processing. In *Lecture Notes in Computer Science*, volume 3070, pages 1002–1007, 2004.

[70] Y. Hatanaka, T. Hara, H. Fujita, S. Kasai, T. Endo, and T. Iwase. Development of an automated method for detecting mammographic masses with a partial loss of region. *IEEE Transactions on Medical Imaging*, 20(12):1209–1214, 2001.

[71] M. Heath, K. Bowyer, D. Kopans, R. Moore, and P. J. Kegelmeyer. The Digital Database for Screening Mammography. In *Proc. International Workshop on Digital Mammography*, pages 212–218, 2000.

[72] M. D. Heath and K. W. Bowyer. Mass detection by relative image intensity. In *Proc. International Workshop on Digital Mammography*, pages 219–225, 2000.

[73] J. J. Heine, M. Kallergi, S. M. Chetelat, and L. P. Clarke. Multiresolution wavelet approach for separating the breast region from the background in high resolution digital mammography. In *Proc. International Workshop on Digital Mammography*, pages 295–298, 1998.

[74] J. J. Heine and R. P. Veltzuihen. A statistical methodology for mammographic density detection. *Medical Physics*, 27(12):2644–2651, 2000.

[75] W. R. Hendee, C. Beam, and E. Hendrick. Proposition: all mammograms should be double-read. *Medical Physics*, 26:115–118, 1999.

[76] S. H. Heywang-Köbrunner, D. D. Dershaw, and I. Schreer. *Diagnostic Breast Imaging. Mammography, sonography, magnetic resonance imaging, and interventional procedures.* Thieme, Stuttgart, Germany, 2001.

[77] R. Highnam and M. Brady. *Mammographic Image Analysis.* Kluwer, Dordrecht, 1999.

[78] R. Highnam, Y. Kita, M. Brady, B. Shepstone, and R. English. Determining correspondence between views. In *Proc. International Workshop on Digital Mammography*, pages 111–118, 1998.

[79] R. Highnam, X. Pan, R. Warren, M. Jeffreys, G. Davey Smith, and M. Brady. Breast composition measurements using retrospective standard mammogram form (SMF). *Physics in Medicine and Biology*, 51:2695–2713, 2006.

[80] W. T. Ho and P. W. T. Lam. Clinical performance of computer-assisted detection (CAD) system in detecting carcinoma in breasts of different densities. *Clinical Radiology*, 58:133–136, 2003.

[81] Z. Huo, M. L. Giger, C. J. Vyborny, U. Bick, P. Lu, D. E. Wolverton, and R. A. Schmidt. Analysis of spiculation in the computerized classification of mammographic masses. *Medical Physics*, 22(10):1569–1579, 1995.

[82] iCAD Second Look. http://www.icadmed.com. *accessed 01/01/07.*

[83] A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: A review. *ACM: Computing Surveys*, 31(3):264–323, 1999.

[84] A. K. Jain, Y. Zhong, and M. P. Dubuisson-Jolly. Deformable template models: A review. *Signal Processing*, 71:109–129, 1998.

[85] A. K. Jain, Y. Zhong, and S. Lakshmanan. Object matching using deformable templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(3):267–278, 1996.

[86] H. Jiang, W. Tiu, S. Yamamoto, and S. Iisaku. A method for automatic detection of spicules in mammograms. *Journal of Computed Aided Diagnostics of Medical Images*, 2(4):1–8, 1998.

[87] M. Kallergi, G. M. Carney, and J. Gaviria. Evaluating the performance of detection algorithms in digital mammography. *Medical Physics*, 26:267–275, 1999.

[88] N. Karssemeijer. Automated classification of parenchymal patterns in mammograms. *Physics in Medicine and Biology*, 43:365–378, 1998.

[89] N. Karssemeijer and G. M. te Brake. Detection of stellate distortions in mammograms. *IEEE Transactions on Medical Imaging*, 15(5):611–619, 1996.

[90] N. Karssemeijer and G. M. te Brake. Combining single view features and asymmetry for detection of mass lesions. In *Proc. International Workshop on Digital Mammography*, pages 95–102, 1998.

[91] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. In *Proc. IEEE International Conference on Computer Vision*, pages 259–268, 1987.

