

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO



FEUP

Computer Aided Detection in Mammography

João Paulo da Silva Ferreira Monteiro

Mestrado Integrado em Engenharia Electrotécnica e Computadores
Major Telecomunicações

Orientador: Prof. Dr. Jaime Santos Cardoso

Co-orientador: Me. Jakub Tkaczuk (INESC-Porto)

Janeiro de 2011

Resumo

O cancro da mama é um problema sério de saúde pública sendo responsável pelo segundo maior número de mortes causadas por cancro em mulheres. A detecção precoce é por isso crucial para ajudar a reduzir o número de mortes. Até ao momento, a mamografia é a melhor ferramenta para fazer o rastreio do cancro da mama. No entanto a detecção de lesões em mamografias é um procedimento difícil devido a vários factores sendo os mais proeminentes o grande número de estruturas não-patológicas presentes na imagem. A utilização dos computadores e redes em ambiente médico para além de simplificar o acesso aos dados do paciente contribuiu fortemente para melhorar a qualidade e as tomadas de decisão nos cuidados de saúde. Os arquivos médicos digitais e bases de dados são algumas das melhorias mais importantes que apesar de terem algumas limitações, contribuem para uma prática médica com mais qualidade e menores custos. Os sistemas de diagnóstico assistido por computador são um exemplo destes sistemas que podem representar uma melhoria na eficiência do diagnóstico na área do cancro da mama.

Esta tese foi desenvolvida no âmbito do projecto SPACS do INESC-Porto e também contou com a orientação na empresa do investigador Jakub Tkaczuk (MSc). O objectivo do projecto é desenvolver um módulo de software para detectar e diagnosticar lesões suspeitas em mamografias. O objectivo desta dissertação é o de fornecer uma contribuição importante ao projecto sobretudo nas áreas de métricas de avaliação de algoritmos de detecção e na melhoria dos algoritmos em si.

Este trabalho introduziu com sucesso novas métricas de avaliação de algoritmos que são robustas, rigorosas e previsíveis. Espera-se que estas métricas venham a contribuir para uma melhor uniformização das métricas usadas pela comunidade científica dedicada ao processamento de imagem em imagens médicas. O método proposto de segmentação de massas obteve resultados encorajadores visto que em relação ao projecto representa uma melhoria de uma área segmentada correctamente de cerca de 20%. A abordagem implementada de detecção de microcalcificações foi comparada com a utilizada pelo SPACS o que permitiu destacar algumas limitações de ambos os métodos.

Abstract

Breast cancer is a major health concern being responsible for the second most number of deaths caused by cancer in women. Early detection is therefore crucial to help reduce the number of deaths. At the moment, mammography is the best tool available to screen for early breast cancer. However, the detection of lesions in mammograms is a difficult procedure due to several factors the most prominent one being the high number of non-pathological structures present in the image. The arrival of computers and data networks in the medical environment, besides simplifying the access to patient data and related services, has strongly contributed to improve the quality and decision taking in healthcare. Digital Medical archives and databases are one of the most important improvements that, although still presenting a few limitations, significantly contribute to better medicine at a lower cost. Computer Aided Detection systems are an example of these systems that can help increase the diagnostic efficiency of breast cancer.

This thesis was developed under the SPACS project currently ongoing at INESC-Porto and also counted with the orientation in the company of researcher Jakub Tkaczuk (MSc). The project's aim is to develop a software module to integrate with current Picture Archiving and Communication Systems (PACS) that acts as a CADe and CADx of suspicious lesions in mammograms. The objective of this dissertation is to provide a valuable contribution to the SPACS project mainly in the areas of the evaluation metrics of the detection/segmentation algorithms and also in improving the performance of the algorithms themselves.

This work has successfully introduced a new framework to evaluate the performance of microcalcifications or mass detection algorithms that are robust rigorous and predictable. These metrics will hopefully contribute to a better uniformisation of the performance evaluation metrics used within the medical image processing community. The proposed segmentation method obtained some encouraging results because it meant a significant improvement relative to the region growing algorithm since it obtained an additional 20% of overlap area in the enhanced version of the ROI. The implemented microcalcifications approach was compared to the current detection algorithm which highlighted some limitations of both approaches and allowed to devise some directions in the future work.

Contents

1	Introduction	1
1.1	Motivation	1
1.1.1	Breast Cancer	1
1.1.2	Breast Cancer Screening	1
1.1.3	Computer-Aided Diagnosis	2
1.2	Mammography	2
1.2.1	Full Field Digital Mammography	2
1.2.2	Screening Mammography	3
1.2.3	Types of lesions	3
1.3	Classification	4
1.4	Database	4
1.5	Objectives	6
1.6	Structure of the Dissertation	7
2	Literature Review	11
2.1	Overview	11
2.1.1	Typical CADx system	11
2.1.2	Performance evaluation metrics	11
2.2	Detection and Classification Algorithms of Microcalcifications	12
2.3	Detection and Classification Algorithms of Masses	13
2.3.1	Segmentation of masses	14
2.4	Summary and Discussion	15
2.4.1	Microcalcifications	15
2.4.2	Masses	15
2.5	Fundamentals	15
2.5.1	Spacial Filtering	15
2.5.2	Frequency Filtering	19
3	Performance Evaluation Metrics	21
3.1	Motivation and Overview	21
3.2	Performance Evaluation of Microcalcifications	22
3.2.1	Algorithm Description	22
3.2.2	Illustrative example	23
3.2.3	Cluster Detection	25
3.3	Performance Evaluation of Masses	25
3.3.1	Evaluation of Mass Segmentation Algorithms	26

4	Mass Segmentation Method	27
4.1	Base Algorithm	27
4.2	Evaluation	31
4.2.1	Procedure	31
4.2.2	Results	32
5	Microcalcifications Detection	35
5.1	SPACS method	35
5.2	Implemented method	36
5.3	Evaluation	37
5.3.1	Procedure	37
5.3.2	Results	37
6	Conclusion and Future Work	41
6.1	Contributions	41
6.2	Future Work	41

List of Figures

1.1	Example of a CC view (left) and MLO view (right) of a breast	4
1.2	Example of a cluster of microcalcifications	6
1.3	Example of several types of masses	7
1.4	Example of an architectural distortion	8
1.5	Example of an annotation of the HSJ database	9
2.1	Graphical representation of the Thresholding transform	16
2.2	Illustration of a filtering operation	17
2.3	Examples of masks used for averaging	18
2.4	Mask used to implement the Laplacian	18
2.5	Mask used to implement the Prewitt operator	19
3.1	Centroids of the GT image	23
3.2	Centroids of the OA image	23
3.3	Centroids of the OA and GT in the same image	24
3.4	Graphical representation of the distance transform	25
4.1	ROI	27
4.2	ROI in polar coordinates	28
4.3	Gradient of the ROI in polar coordinates	29
4.4	Gradient of the ROI in polar coordinates after post-processing	30
4.5	Shortest path of the ROI in polar coordinates	30
4.6	Contour of the mass in Cartesian coordinates	31
4.7	Example of a segmentation with good results	33
4.8	Example of a segmentation with poor results	34
5.1	Example of a detection - Implemented algorithm	38
5.2	Example of a detection - SPACS	39
5.3	Example of a detection - Implemented algorithm	40

List of Tables

1.1	BI-RADS Assessment Categories.	5
1.2	Different classes of the database and corresponding number of cases	5
3.1	Distance Matrix	24
3.2	Distance Matrix after the distance transform is applied	25
4.1	Mean values of the distances and overlap - Shortest-Path Method	32
4.2	Mean values of the distances and overlap - Region-growing algorithm	32
5.1	Detection results comparisson	37

Abbreviations and Symbols

CADe	Computer-Aided Detection
CADx	Computer-Aided Diagnosis
HSJ	Hospital São João
SPACS	Semantic Picture Archiving and Communication Systems
GT	Ground Truth Image
OA	Output Algorithm Image
ROI	Region Of Interest

Chapter 1

Introduction

In the following sections, we introduce the work developed within the scope of this thesis. We describe the motivation for this thesis as well as a few concepts and terminology that the reader should be acquainted before reading further chapters. We also define the objectives that should be achieved. Finally, in the last section, an overview of the thesis and its structure will be presented.

1.1 Motivation

1.1.1 Breast Cancer

According to the World Health Organization, breast cancer was responsible for approximately 519 000 deaths in 2004 comprising 16% of all cancer incidence among women [1]. It constitutes the most common form of cancer in women worldwide and is the second leading cause of cancer deaths after lung cancer. In Portugal 1500 women die every year due to breast cancer whereas in the European Union it causes one in six of all deaths from cancers in women [2]. Some of the most important risk factors are reproductive (early menarche, late menopause, late age at first childbirth) [3] as well as alcohol consumption, obesity and physical inactivity [4]. The number of breast cancer cases is particularly rising in developing countries where most of the cases are diagnosed in advanced stages. Early detection is therefore essential in order to achieve a reduction in mortality rates.