[92] L. A. L. Khoo, P. Taylor, and R. M. Given-Wilson. Computer-aided detection in the United Kingdom national breast screening programme: prospective study. *Radiology*, 237(2):444–449, 2005.

[93] J. Kittler. Feature selection and extraction. In T. Y. Young and K. S. Fu, editors, *Handbook of Pattern Recognition and Image Processing*, pages 59–83. Academic Press, 1986.

[94] H. Kobatake and M. Murakami. Adaptive filter to detect rounded convex regions: Iris filter. In *Proc. IAPR International Conference on Pattern Recognition*, volume 2, pages 340–345, 1996.

[95] H. Kobatake, M. Murakami, H. Takeo, and S. Nawano. Computerized detection of malignant tumors on digital mammograms. *IEEE Transactions on Medical Imaging*, 18(5):369–378, 1999.

[96] H. Kobatake, H. Takeo, and S. Nawano. Tumor detection system for full-digital mammography. In *Proc. International Workshop on Digital Mammography*, pages 87–94, 1998.

[97] H. Kobatake and Y. Yoshinaga. Detection of spicules on mammogram based on skeleton analysis. *IEEE Transactions on Medical Imaging*, 15(3):235–245, 1996.

[98] H. Kobatake, Y. Yoshinaga, and M. Murakami. Automatic detection of malignant tumors on mammogram. In *Proc. IEEE International Conference on Image Processing*, volume 1, pages 407–410, 1994.

[99] Kodak. http://www.kodak.com. *accessed 01/01/07.*

[100] S. L. Kok-Wiles, M. Brady, and R. Highman. Comparing mammogram pairs for the detection of lesions. In *Proc. International Workshop on Digital Mammography*, pages 103–110, 1998.

[101] D. Kopans. *Breast Imaging.* Lippincott-Raven, Philadelphia, 1998.

[102] M. A. Kupinski and M. L. Giger. Automated seeded lesion segmentation on digital mammograms. *IEEE Transactions on Medical Imaging*, 17(4):510–517, 1998.

[103] S. M. Kwok, R. Chandrasekhar, and Y. Attikiouzel. Adaptation of the Daugman-Downing texture demodulation to highlight circumscribed mass lesions on mammograms. In *Proc. International Conference on Digital Signal Processing*, pages 449–452, 2002.

[104] S. M. Kwok, R. Chandrasekhar, Y. Attikiouzel, and M. T. Rickard. Automatic pectoral muscle segmentation on mediolateral oblique view mammograms. *IEEE Transactions on Medical Imaging*, 23(9):1129–1140, 2004.

[105] S. M. Lai, X. Li, and W. F. Bischof. On techniques for detecting circumscribed masses in mammograms. *IEEE Transactions on Medical Imaging*, 8(4):377–386, 1989.

[106] J. R. Landis and G. G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(3):159–174, 1977.

[107] T. Lau and W. F. Bischof. Automated detection of breast tumors using the asymmetry approach. *Computers and Biomedical Research*, 24(3):273–295, 1991.

[108] K. Laws. Rapid texture identification. *SPIE, Image Processing for Missile Guidance*, 238:376–380, 1980.

[109] C. D. Lehman, E. White, S. Peacock, M. J. Drucker, and N. Urban. Effect of age and breast density on screening mammograms with false-positive findings. *American Journal of Roentgenology*, 173:1651–1655, 1999.

[110] H. Li, Y. Wang, K. J. R. Liu, S. C. B. Lo, and M. T. Freedman. Computerized radiographic mass detection - Part I: Lesion site selection by morphological enhancement and contextual segmentation. *IEEE Transactions on Medical Imaging*, 20(4):289–301, 2001.

[111] H. D. Li, M. Kallergi, L. P. Clarke, V. K. Jain, and R. A. Clark. Markov random field for tumor detection in digital mammography. *IEEE Transactions on Medical Imaging*, 14(3):565–576, 1995.

[112] L. Li, R. A. Clark, and J. A. Thomas. Computer-aided diagnosis of masses with full-field digital mammography. *Academic Radiology*, 9(1):4–12, 2002.

[113] T. Lindeberg. Edge detection and ridge detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):117–156, 1998.

[114] J. B. MacQueen. Some methods of classification and analysis of multivariate observations. In *Proc. of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297, 1967.

[115] I. E. Magnin, F. Cluzeau, C. L. Odet, and A. Bremond. Mammographic texture analysis: an evaluation of risk for developing breast cancer. *Optical Engineering*, 25:780–784, 1986.

[116] A. Malich, D. R. Fischer, and J. Böttcher. CAD for mammography: the technique, results, current role and further developments. *Epidemiologic Reviews*, 16:1449–1460, 2006.

[117] J. Martí, J. Español, E. Golobardes, J. Freixenet, R. García, and M. Salamó. Classification of microcalcifications in digital mammograms using case-based reasoning. In *Proc. International Workshop on Digital Mammography*, pages 285–294, 2000.

[118] J. Martí, J. Freixenet, M. Peracaula, A. Oliver, D. Raba, J. Espunya, J. Pont, and R. Martí. Automatic segmentation of microcalcifications based on the fusion of different algorithms over CC and MLO views. In *Proc. International Workshop on Digital Mammography*, 2004.

[119] J. Martí, J. Freixenet, D. Raba, A. Bosch, J. Pont, J. Español, R. Bassaganyas, E. Golobardes, and X. Canaleta. HRIMAC: Una herramienta de recuperación de imágenes mamográficas por análisis de contenido para el asesoramiento en el diagnóstico del cáncer de mama. In *Inforsalud: VI Congreso Nacional de Informática de la Salud*, pages 125–132, 2003.

[120] R. Martí, D. Raba, A. Oliver, and R. Zwiggelaar. Mammographic registration: proposal and evaluation of a new approach. In *Lecture Notes in Computer Science*, volume 4046, pages 213–220, 2006.

[121] R. Martí, R. Zwiggelaar, C. Rubin, and E. R. E. Denton. Automatic point correspondence and registration based on linear structures. *International Journal of Pattern Recognition and Artificial Intelligence*, 16(3):331–340, 2002.

[122] R. Martí, R. Zwiggelaar, and C. M. E. Rubin. Tracking mammographic structures over time. In *Proc. British Machine Vision Conference*, pages 143–152, 2001.

[123] K. E. Martin, M. A. Helvie, C. Zhou, M. A. Roubidoux, J. E. Bailey, C. Paramagul, C. E. Blane, K. A. Klein, S. S. Sonnad, and H. P. Chan. Mammographic density measured with quantitative computer-aided method: comparison with radiologists' estimates and BI-RADS categories. *Radiology*, 240:656–665, 2006.

[124] T. Matsubara, H. Fujita, T. Hara, S. Kasai, O. Otsuka, Y. Hatanaka, and T. Endo. Development of a new algorithm for detection of mammographic masses. In *Proc. International Workshop on Digital Mammography*, pages 139–142, 1998.

[125] T. Matsubara, H. Fujita, S. Kasai, M. Goto, Y. Tani, T. Hara, and T. Endo. Development of new schemes for detection and analysis of mammographic masses. In *Proc. International Conference on Information and Intelligent Systems*, pages 63–66, 1997.

[126] P. McKenzie and M. Alder. Initializing the EM algorithm for use in Gaussian mixture modeling. In *Proc. on Pattern Recognition in Practice IV*, pages 91–105, 1994.

[127] G. J. McLachlan. *Discriminant Analysis and Statistical Pattern Recognition*. John Wiley and Sons, New York, 1992.

[128] A. J. Méndez, M. Souto, P. G. Tahoces, and J. J. Vidal. Computer aided diagnosis for breast masses detection on a telemammography system. *Computerized Medical Imaging and Graphics*, 27:497–502, 2003.

[129] A. J. Méndez, P. G. Tahoces, M. J. Lado, M. Souto, and J. J. Vidal. Computerized-aided diagnosis: Automatic detection of malignant masses in digitized mammograms. *Medical Physics*, 25(6):957–964, 1998.

[130] C. E. Metz. Evaluation of digital mammography by ROC analysis. In *Proc. International Workshop on Digital Mammography*, pages 61–68, 1996.

[131] P. Miller and S. M. Astley. Classification of breast tissue by texture analysis. *Image and Vision Computing*, 10:277–282, 1992.