1.1.2 Breast Cancer Screening

Breast cancers that are found because they are causing symptoms tend to be larger and are more likely to have already spread beyond the breast. On the other hand, breast cancers found during screening exams (exams performed on asymptomatic woman to detect early, clinically unsuspected breast cancer) are more likely to be small and still confined to the breast. Whether a

tumour is in situ or invasive along with its size are some of the most important factors in predicting the outlook (prognosis) of a woman with this disease. The most reliable means of detecting breast cancer in its early stages is, under the best scientific evidence, to conduct regular screening of presumably healthy women. Studies show that early detection tests for breast cancer results in a reduced mortality rate of 15-20% [5]. However, even though screening mammography increases breast cancer survival rate, it is not perfect. Estimates of the numbers of cancers missed by mammography are usually around 20% [6]. Reasons for not seeing the cancer include poor positioning or technique, dense tissue obscuring a lesion, subtle features of malignancy and observer error (perception error, incorrect interpretation of a suspect finding).

Furthermore, the classification of lesions as benign or malignant based on their appearance in mammograms is a difficult task even for experienced radiologists [7]. The considerable amount of overlap of breast tissue sometimes produces patterns that appear like suspicious masses to a radiologist or alters the appearance of real abnormalities. Mammography has a positive predictive value (PPV) of less than 35% [8], where the PPV is defined as the percentage of lesions subjected to biopsy that were found to be cancer. Thus, a high proportion of biopsies are performed on benign lesions. This causes stress, anxiety and discomfort for the patients as well as unnecessary expenses spent in examinations.

1.1.3 Computer-Aided Diagnosis

In order to increase diagnostic efficiency, clinical decision support systems based on advanced image processing and machine learning techniques can be used to locate and classify possible lesions. Computer-aided detection (CAdE) systems aid radiologists in detecting mammographic lesions that may indicate the presence of breast cancer and computer-aided diagnosis systems (CAdx) assist the radiologist in the classification of mammographic lesions as benign or malignant. These systems act only as a second reader so the final decision is made by the radiologist [9].

CAdE systems help scan digital images for typical appearances and to highlight conspicuous sections. This allows the radiologist to draw more accurate conclusions about the condition of the pathology by focusing his attention on the marked area. CAdx systems support radiologists in the discrimination between benign and malignant mammographic abnormalities by presenting a diagnosis suggestion.

1.2 Mammography

1.2.1 Full Field Digital Mammography

Full Field Digital Mammography is a medical imaging modality in digital format. With digital mammography, low energy x-rays pass through the breast exactly like screen-film mammograms but are recorded by an electronic sensor instead of the film. Digital mammography offers several advantages over film screen mammography by improving resolution, contrast and signal to noise

ratios which can lead to higher detection rates [10]. Some other advantages are the absence of developing or handling artefacts, near instantaneous image acquisition, low patient radiation and the ability to transmit images electronically. The most important application however is the possibility to use image processing techniques (such as CAdE) to manipulate the image and better visualize suspicious regions that would be difficult to see on conventional mammography. In fact, studies show that the development of FFDM was a catalyst in the increase of such computer systems [11]. The adoption of digital mammography has also allowed the creation of publicly available databases for use by the medical and mammographic image analysis research community. Despite this fact, some of the most widely used data sets are digitized screen-film mammograms like the Digital Database for Screening Mammography (DDSM) [12] and the Mammographic Image Analysis Society (MIAS) Mini-Mammographic Database [13].

1.2.2 Screening Mammography

There are two basic types of mammographic examinations: screening mammography and diagnostic mammography [14]. In screening mammograms two views of each breast are recorded; the craniocaudal (CC) view, which is a top-to-bottom view, and a mediolateral oblique (MLO) view, which is a side view taken at an angle (Figure 1.1). Diagnostic mammography is an exam adapted to the individual patient performed to evaluate a breast complaint or abnormality detected by physical exam or routine screening mammography. Some of the quality assessment criteria include symmetry of right and left mammograms, no motion blurring, no skin folds but most importantly, the adequate visualization of the glandular tissue must be guaranteed [15]. In fact it is crucial that these criteria are met since, as mentioned earlier, poor positioning is one of the factors responsible for a radiologist missing an abnormality.

1.2.3 Types of lesions

There are a large number of different types of mammographic abnormalities that may indicate the presence of breast cancer. Some of the most important signs are microcalcifications and masses as well as architectural distortions. Calcifications are tiny deposits of calcium, which appear as small bright spots on the mammogram (Figure 1.2). They do not indicate the presence of cancer by themselves but because many breast cancers contain calcifications, it is important to determine whether or not they are within cancerous tissue. They are characterized by their type and distribution properties.

A mass is a generic term to indicate a localised swelling, protuberance, or lump in the breast and it is characterised by its location, size, density (fat containing masses, low density, isodense, high density), shape (round, oval, lobular, irregular), margin (circumscribed, microlobular, obscured, indistinct, spiculated), and associated findings [16, 17] (Figure 1.3).

Architectural distortion (Figure 1.4) is a distortion of the normal architecture with no definite mass visible, including spiculations radiating from a point and focal retraction or distortion at the

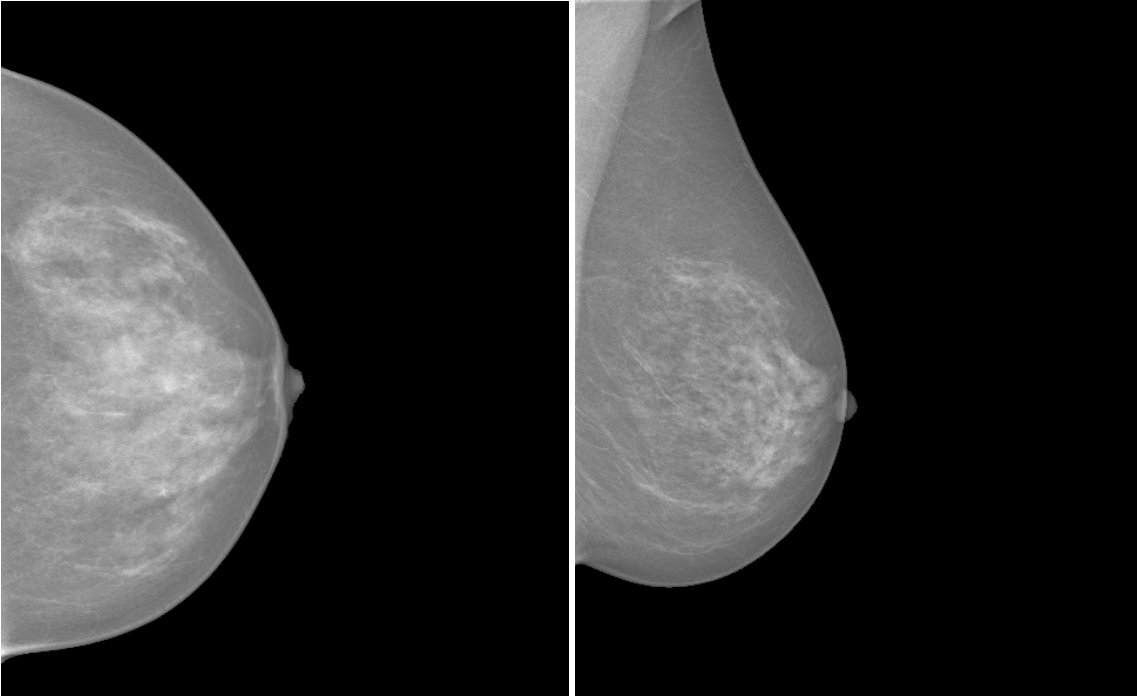


Figure 1.1: Example of a CC view (left) and MLO view (right) of a breast (selected from the HSJ database).

edge of the parenchyma. Other predictors of breast cancer include bilateral asymmetry, curvilinear structures (CLS), and breast tissue density.

1.3 Classification

Generally mammograms are classified under the Breast Imaging Reporting and Database System (BI-RADS TM) developed by the American College of Radiology (ACR). It contains a lexicon for standardized terminology (descriptors) for breast imaging, as well as Standard Reporting with Final Assessment Categories and guidelines for Follow-up and Outcome Monitoring [17]. There are 7 Assessment Categories ranging from 0 to 6 as shown in Table 1.1.

1.4 Database

The work developed in the scope of this thesis was tested on the database created by the Semantic Picture Archiving and Communication System (SPACS) project at INESC-Porto. The mammograms were collected at Hospital de São João (HSJ), Porto and consist of 410 anonymized digital images. The images were saved in the DICOM format and are of size 3328x4084 or 2560x3328 (depending on the breast size of the patient). The pixel size is $70\ \mu m$ and 14bit contrast resolution. In addition, the medical reports corresponding to each mammogram were also saved. All the mammograms were divided into six classes (Table 1.2).