[132] N. R. Mudigonda, R. M. Rangayyan, and J. E. L. Desautels. Detection of breast masses in mammograms by density slicing and texture flow-field analysis. *IEEE Transactions on Medical Imaging*, 20(12):1215–1214, 2001.

[133] I. Muhimmah, A. Oliver, E. R. E. Denton, J. Pont, E. Pérez, and R. Zwiggelaar. Comparison between Wolfe, Boyd, BI-RADS and Tabár based mammographic risk assessment. In *Lecture Notes in Computer Science*, volume 4046, pages 407–415, 2006.

[134] X. Muñoz, J. Freixenet, X. Cufí, and J. Martí. Strategies for image segmentation combining region and boundary information. *Pattern Recognition Letters*, 24(1-3):375–392, 2003.

[135] National Electrical Manufacturers Association. *Digital Imaging and Communications in Medicine (DICOM).* National Electrical Manufacturers Association, 3.1 edition, 2006.

[136] R. M. Nishikawa and M. Kallergi. Computer-aided detection, in its present form, is not an effective aid for screening mammography. *Medical Physics*, 33:811–814, 2006.

[137] Oklahoma Heart Hospital. http://www.okheart.com. *accessed 01/01/07.*

[138] A. Oliver, J. Freixenet, A. Bosch, D. Raba, and R. Zwiggelaar. Automatic classification of breast tissue. In *Lecture Notes in Computer Science*, volume 3523, pages 431–438, 2005.

[139] A. Oliver, J. Freixenet, M. Peracaula, and J. Martí. Mass segmentation using a pattern matching approach with a mutual information based metric. In *Proc. of the Congrés Català d'Intel·ligència Artificial*, pages 115–122, 2005.

[140] A. Oliver, J. Freixenet, and R. Zwiggelaar. Automatic classification of breast density. In *Proc. IEEE International Conference on Image Processing*, volume 2, pages 1258–1261, 2005.

[141] A. Oliver, J. Martí, R. Martí, A. Bosch, and J. Freixenet. A new approach to the classification of mammographic masses and normal breast tissue. In *Proc.*

IAPR International Conference on Pattern Recognition, volume 4, pages 707–710, 2006.

[142] X. Pan, M. Brady, R. Highnam, and J. Declerck. The use of multi-scale monogenic signal on structure orientation identification and segmentation. *Lecture Notes in Computer Science*, 4046:601–608, 2006.

[143] L. F. Parr, A. L. Anderson, B. R. Glennon, and P. Fetherston. Quality control issues on high resolution diagnostic monitors. *Journal of Digital Imaging*, 14:22–26, 2001.

[144] M. Penedo, M. Souto, P. G. Tahoces, J. M. Carreira, J. Villalón, G. Porto, C. Seoane, J. J. Vidal, K. S. Berbaum, D. P. Chakraborty, and L. L. Fajardo. Free-response receiver operating characteristic evaluation of lossy JPEG2000 and object-based set partitioning in hierarchical trees compression of digitized mammograms. *Radiology*, 237:450–457, 2005.

[145] N. Petrick, H. P. Chan, B. Sahiner, and M. A. Helvie. Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms. *Medical Physics*, 26(8):1642–1654, 1999.

[146] N. Petrick, H. P. Chan, B. Sahiner, and D. Wei. An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection. *IEEE Transactions on Medical Imaging*, 15(1):59–67, 1996.

[147] N. Petrick, H. P. Chan, D. Wei, B. Sahiner, M. A. Helvie, and D. D. Adler. Automated detection of breast masses on mammograms using adaptive contrast enhancement and texture classification. *Medical Physics*, 23(10):1685–1696, 1996.

[148] S. Petroudi, T. Kadir, and M. Brady. Automatic classification of mammographic parenchymal patterns: A statistical approach. In *Proc. International Conference IEEE Engineering in Medicine and Biology Society*, volume 1, pages 798–801, 2003.

[149] E. D. Pisano, E. B. Cole, E. O. Kistner, K. E. Muller, B. M. Hemminger, M. L. Brown, R. E. Johnston, C. M. Kuzmiak, M. P. Braeuning, R. I. Freimanis,

M. S. Soo, J. A. Baker, and R. Walsh. Interpretation of digital mammograms: Comparison of speed and accuracy of soft-copy versus printed-film display. *Radiology*, 223(2):483–488, 2002.

[150] W. E. Polakowski, D. A. Cournoyer, S. K. Rogers, M. P. DeSimio, D. W. Ruck, J. W. Hoffmeister, and R. A. Raines. Computer-aided breast cancer detection and diagnosis of masses using Difference of Gaussian and derivative-based feature saliency. *IEEE Transactions on Medical Imaging*, 16(6):811–819, 1997.

[151] H. Qi and W. E. Snyder. Lesion detection and characterization in digital mammography by Bézier histograms. In *Proc. International Conference IEEE Engineering in Medicine and Biology Society*, volume 2, pages 1021–1024, 1998.

[152] W. Qian, L. Li, L. Clarke, F. Mao, R. A. Clark, and J. Thomas. A computer assited diagnostic system for mass detection. In *Proc. International Workshop on Digital Mammography*, pages 79–86, 1998.

[153] W. Qian, L. Li, and L. P. Clarke. Image feature extraction for mass detection in digital mammography: Influence of wavelet analysis. *Medical Physics*, 26(3):402–408, 1999.

[154] W. Qian, L. Li, L. P. Clarke, M. Fei, and R. A. Clark. Adaptive CAD modules for mass detection in digital mammography. In *Proc. International Conference IEEE Engineering in Medicine and Biology Society*, volume 2, pages 1013–1016, 1998.

[155] W. Qian, X. Sun, D. Song, and R. A. Clarke. Digital mammography - wavelet transform and kalman-filtering neural network in mass segmentation and detection. *Academic Radiology*, 8(11):1074–1082, 2001.

[156] J. R. Quinlan. Induction of decision trees. *Machine Learning*, 1(1):81–106, 1986.

[157] J. R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, New York, 1993.

[158] J. R. Quinlan. Bagging, boosting, and C4.5. In *Nat. Conf. on Art. Intell.*, pages 725–730, 1996.

[159] R2 ImageChecker. http://www.r2tech.com. *accessed 01/01/07.*

[160] D. Raba, J. Martí, R. Martí, and M. Peracaula. Breast mammography asymmetry estimation based on fractal and texture analysis. In *Proc. of the Computed Aided Radiology and Surgery*, page 1398, 2005.

[161] D. Raba, A. Oliver, J. Martí, M. Peracaula, and J. Espunya. Breast segmentation with pectoral muscle suppression on digital mammograms. In *Lecture Notes in Computer Science*, volume 3523, pages 471–478, 2005.

[162] R. M. Rangayyan, E. N. M. El-Faramawy, J. E. L. Desautels, and O. A. Alim. Measures of acutance and shape for classification of breast tumors. *IEEE Transactions on Medical Imaging*, 16(6):799–810, 1997.

[163] F. J. P. Richard and L. D. Cohen. A new image registration technique with free boundary constraints: application to mammography. *Computer Vision and Image Understanding*, 89:166–196, 2003.

[164] L. G. Roberts. Machine perception of three-dimensional solids. In J. Tippet, D. Berkowitz, L. Clapp, C. Koester, and A. Vanderburgh, editors, *Optical and Electro-Optical Information Processing*, pages 159–197. MIT Press, Cambridge, Massachussets, 1965.

[165] A. A. J. Roelofs, S. van Woudenberg, J. D. M. Otten, J. H. C. L. Hendriks, A. Bödicker, C. J. G. Evertsz, and N. Karssemeijer. Effect of soft-copy display supported by CAD on mammography screening performance. *Epidemiologic Reviews*, 16:45–52, 2006.

[166] P. K. Saha, J. K. Udupa, E. F. Conant, P. Chakraborty, and D. Sullivan. Breast tissue density quantification via digitized mammograms. *IEEE Transactions on Medical Imaging*, 20(8):792–803, 2001.

[167] B. Sahiner, H. P. Chan, N. Petrick, M. A. Helvie, and M. M. Goodsit. Computerized characterization of masses on mammograms: the rubber band straightening transform and texture analysis. *Medical Physics*, 25(4):516–526, 1998.

[168] B. Sahiner, H. P. Chan, N. Petrick, M. A. Helvie, and L. M. Hadjiiski. Improvement of mammographic mass characterization using spiculation measures and morphological features. *Medical Physics*, 28(7):1455–1465, 2001.