Table 1.1: BI-RADS Assessment Categories.

Category	Description
0	Need Additional Imaging Evaluation and/or Prior Mammograms For Comparison
1	Negative
2	Benign Finding(s)
3	Probably Benign Finding(s) Initial Short-Interval. Follow-Up Suggested
4	Suspicious Abnormality Biopsy Should Be Considered
5	Highly Suggestive of Malignancy Appropriate Action Should Be Taken
6	Known Biopsy Proven Malignancy Appropriate Action Should Be Taken

Table 1.2: Different classes of the database and corresponding number of cases

Classes	Number of cases
Assymetry	12
Calcifications	221
Multiple Findings	79
Arquitectural Distortions	3
Normal	67
Masses	28
Total	410

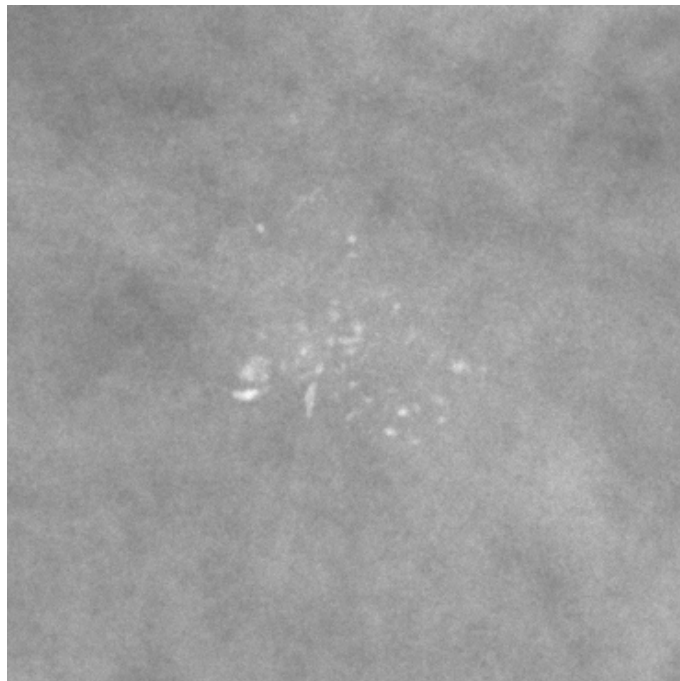


Figure 1.2: Example of a cluster of microcalcifications (selected from the HSJ database).

The motivation behind the creation of this database is that most of the databases available have some limitations such as poor annotations (whether they are not precise or just consist in a circle around the ROI) which makes them unsuited for CADe/CADx algorithms to be tested in. In this case all the lesions were manually segmented by an specialist in the field, and validated by a second specialist (Figure 1.5). The aim of the project is also to make the HSJ database public in a near future, but at the moment, it still remains private to be used only within the SPACS project.

1.5 Objectives

This thesis is being developed under the SPACS project currently ongoing at INESC-Porto. The project's aim is to develop a software module to integrate with current Picture Archiving and Communication Systems (PACS) that acts as a CADe and CADx of suspicious lesions in mammograms. The system also supports automatic, semantic based, description and search methods directly over medical images but the main focus of this dissertation will be related to the Computer Aided Detection (CADe) module.

The objective of this dissertation is to provide a valuable contribution to the SPACS project mainly in the areas of the evaluation metrics of the detection algorithms and also in improving the performance of the algorithms themselves.

It is important to realize that mammographic image analysis is an extremely difficult task due to several factors. The large variability in the appearance of lesions and the fact that they are often

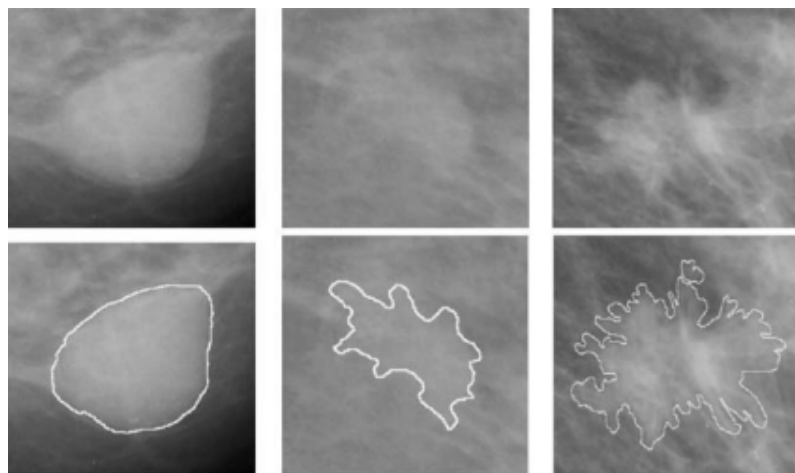


Figure 1.3: Example of several types of masses: round and circumscribed mass (left), lobular and well-defined mass (centre) and a spiculated shape and ill-defined margin (right) (retrieved from [16]).

occluded in dense breast tissue makes detection very difficult. Also, since the efficacy of CAD systems can have very serious implications, there is a need for near perfection.

1.6 Structure of the Dissertation

Besides this introductory chapter, there are more five chapters in this dissertation. In chapter 2, a review of the state-of-the-art will be made. Additionally, the results obtained will be discussed comparing the results and highlighting the advantages and disadvantages of each one. Chapter 3 introduces the metrics used to evaluate the performance of the detection algorithms while chapter 4 deals with the segmentation of masses. Chapter 5 discusses the detection of microcalcifications and finally chapter 6 presents the conclusions and future work.

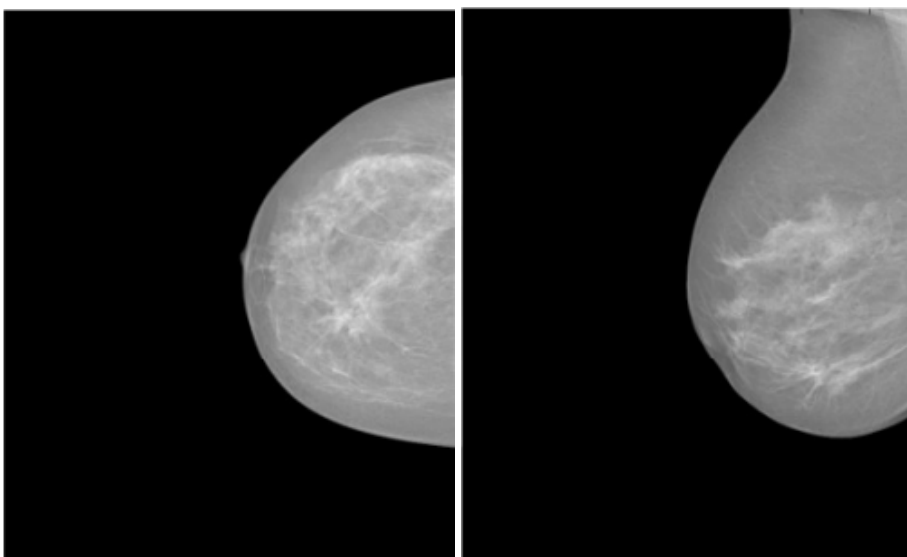


Figure 1.4: Example of an architectural distortion (selected from the DDSM database).

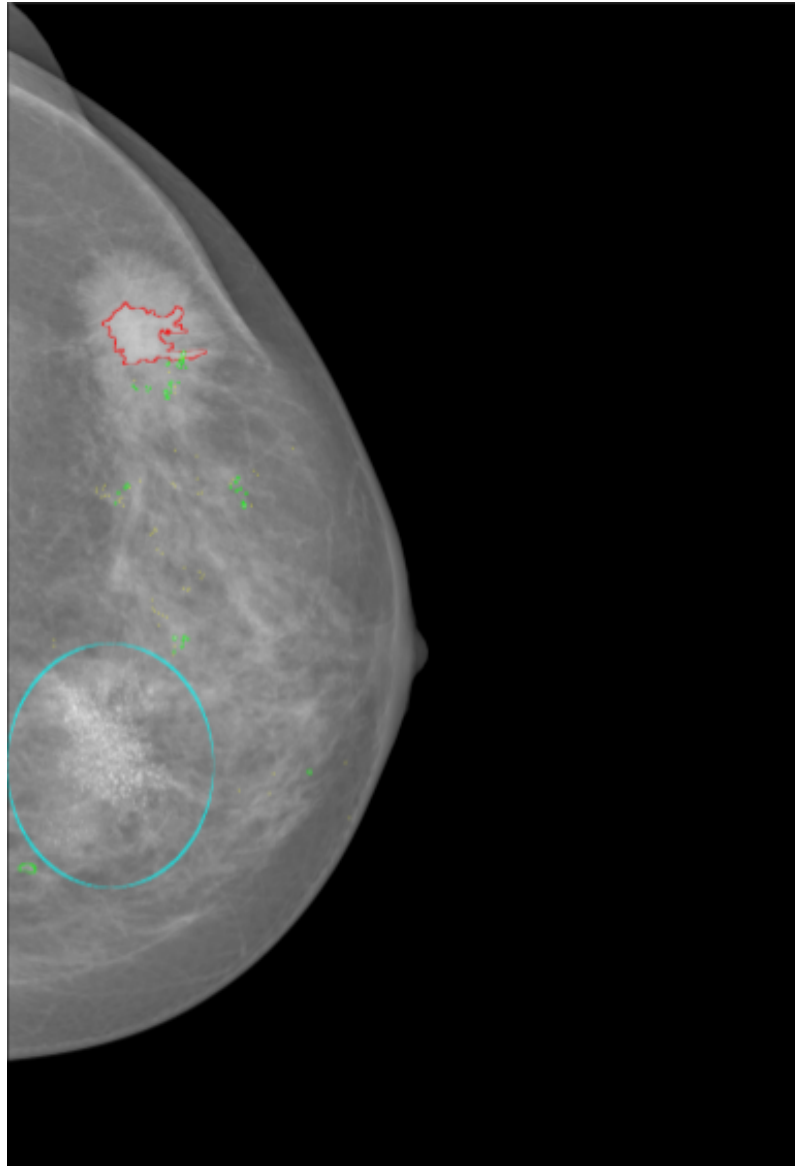


Figure 1.5: Example of a mammogram annotation of the HSJ database. Masses are in red, small microcalcifications in yellow, large microcalcifications in green, and clusters of microcalcifications in blue.

Chapter 2

Literature Review

This chapter is divided in four sections. The first section provides an overview of a typical CADx system and presents the metrics used to evaluate the performance of these systems. Sections 2 and 3 discuss the detection and classification of calcifications and masses respectively and were based on the work done by Rangayyan et al. [18]. Section 4 provides a summary and discussion. Section 5 provides some background knowledge on image processing techniques used in this work.

2.1 Overview

2.1.1 Typical CADx system

Most CADx systems have a five-stage process: pre-processing, segmentation, feature extraction, feature selection, and classification. The goal of the pre-processing stage is to enhance the quality of a mammogram by increasing the signal-to-noise ratio, dynamic range, and contrast. In the segmentation stage, the region of interest (ROI) is extracted. ROI is typically a rectangular sub image from the mammogram that contains lesions as well as normal tissue. The segmentation of the lesions from the normal tissue is also done in this step. In the feature extraction stage, the features are determined from the characteristics of the abnormalities segmented in the previous step. In the feature selection phase, the best subset features is selected so that the lesion can be classified in the following stage.

2.1.2 Performance evaluation metrics

The metrics used to evaluate the performance of detection algorithms are sensitivity (Equation 2.1) and the number of false positives per image (Equation 2.2). Sensitivity measures the proportion of marks made by the detection system that correspond to the actual location of a lesion. On the other hand, a false-positive mark is a mark made by the CAD system that does not correspond

to the location of a lesion. A plot of sensitivity versus the number of false positives per image (FPI) is called a receiver operating characteristic (ROC) plot [19] and this is generally used to report the performance of the detection algorithm.

$$Sensitivity = \frac{\text{Number of True-Positive Marks}}{\text{Total Number of Lesions}} \quad (2.1)$$

$$FPI = \frac{\text{Number of False-Positive Marks}}{\text{Total Number of Lesions}} \quad (2.2)$$

The metrics used to report the accuracy of classification algorithms are sensitivity and specificity as well as FPI. Sensitivity, in this context, is defined as a lesion for which the CADx predicts to be cancerous and is actually found to be malignant (Equation 2.3). Specificity is the fraction of benign lesions that are correctly identified by the CAD as being benign (Equation 2.4).

$$Sensitivity = \frac{\text{Number of True-Positive Classifications}}{\text{Number of Malignant Lesions}} \quad (2.3)$$

$$Specificity = \frac{\text{Number of True-Negative Classifications}}{\text{Number of Benign Lesions}} \quad (2.4)$$

2.2 Detection and Classification Algorithms of Microcalcifications

Calcifications are tiny deposits of calcium, which appear as small bright spots of various sizes and shapes. Calcifications associated with benign diseases are generally large, round, small in number, diffusely distributed, and homogeneous in size and shape. Malignant calcifications tend to be numerous, clustered, small, varying in size and shape, angular, irregularly shaped, and branching in orientation [20, 21].

Shen et al. [22] reported a method for the detection and classification of calcifications. In the segmentation stage, a multi-tolerance region growing procedure for the detection of potential calcification regions and the extraction of contours is used. Next, in the feature extraction step, three measures of shape features, including compactness, moments, and Fourier descriptors are computed for each region. The feature vectors obtained are then classified using a neural network. 94% of the benign calcifications were correctly classified and the correct classification rates malignant calcifications were 87%.

Yu et al. [23], proposed a method for the detection of calcification based in two steps. In the first one, potential microcalcification pixels are segmented, and in the second stage, individual microcalcifications are detected. The first step, wavelet features and gray level statistical features are used to segment the microcalcification pixels. In the second step, potential individual microcalcification objects are classified as true or false individual microcalcification objects using a set of 31 features followed by a general regression neural network classifier. A sensitivity of 90% was obtained with 0.5 false positive per image in a database with 40 mammograms.

Qian et al. [24] reported a method for segmenting microcalcifications based on the wavelet transform. In the first step, a tree structured non-linear filter is used to suppress image noise. In the second stage the ROI is decomposed into its sub bands using the wavelet transformation and the coefficients of the sub-bands are weighted so that microcalcifications are enhanced, and background tissue is suppressed once the inverse wavelet transform is applied to the data. In the final step, automatic linear scaling is used to display the resulting image. These algorithms were tested in a database of 100 single view mammograms and obtained a sensitivity of 96% at 1.6 false-positives per image.

Wei et al. [25] presented a method of detecting calcifications using relevance vector machine. This machine learning technique is based on Bayes estimation theory and as a feature it can yield a decision function that depends on only a very small number of relevance vectors. The proposed method was evaluated in a database of 141 mammograms and obtained a sensitivity of 90% at 1 false positive per image. The statistical performance of the classifier was essentially the same as a support vector machine classifier developed previously. However, the RVM classifier greatly reduces the computational complexity.

Yu et al. [26] proposed a detection system based in two stages. In the first stage, suspicious microcalcifications are detected by thresholding a filtered mammogram via a wavelet filter according to the MPV (mean pixel value) of that image. The second stage is designed to decrease the false positive rate. Markov random field parameters are extracted from the neighbourhood of the suspicious microcalcifications along with three auxiliary texture features which then serve as input to the classifiers. These algorithms were tested on a database of 25 mammograms containing 25 areas of clustered microcalcifications. The method achieved a sensitivity of 92%, with 0.75 false positives per image.

2.3 Detection and Classification Algorithms of Masses

A mass is a three-dimensional structure demonstrating convex outward borders, usually evident on two orthogonal views. Masses are characterised by its location, size, shape, margin, and associated findings. Malignant masses generally have a spiculated, rough and blurry boundary. Benign masses are usually round and oval shaped with smooth and circumscribed margins.

Mudigonda et al. [27] proposed a method for segmenting suspicious masses based on the analysis of iso-intensity contour groups. False positives are then eliminated using features based on flow orientation in adaptive ribbons of pixels across the margins of masses. This procedure was tested on 56 images from the Mini-MIAS database and obtained a sensitivity of 81% at 2.2 false positives per image. Furthermore, the masses were classified as benign or malignant using five texture features based on gray-level co-occurrence matrices (GCM's) and using features on a logistic regression method. Benign vs. malignant classification resulted in an accuracy of 0.79 with 19 benign and 13 malignant cases.

Liu et al. [28] presented a method for detecting spiculated masses using the wavelet transform to obtain a multi-resolution representation of the original image. A set of features is then extracted

at each resolution and serve as input to a binary tree classifier. The accuracy of the algorithm on the MIAS database achieved 84.2% true positive detection at less than 1 false positive per image.

Zheng et al. [29] proposed a procedure that combines several artificial intelligence techniques with the discrete wavelet transform (DWT). A dimension analysis serves to determine the ROI. The segmentation is done using a multi-resolution Markov random field algorithm and finally a tree type classification strategy is applied. The algorithm was tested in the Mini-MIAS database and has a sensitivity of 97.3% with 3.9 false positives per image.