[169] B. Sahiner, H. P. Chan, D. Wei, N. Petrick, M. A. Helvie, D. D. Adler, and M. M. Goodsit. Image feature selection by a genetic algorithm: Application to classification of mass and normal breast tissue. *Medical Physics*, 23:1671–1684, 1996.

[170] B. Sahiner, N. Petrick, H. P. Chan, L. M. Hadjiiski, C. Paramagul, M. A. Helvie, and M. N. Gurcan. Computer-aided characterization of mammographic masses: Accuracy of mass segmentation and its effects on characterization. *IEEE Transactions on Medical Imaging*, 20(12):1275–1284, 2001.

[171] M. Sallam and K. Bowyer. Registration and difference analysis of corresponding mammogram images. *Medical Image Analysis*, 3(2):103–118, 2001.

[172] S. Sanjay-Gopal, H. P. Chan, T. Wilson, M. Helvie, N. Petrick, and B. Sahiner. A regional registration technique for automated interval change analysis of breast lesions on mammograms. *Medical Physics*, 26(12):2669–2679, 1999.

[173] S. E. Selvan, C. C. Xavier, N. Karssemeijer, J. Sequeira, R. A. Cherian, and B. Y. Dhala. Parameter estimation in stochastic mammogram model by heuristic optimization techniques. *IEEE Transactions on Information Technology in Biomedicine*, 10:685–695, 2006.

[174] L. Shen, R. M. Rangayyan, and J. E. L. Desautels. Application of shape analysis to mammographic calcifications. *IEEE Transactions on Medical Imaging*, 13:263–274, 1994.

[175] E. A. Sickles. Breast cancer screening outcomes in women ages 40-49: clinical experience with service screening using modern mammography. *Journal of the National Cancer Institute: Monographs*, 22:99–104, 1997.

[176] E. A. Sickles, W. N. Weber, and H. B. Galvin. Baseline screening mammography: one vs two views per breast. *American Journal of Roentgenology*, 147:1149–1153, 1986.

[177] E. Siegel, E. Krupinski, E. Samei, M. Flynn, K. Andriole, B. Erickson, J. Thomas, A. Badano, J. A. Seibert, and E. D. Pisano. Digital mammography image quality: Image display. *Journal of the American College of Radiology*, 3:615–627, 2006.

[178] R. Sivaramakrishna, N. A. Obuchowski, W. A. Chilcote, and K. A. Powell. Automatic segmentation of mammographic density. *Academic Radiology*, 8(3):250–256, 2001.

[179] A. P. Smith. Fundamentals of digital mammograhpy: physiscs, technology and practical considerations. *Radiology Management*, 25(5):18–31, 2003.

[180] E. A. Stamatakis, I. W. Ricketts, A. Y. Cairns, C. Walker, and P. E. Preece. Detecting abnormalities on mammograms by bilateral comparison. In *Proc. IEE Colloquium on Digital Mammography*, pages 12/1–12/4, 1996.

[181] T. Stathaki and A. G. Constantinides. Neural networks and higher order spectra for breast cancer detection. In *Proc. IEEE Workshop on Neural Networks for Signal Processing*, pages 473–481, 1994.

[182] J. Suckling, D. R. Dance, E. Moskovic, D. J. Lewis, and S. G. Blacker. Segmentation of mammograms using multiple linked self-organizing neural networks. *Medical Physics*, 22:145–152, 1995.

[183] J. Suckling, J. Parker, D. R. Dance, S. M. Astley, I. Hutt, C. R. M. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. L. Kok, P. Taylor, D. Betal, and J. Savage. The Mammographic Image Analysis Society digital mammogram database. In *Proc. International Workshop on Digital Mammography*, pages 211–221, 1994.

[184] X. Sun, D. Qian, W. Song, and A. C. Robert. Ipsilateral multi-view CAD system for mass detection in digital mammography. In *IEEE Workshop on Mathematical Methods in Biomedicine and Image Analysis*, pages 19–26, 2001.

[185] X. Sun, W. Qian, and D. Song. Ipsilateral-mammogram computer-aided detection of breast cancer. *Computerized Medical Imaging and Graphics*, 28:151–158, 2004.