Zwiggelaar et al. [30] focuses on the detection of spiculated lesions in mammograms. A generic method of representing patterns of linear structures is described, which relies on the use of factor analysis to separate the systematic and random aspects of a class of patterns. The appearance of central masses is modelled using local scale-orientation signatures based on recursive median filtering. For lesions of 16 mm and larger this approach obtained a sensitivity of 80% at 0.014 false positives per image, whilst the mass detection approach results in a sensitivity 80% at 0.23 false positives per image.

Rangayyan et al. [31] presented a method for classifying as benign or malignant spiculated or circumscribed masses. To do so, it uses a set of shape factors (compactness, Fourier descriptors, moments, and chord-length statistics) and edge acutance. The algorithm was tested on 39 images from the MIAS database and an additional set of 15 local cases containing 16 circumscribed benign, seven circumscribed malignant, 12 spiculated benign, and 19 spiculated malignant lesions. It obtained a classification accuracy of 95%.

2.3.1 Segmentation of masses

This subsection discusses the segmentation of masses which can be done given the full image or just a small region of the mammogram. A popular approach to the segmentation of masses is to use region growing algorithms. The problem resides in the choice of the initial seed points. Zhang et al. [32] proposes a method that uses a modified region-growing technique where instead of the traditional threshold of gray-level difference it uses three different criteria. The initial seed points corresponded to the local maximum of small regions. It was tested on 40 mammograms of the MIAS database and obtained a true positive rate of 90% at 1.3 False Positives per image.

Some methods use the watershed transform to segment detected masses. If we think of a gray-scale image as a topological surface, the catchment basins would correspond to the low points in intensity. If rain would keep falling on these catchment basins the water level would rise until it would merge on the watershed ridge. The watershed transform computes the catchment basins and ridge lines of a given image. However, not always the objects we want to segment correspond to the catchment basins of an image. Most of the methods have a pre-processing step to address this problem. Sheshadri et al. [33] applied a pre-processing scheme that uses a local contrast enhancement of a ROI containing a suspicious finding then performs a morphological smoothing of the contrast enhanced image and finally computes the morphological gradient of the smoothed image.

Other methods rely on the boundaries of the regions to extract the contours. The method by [34] utilizes a geometric active contour model that minimizes an energy function based on the homogeneities inside and outside of the evolving contour. This method was tested on a FFDM database of which 85% were correctly segmented.

2.4 Summary and Discussion

2.4.1 Microcalcifications

Despite the wide variety of approaches and their apparent success, the problem of detecting and classifying microcalcifications is not fully solved. The segmentation of objects that have a very low contrast against the background still remains a difficult task as well as the detection of calcifications in mammograms with high noise levels and curvilinear structures (CLS).

Furthermore, the results presented above can be misleading as it is very difficult to make a fair comparison of the different algorithms as they are often evaluated with private or local databases. There is a lack of comparative studies that evaluate the performance of these algorithms on the same datasets since the performance of a CADx algorithm can vary dramatically on different databases.

2.4.2 Masses

The task of efficiently detecting mammographic masses is far from solved and still needs improvement. This may be due to the large variability in lesion shape, size and surrounding breast tissue type. State of the art algorithms find masses with ill-defined margins the most difficult to detect followed by spiculated margins and finally circumscribed margins. As previously discussed in the section on calcification segmentation a direct comparison of these methods is very hard since they are often tested on different databases. There are almost no comparative studies because rarely source code for novel methods is made publicly available for research purposes.

2.5 Fundamentals

2.5.1 Spatial Filtering

There are two main types of processing operations in the spatial domain: intensity transformations and spatial filtering. In the intensity transformations the value of each pixel is altered depending of its intensity alone. The value of a given pixel in the output image s is computed by applying a transform T to the pixel of the original image r (Equation 2.5).

$$s = T(r) \quad (2.5)$$

One common example of intensity transform is the Thresholding operation (Figure). Segmentation through this technique is accomplished by scanning the image pixel by pixel and labelling

each pixel as object or background, depending whether the gray level of that pixel is greater or less than the value of the threshold m .

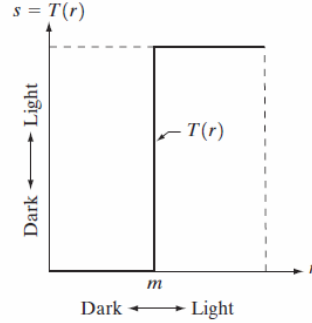


Figure 2.1: Graphical representation of the Thresholding transform (retrieved from [35])

In spatial filtering operations the value of a given pixel in the output image depends of the intensity of the pixels in its neighbourhood and the corresponding values of a subimage that has the same dimensions as the neighbourhood. This is why sometimes these operations are referred as neighbourhood operations or spatial convolution. In fact, intensity transformations are the simplest form of these neighbourhood operations where the neighbourhood's size 1×1 .

As shown in Figure 2.2, at each point (x,y) the response of the linear filter at that point is the sum of products of the filter coefficients and the corresponding neighbourhood pixels in the area covered by the filter mask. Formally, the filtering of an image f of size $M \times N$ with a filter mask of size $m \times n$ is given by the convolution of the original image $f(x,y)$ with the mask $w(x,y)$ as shown in Equation 2.6 where $a = \frac{(m-1)}{2}$ and $b = \frac{(n-1)}{2}$.

$$g(x,y) = f(x,y) \otimes w(x,y) = \sum_{s=-a}^a \sum_{t=-b}^b w(s,t) f(x+s,y+t) \quad (2.6)$$

Figure 2.3 shows an example of two masks used in averaging filters. Substituting the value of each pixel by the average of intensities of its neighbourhood results in an image with the high frequencies attenuated or smoothed. Since random noise is typically of high frequencies, one obvious application of these averaging filters is noise reduction. However, these types of filters have a disadvantage: they blur the contours present in the image since they are also characterized by high frequencies. In Figure 2.3 the mask on the left represents the arithmetic mean whereas the mask on the right, represents a weighted average where the contribution of the neighbor pixels have different weights depending on the distance to the central pixel.

Depending on the coefficients of the mask, the filter can also be of two common types: *low-pass* or *high-pass*. The low-pass filters are used to remove or smooth small details such as noise. The high-pass filters mainly enhances the high frequency transitions in the image. These filters are sensitive to discontinuities in the image therefore are based on derivatives. The amplitude of the filter is proportional to the degree of discontinuity in the center pixel. Thus, image differentiation

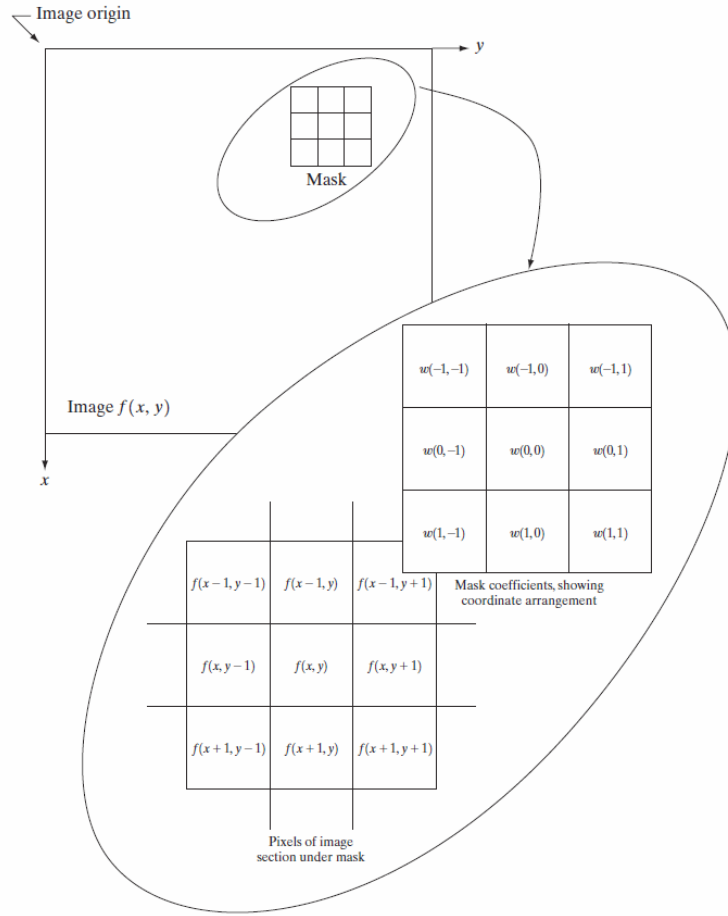


Figure 2.2: Illustration of a filtering operation (retrieved from [35])

enhances edges and other discontinuities (such as noise) and deemphasizes areas with slowly varying gray-level values.

The derivatives of a digital function $f(x)$ are defined in terms of differences (Equation 2.7). First-order derivatives generally produce thicker edges in an image and also have a stronger response to a gray-level step.

$$\frac{\partial f}{\partial x} = f(x+1) - f(x) \quad (2.7)$$

Second-order derivatives (Equation 2.8) have a stronger response to fine detail, such as thin lines and isolated points and produce a double response at step changes in gray level.

$$\frac{\partial^2 f}{\partial x^2} = f(x+1) + f(x-1) - 2f(x) \quad (2.8)$$

The *Laplacian* is a second order operator that is used for enhancement and is defined by Equation 2.9.

$\frac{1}{9} \times$	1	1	1
	1	1	1
	1	1	1

$\frac{1}{16} \times$	1	2	1
	2	4	2
	1	2	1

Figure 2.3: Examples of masks used for averaging: arithmetic mean (left) and weighted average (right) (retrieved from [35])

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \quad (2.9)$$

Where $\frac{\partial^2 f}{\partial x^2}$ is defined by Equation 2.10 and $\frac{\partial^2 f}{\partial y^2}$ is defined by 2.11.

$$\frac{\partial^2 f}{\partial x^2} = f(x+1, y) + f(x-1, y) - 2f(x, y) \quad (2.10)$$

$$\frac{\partial^2 f}{\partial y^2} = f(x, y+1) + f(x, y-1) - 2f(x, y) \quad (2.11)$$

Substituting Equation 2.10 and 2.11 in 2.9 yields Equation 2.12

$$\nabla^2 f = f(x+1, y) + f(x-1, y) + f(x, y+1) + f(x, y-1) - 4f(x, y) \quad (2.12)$$

The diagonal directions can be incorporated in Equation 2.12 by adding two more terms, one for each of the two diagonal directions. The end result is implemented with the mask in Figure 2.4.

1	1	1
1	-8	1
1	1	1

Figure 2.4: Mask used to implement the Laplacian (retrieved from [35])

Since the Laplacian is a second order derivative operator, it will enhance discontinuities in an image and smooth regions with low intensity variations. Therefore, the result of applying this operator to an image results in an gray scale image with contours and other discontinuities. By adding this image with the original one, the result is an enhanced version of the original one where the edges and fine detail are highlighted.

First order derivatives are based an approximation of the 2-D gradient of an image and are used to detect edges. The gradient of a given function $f(x,y)$ is given by equation 2.13 and its magnitude by 2.14.

$$\nabla f = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \quad (2.13)$$

$$G = \sqrt{G_x^2 + G_y^2} \quad (2.14)$$

Unlike the Laplacian, the partial derivatives are not isotropic (rotation invariant) but the magnitude is. This operator can be implemented by using the Prewitt operator (Figure 2.5). The left filter emphasizes horizontal edges and the right filter emphasizes vertical edges.

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

Figure 2.5: Mask used to implement the Prewitt operator (retrieved from [35])

2.5.2 Frequency Filtering

Working in the frequency domain, offers several advantages like the flexibility in designing and implementing filtering solutions in enhancement and restoration problems.

The computation of the Fourier transform of a digital image $f(x,y)$ of size $M \times N$ is achieved by computing the Discrete Fourier Transform (implemented by the Fast Fourier Transform algorithm) which is given by Equation 2.15.

$$F(u,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) e^{-j2\pi(\frac{ux}{M} + \frac{vy}{N})} \quad (2.15)$$

Filtering in the frequency domain is simple conceptually: it consists in multiplying the Fourier transform of the original image $F(u,v)$ by the filter in the frequency domain $H(u,v)$ and then

compute the inverse Fourier transform of the result in order to obtain the image in the spatial domain $f(x, y)$ (Equation 2.16).

$$f(x, y) = \mathcal{F}^{-1}[H(u, v)F(u, v)] \quad (2.16)$$

Chapter 3

Performance Evaluation Metrics

In this chapter, a novel performance evaluation framework for microcalcifications and masses will be introduced. In section 1, the motivation for it will be introduced as well as a general overview of the method. As it will become evident in later sections the proposed framework can be applied in the evaluation of both microcalcifications and masses. Therefore, these two subjects will be dealt separately in sections 4 and 5 for microcalcifications and for masses respectively.

3.1 Motivation and Overview

One of the problems in this line of research is the fact that often, detection algorithms are evaluated using different metrics which produce results that cannot be compared between them. A general consensus in how to evaluate the performance of a detection algorithm is still lacking in the medical image processing community.

The output of the detection/segmentation algorithm (OA) and the manual annotations of the radiologist (Ground Truth - GT) are binary images containing the outline of the finding. The output of the performance evaluation metric is given in the form of number of false positives and missed lesions. These metrics are computed by comparing the results obtained by the algorithm with the manual annotations done by the radiologist.

In most of the metrics, as in [36], this is done at looking individually at each lesion in the GT and finding the correspondent one in the OA within its neighborhood. This kind of local decision has however a disadvantage: if two detected findings in the OA are close to a same individual lesion in the GT, they may be marked as true positives. By introducing the concept of global decision, we propose to solve this problem by considering all the lesions instead of each one individually. The end result is metric that is simple, rigorous and widely applicable, since the existing ones not always obey these standards.

3.2 Performance Evaluation of Microcalcifications

3.2.1 Algorithm Description

Microcalcifications are very small structures, so it's a valid assumption to treat them as singular points. Therefore, the first preliminary step consists in approximating the lesions by their centroids.

The main steps involved in the algorithm are described bellow:

1. **Compute the *Dissimilarity Matrix*** — Determines all the costs between the centroids found in the GT and OA and arranges them in a Matrix.
2. **Compute the optimal assignment** — Assigns each microcalcification in the GT to another in the OA using the Hungarian Algorithm.
3. **Applies the detection threshold** — According to the detection threshold, each microcalcification is classified as detected or not.
4. **Compute the performance indexes** — The number of false positives and the missed rate is determined.

Say all centroids of the GT and OA are in the same image. The first step of the algorithm consists in, for each centroid belonging to the GT, compute a dissimilitude measure between that centroid and all the other centroids belonging to the OA image and arranging them in a Matrix. Let m be the number of centroids in the GT and n be the number of centroids in the OA: the result will be a $m \times n$ matrix (DM). Generically, the entry $DM_{i,j}$ corresponds to the dissimilitude metric or cost between the centroid i in the GT and the centroid j of the OA.

The first step consists in matching each lesion in the GT (rows) with a lesion in the OA (columns). The criteria for matching would be of course the distance. This is, in fact, an combinatorial optimal matching problem: It is necessary to match every lesion in the GT (agent) to exactly one lesion in the OA (task) in such a way that the total cost of the assignment is minimized. This assignment problem is easily solved using the Hungarian or Munkres algorithm.

After the second step of the algorithm, each lesion of the GT will have one lesion in the OA assigned to them. This means that if no other measure was taken, no matter how far a centroid would be from the GT, they would always be considered as a detection. This is not correct especially if there are false positives present, therefore a definition of a detected finding is needed. In order to classify each microcalcification as detected or not, it is therefore necessary to define a threshold: a value above which the finding will not be considered as detected. The choice of this threshold will depend on how precise one would want the detection to be (typical values will be suggested later on).

The final step consists in calculating the performance indexes: missed rate and false positives. From the previous step the number of detected findings is known so the number of false positives is easily calculated by computing the difference between the total number of findings in the OA

and the number of detected findings. As to the missed rate, it consists in the number of lesions in GT that were not detected, so it corresponds to the difference between the number of findings in the GT and the total number of detected findings.

3.2.2 Illustrative example

Consider the following case, where the centroids of the GT image are represented in Figure 3.1 and the centroids of the OA image are in the Figure 3.2. This is of course as illustrative example where only a limited number of lesions is considered, in a real mammogram there might be tens or even hundreds of microcalcifications present.

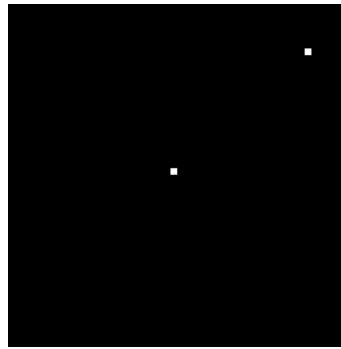


Figure 3.1: Centroids of the GT image - illustrative example.

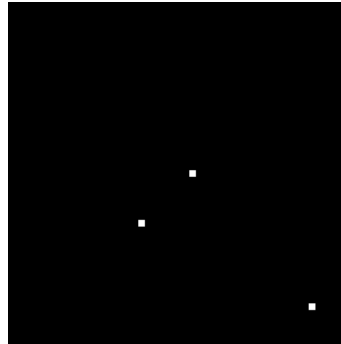


Figure 3.2: Centroids of the OA image - illustrative example.