[186] L. Tabar, C. J. G. Fagerbert, A. Gad, L. Baldetorp, L. H. Holmberg, O. Grontoft, U. Ljungquist, B. Lundstrom, J. C. Manson, G. Erklung, N. E. Day, and F. Pettersson. Reduction in mortality from breast cancer after mass screening with mammography. randomised trial from the breast cancer screening working group of the swedish national board of health and welfare. *Lancet*, 1:829–832, 1985.

[187] P. G. Tahoces, J. Correa, M. Soutu, L. Gómez, and J. J. Vidal. Computer-assisted diagnosis: the classification of mammographic breast parenchymal patterns. *Physics in Medicine and Biology*, 40:103–117, 1995.

[188] P. Taylor, J. Champness, R. Given-Wilson, K. Johnston, and H. Potts. Impact of computer-aided detection prompts on the sensitivity and specificity of screening mammography. *Health Technology Assessment*, 9(6):1–58, 2005.

[189] P. Taylor, S. Hajnal, M. H. Dilhuydy, and B. Barreau. Measuring image texture to separate "difficult" from "easy" mammograms. *British Journal of Radiology*, 67:456–463, 1994.

[190] G. M. te Brake and N. Karssemeijer. Comparison of three mass detection methods. In *Proc. International Workshop on Digital Mammography*, pages 119–126, 1998.

[191] G. M. te Brake and N. Karssemeijer. Single and multiscale detection of masses in digital mammograms. *IEEE Transactions on Medical Imaging*, 18(7):628–639, 1999.

[192] E. L. Thurfjell, K. A. Lernevall, and A. A. S. Taube. Benefit of independent double reading in a population-based mammography screening program. *Radiology*, 191:241–244, 1994.

[193] S. Timp and N. Karssemeijer. A new 2D segmentation method based on dynamic programming applied to computer aided detection in mammography. *IEEE Transactions on Medical Imaging*, 31(5):958–971, 2004.

[194] G. D. Tourassi, N. H. Eltonsy, J. H. Graham, C. E. Floyd, and A. S. Elmaghraby. Feature and knowledge based analysis for reduction of false pos-

itives in the computerized detection of masses in screening mammography. In *Proc. International Conference IEEE Engineering in Medicine and Biology Society*, pages 6524–6527, 2005.

[195] G. D. Tourassi, R. Vargas-Vorecek, D. M. Catarious, and C. E. Floyd. Computer-assisted detection of mammographic masses: A template matching scheme based on mutual information. *Medical Physics*, 30(8):2123–2130, 2003.

[196] M. A. Turk and A. P. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[197] P. Undrill, R. Gupta, S. Henry, and M. Downing. Texture analysis and boundary refinement to outline mammography masses. In *Proc. IEE Colloquium on Digital Mammography*, pages 511–516, 1996.

[198] P. M. Vacek, B. M. Geller, D. L. Weaver, and R. S. Foster. Increased mammography use and its impact on earlier breast cancer detection in Vermont. *Cancer*, 94(8):2160–2168, 2002.

[199] S. van Engeland, P. Snoeren, N. Karssemeijer, and J. H. C. L. Hendriks. Optimized perception of lesion growth in mammograms using digital display. In *Proc. SPIE Medical Imaging*, volume 5034, pages 25–31, 2003.

[200] S. van Engeland, P. R. Snoeren, H. Huisman, C. Boetes, and N. Karssemeijer. Volumetric breast density estimation from full-field digital mammograms. *IEEE Transactions on Medical Imaging*, 25:273–282, 2006.

[201] R. P. Velthuizen. Computer diagnosis of mammographic masses. In *Proc. of the App. Imagery Pattern Recognition Workshop*, pages 166–172, 2000.

[202] L. A. Venta and R. E. Hendrick. Mammographic density categories – reply. *American Journal of Roentgenology*, 178:242–243, 2002.

[203] M. Viergever and S. Lobregt. A discrete dynamic contour model. *IEEE Transactions on Medical Imaging*, 14:12–24, 1995.

[204] N. Vujovic and D. Brzakovic. Establishing the correspondence between control points in pairs of mammographic images. *IEEE Transactions on Medical Imaging*, 6(10):1388–1399, 1997.