For illustration purposes all the lesions are represented in the same image (Figure 3.3). From observing this image, it becomes evident that: the lesion B1 will be detected as being lesion A1; lesion A2 will not be detected; B2 and B3 will be marked as false positives.

First it is necessary to define which dissimilitude metric or cost will be used. The metric that immediately comes to mind is of course the distance between the coordinates of the centroid. The distance between pixels can be of various types: *Euclidean*, *cityblock* or *chessboard* to name a few. In this example, the Euclidean distance was chosen. It is also necessary to define the threshold

Table 3.1: Distance Matrix - illustrative example.

	B1	B2	B3
A1	5	20	60
A2	40	60	80

bellow which a finding is considered detected or not. The most typical value is the width of the microcalcification in pixels (in this example it is considered to be five).

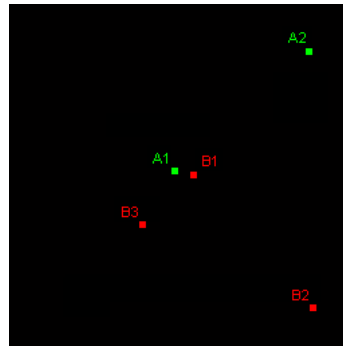


Figure 3.3: Centroids of the GT (in green) and OA (in red) in the same image - illustrative example.

The first step is to determine the distances between centroids to build the corresponding distance matrix (Table 3.1). Note that for simplicity, only integer values for the distance were considered.

Next, the optimal assignment will be computed. But since the objective is to minimize the costs, the Hungarian algorithm will wrongfully match lesion A1 with B2 and A2 with B1 (with a total cost of 60) instead of assigning lesion A1 to B1 and A2 to B2 (with a total cost of 65). This is of course false since the lesion A1 is closer to the lesion B1 than it is to B2. This problem is solved by imposing a maximum distance (Saturation threshold) over which the distance transcribed to the Distance Matrix is always the same (Saturation value). Figure 3.4 shows a graphical representation of this transform. The intuition for this method is that it is as penalizing to error by a distance of 50 as it is by a distance of 500.

Now once again it is necessary to define the saturation threshold (and value which is the same). A typical value would be choosing a saturation threshold which is ten times bigger than the detection one.

The table 3.2 contains the result of this transformation applied to the original distance matrix where the $St=Sv=50$. Now the Hungarian algorithm will work as expected since it will assign lesion A1 to B1 and A2 to B2. Applying the detection threshold of 5 pixels, it gives that only lesion A1 is detected. Therefore, the number of false positives is two: B2 and B3 and the missed rate is one: lesion A2 wasn't detected.

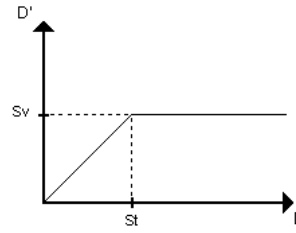


Figure 3.4: Graphical representation of the distance transform. St: Saturation Threshold; Sv: Saturation Value

3.2.3 Cluster Detection

There might be the case where there are clusters of microcalcifications present in an image. In the HSJ database the clusters are annotated with a circle around them unlike isolated microcalcifications whose borders are manually segmented. Since from looking just at the reference there is no way of knowing how many microcalcifications are present, the detection scheme processes the images with clusters normally just like described above but with one additional stage. This last stage consists in comparing whether the marked false positives are inside the cluster or not, and if so, they are no longer considered as false positives.

3.3 Performance Evaluation of Masses

It is now evident that the same method can be applied to evaluate the performance of Mass detection methods. Nevertheless, since Masses are much bigger structures than microcalcifications, treating them as singular points is not a reasonable approximation. For this reason, a few changes were made to the previous method but since they use the same framework, instead of explaining the whole method, only the differences between them will be highlighted.

The main difference compared to the microcalcifications method, resides in the first step. Instead of building a matrix with all the distances between centroids, the metric used is a measure of overlap between the masses. Say A_k is the area occupied by the lesion k . The measure of overlap between the mass i of the GT and the mass j of the OA is given by Equation 3.1.

Table 3.2: Distance Matrix after the distance transform is applied - illustrative example.

	B1	B2	B3
A1	5	20	50
A2	40	50	50

$$O = \frac{A_i \cap A_j}{A_i \cup A_j} \quad (3.1)$$

The next step is to match each finding in the OA to the corresponding in the GT. In this case, the goal is not to minimize the cost (distances) but to maximize the benefit (the amount of overlap) so one might think that the Hungarian algorithm is unsuitable for this case. However, it is valid to say that $\max O \equiv \min(1 - O)$, $O \in [0, 1]$. Since the amount of overlap comes in percentage, it is therefore possible to apply the Hungarian algorithm to the matrix $\hat{O}M = (1 - OM)$.

It is also necessary to define a threshold which, in this case, comes under the form as the minimum percentage of overlap above which the finding will be considered detected. As to the computation of the performance indexes, they are performed in a similar fashion to the microlocalifications. There might be the case where there are in fact detected masses in the OA but none of them overlap with any mass in the GT, in which case the number of false positives is given by the number of findings in the OA, and the number of missed lesions is given by the number of lesions in the GT.

3.3.1 Evaluation of Mass Segmentation Algorithms

In regard to masses, this work focuses more on segmentation rather than detection, so this section will discuss the metrics to assess the quality of boundary extraction of masses.

At first, it becomes evident that the overlap measure defined in Equation 3.1 may also be seen as an performance index. This because it expresses the percentage of overlap between the mass segmented by the algorithm and the mass present in the ground truth.

Also, a another way to evaluate the quality of the segmentation is to evaluate the distance between the contours which is done by using two different kinds of metrics: Hausdorff distance and Average distance. The Hausdorff distance between contour I and J can be seen as the distance between the furthest point in I and the point of J that is closest to I . Formally it is defined by:

$$H(I, J) = \max(h(I, J), h(J, I))$$

$$h(I, J) = \max_{i \in I} \min_{j \in J} ||i - j||$$

The advantage of this metric is that it highlights the worst cases whereas in the Average metric any big difference in the contour has the same “weight” in the end result.

Chapter 4

Mass Segmentation Method

This chapter is concerned with the introduction of a new method of mass segmentation. It was based on a method used to detect the pectoral muscle in mammograms [37]. The base algorithm will be described in section one and section two deals with the evaluation where the procedure and the results obtained will be presented.

4.1 Base Algorithm

The segmentation algorithm was applied not to the whole image, but to a manually selected region of interest (ROI) of size 1024x1024 that contains a single mass. Figure 4.1 shows a ROI selected from the database that will be used to explain the segmentation algorithm.

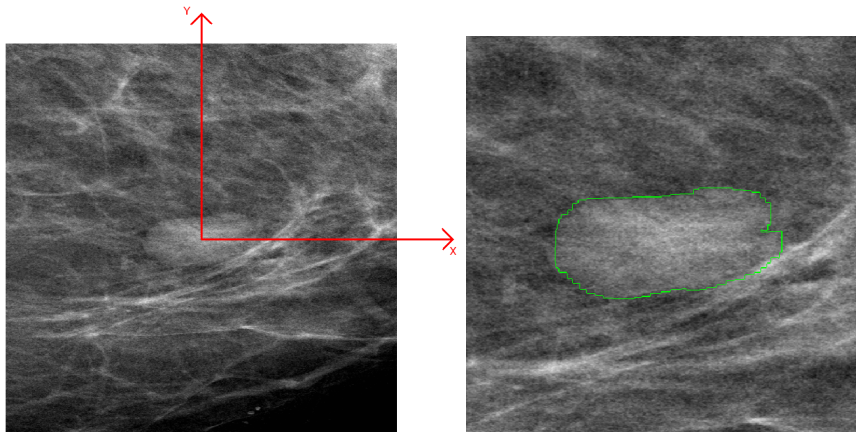


Figure 4.1: ROI (left) Detail of the manually segmented contour of the mass (right)

Masses are usually round structures which most prominent characteristic is the discontinuity of the gray levels between one side and the other of the border. The algorithm explores this characteristic: the main idea is that the border corresponds to a closed path along the edge pixels

and that if the edge pixels are assigned the appropriate weight, the border will correspond to the shortest (in the sense of minimum-cost) path. The main steps of the algorithm are bellow:

1. **Transform into polar coordinates**
2. **Compute the gradient**
3. **Compute the shortest path**
4. **Transform back into Cartesian coordinates**

Since masses are closed round structures, if we were to find the shortest path in the Cartesian coordinates we had to impose to the algorithm to find a round path along the mass. However, by transforming the image in polar coordinates, the mass is being “extended” or “stretched” along a straight line (intuitively speaking) and therefore the edges will fall from top to bottom of the image. This in turn makes it easier to find the shortest path since the computation of the costs to be assigned to the edges is simpler. But to achieve this, it is essential that the origin of the polar coordinates system falls inside the mass. This is done by choosing a ROI whose center is inside the mass and choosing the center as the origin of the coordinates. One must also make sure that the whole mass is inside the ROI.