[205] S. K. Warfield, K. H. Zou, and W. M. Wells. Simultaneous truth and performance level estimation (staple): An algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging*, 23:903–921, 2004.

[206] J. Wei, H. P. Chan, B. Sahiner, L. M. Hadjiiski, M. A. Helvie, M. A. Roubidoux, C. Zhou, and J. Ge. Dual system approach to computer-aided detection of breast masses on mammograms. *Medical Physics*, 33(11):4157–4168, 2006.

[207] J. Wei, B. Sahiner, L. M. Hadjiiski, H. P. Chan, N. Petrick, M. A. Helvie, M. A. Roubidoux, J. Ge, and C. Zhou. Computer-aided detection of breast masses on full field digital mammograms. *Medical Physics*, 32(9):2827–2838, 2005.

[208] M. B. Williams, M. J. Yaffe, A. D. A. Maidment, M. C. Martin, J. A. Seibert, and E. D. Pisano. Image quality in digital mammography: Image acquisition. *Journal of the American College of Radiology*, 3:589–608, 2006.

[209] F. Winsberg, M. Elkin, J. Macy, V. Bordaz, and W. Weymouth. Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis. *Radiology*, 89(2):211–215, 1967.

[210] M. Wirth, D. Nikitenko, and J. Lyon. Segmentation of the breast region in mammograms using a rule-based fuzzy reasoning algorithm. *International Journal on Graphics, Vision and Image Processing*, 5(2):45–54, 2005.

[211] J. N. Wolfe. Risk for breast cancer development determined by mammographic parenchymal pattern. *Cancer*, 37:2486–2492, 1976.

[212] J. Yang, D. Zhang, A. F. Frangi, and J. Yang. Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1):131–137, 2004.

[213] F. F. Yin, M. L. Giger, K. Doi, C. J. Vyborny, and R. A. Schmidt. Computerized detection of masses in digital mammograms: Automated alignment of breast images and its effect on bilateral-subtraction technique. *Medical Physics*, 21(3):445–452, 1994.

[214] B. Zheng, W. F. Good, D. R. Armfield, C. Cohen, T. Hertzberg, J. H. Sumkin, and D. Gur. Performance change of mammographic CAD schemes optimized with most-recent and prior image databases. *Academic Radiology*, 10(3):238–288, 2003.

[215] L. Zheng and A. K. Chan. An artificial intelligent algorithm for tumor detection in screening mammogram. *IEEE Transactions on Medical Imaging*, 20(7):559–567, 2001.

[216] C. Zhou, H. P. Chan, N. Petrick, M. A. Helvie, M. M. Goodsitt, B. Sahiner, and L. M. Hadjiiski. Computerized image analysis: Estimation of breast density on mammograms. *Medical Physics*, 28(6):1056–1069, 2001.

[217] S. C. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and Bayes/MDL for multi-band image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9):884–900, 1996.

[218] S. W. Zucker. Region growing: Childhood and adolescence. *Computer Graphics and Image Processing*, 5:382–399, 1976.

[219] R. Zwiggelaar, S. M. Astley, C. R. M. Boggis, and C. J. Taylor. Linear structures in mammographic images: Detection and classification. *IEEE Transactions on Medical Imaging*, 23(9):1077–1086, 2004.

[220] R. Zwiggelaar, L. Blot, D. Raba, and E. R. E. Denton. Set-permutation-occurrence matrix based texture segmentation. In *Lecture Notes in Computer Science*, volume 2652, pages 1099–1107, 2003.

[221] R. Zwiggelaar and E. R. E. Denton. Optimal segmentation of mammographic images. In *Proc. International Workshop on Digital Mammography*, pages 751–757, 2004.

[222] R. Zwiggelaar, I. Muhimmah, and E. R. E. Denton. Mammographic density classification based on statistical gray-level histogram modelling. In *Proc. Medical Image Understanding and Analysis*, pages 183–186, 2005.

[223] R. Zwiggelaar, T. C. Parr, J. E. Schumm, I. W. Hutt, C. J. Taylor, S. M. Astley, and C. R. M. Boggis. Model-based detection of spiculated lesions in mammograms. *Medical Image Analysis*, 3(1):39–62, 1999.