Figure 4.2 shows the result of the transformation into polar coordinates. Note that on the left side of the image it is clearly visible a region where both texture and gray-level characteristics differ from the other parts of the image. Also, on the right side of the image, the black “saw-tooth” like structures, are a result of computing the polar coordinate image with the origin at the centre. The “valleys” correspond to the corners of the image in Cartesian coordinates and the top of the structures correspond to the points of the image where the angle is a multiple integer of $\pi/2rads$.

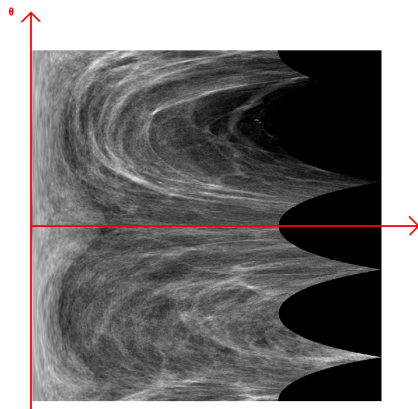


Figure 4.2: ROI in polar coordinates

The next step of the algorithm is to compute the gradient of the image in polar coordinates to highlight the edge pixels. The method used was the Prewitt operator where the gradient in the

x-direction is approximated by equation 4.1 and the gradient in the y-direction by equation 4.2. The magnitude of the gradient is computed using equation 4.3 where $*$ represents the convolution and I the original image. The result is shown on Figure 4.3.

$$G_x = \begin{bmatrix} +1 & +1 & +1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} * I \quad (4.1)$$

$$G_y = \begin{bmatrix} +1 & 0 & -1 \\ +1 & 0 & -1 \\ +1 & 0 & -1 \end{bmatrix} * I \quad (4.2)$$

$$g = \sqrt{G_x^2 + G_y^2} \quad (4.3)$$

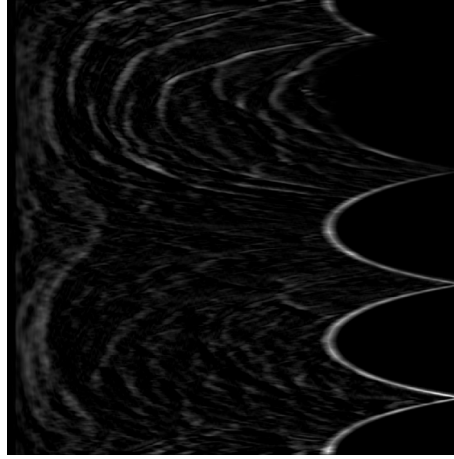


Figure 4.3: Gradient of the ROI in polar coordinates

As expected, there is a strong gradient caused by the structures in the right part of the image. In order to remove them, a set of similar structures was created, and dilated and then used to mask this image. The result of these operations can be seen in Figure 4.4. The image might look a little brighter because, since the brighter parts were removed, the brightness of the remaining areas were adjusted for visualization purposes.

It is from this image, that the algorithm computes the shortest path between the top and bottom rows. The algorithm first assigns costs to the pixels and then works backwards on finding the path. The key aspect of the method is that the costs are assigned in order to favour the path through pixels with high gradient. Most of the times (just like in the example) a mass is not round and so, in polar coordinates, the contour will not correspond to an approximately straight line from top to bottom. To safeguard situations like this, the cost assignment must be such that attributes similar weights to variations to the right and left and allows more than one pixel in each row to be part of the connected path from top to bottom. Also, since the mass has a closed contour, one must

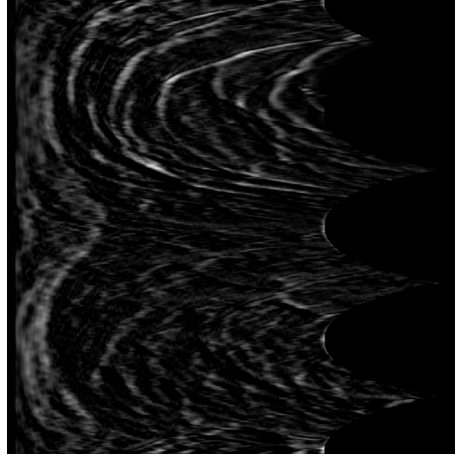


Figure 4.4: Gradient of the ROI in polar coordinates after post-processing

ensure that the contour's start and end point coincide, which in polar coordinates mean the starting and end column of the contour must be the same. This was done by computing the shortest-path starting and ending at the same column for every column of the image. To force that the starting and end columns were the same, the infinite value was assigned to the initial weights of columns that were not being considered. At the end of each calculation for each column the contour and its cost was saved. The shortest path was the one with the contour that had the smaller cost. These three were the major changes to the shortest path algorithm, for a deeper understanding of the algorithm consult [37]. Figure 4.5 shows the shortest path between top and bottom rows.

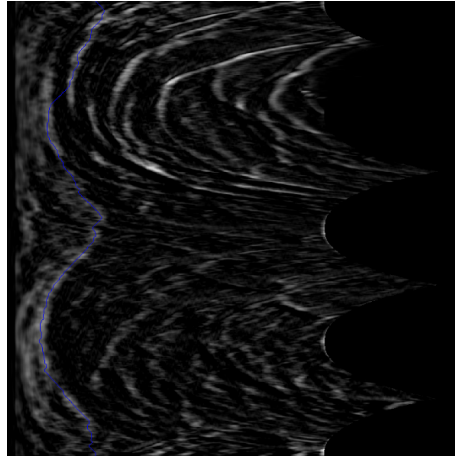


Figure 4.5: Shortest path of the ROI in polar coordinates

The last step consists on transforming the contour in polar coordinates back into Cartesian coordinates. However, after this step the resulting path may not be a single connected component since some approximations are done during the conversion. In those cases, a series of morphological dilation and eroding operations are applied to the contour to connect the gaps. The final result

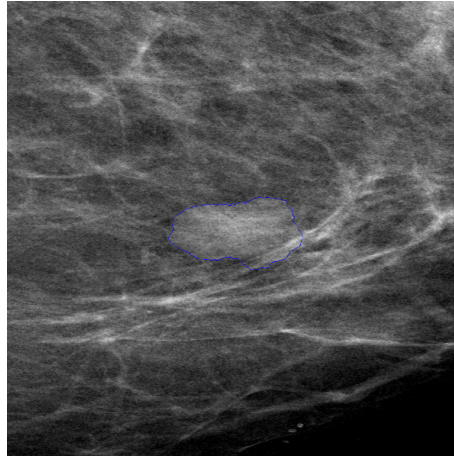


Figure 4.6: Contour of the mass in Cartesian coordinates

of the segmentation is shown on Figure 4.6.

4.2 Evaluation

4.2.1 Procedure

The segmentation algorithm was tested on the total number of individual masses present in the HSJ database (fourty) using three different images. The original ROI, without any enhancement method, an enhanced image and an image created with textural components.

The enhancement method is an adaptation of the method presented in [38]. First Contrast Limited Adaptive Histogram Equalization (CLAHE) applied to the whole image is used to remove noise and enhance mammographic images. Next, a custom algorithm is used to provide non-linear enhancement. First we perform non-linear adjustments based on the image gray level statistics. Then the mean value image is computed using 16 by 16 blocks (size determined experimentally). Finally the image is normalized, thanks to which image pixels are in range [0, 1] and the ROI is extracted.

As to the texture image, it was obtained by dividing the enhanced image into blocks of size 16x16 and 12 texture features [39] are extracted from each block. These features are then averaged, forming one value for each block, which creates a new image, where masses are roughly segmented from the background tissue.

The results of these two evaluation techniques were assessed using the metrics described in section 3.3.1. Note that the distance metrics were normalized in by the size of the diagonal of the ROI ($d = \sqrt{1024^2 + 1024^2}$).

Table 4.1: Mean values of the distances and overlap - Shortest-Path Method

	Original	Enhanced	Texture
Hausdorff	0.118740	0.088359	0.151650
Average	0.023078	0.018113	0.043024
Overlap	0.6214	0.7074	0.4227

4.2.2 Results

The mean values of all the results obtained for each image are summarized In Table 4.1. It is evident that the enhanced image obtains the best results in all metrics. As the shortest-path method takes more advantage of the gray-level characteristics of masses rather than texture, it works substantially better in original and enhanced images which explains the poor performance in the texture images. Table 4.2 shows the performance obtained by the detection method in the SPACS project (a region growing algorithm). The shortest-path algorithm obtained significantly better performance than standard region-growing approach. In addition, the region membership criterion needs to be defined manually. A case were the segmentation obtained good results is shown in Figure 4.7 and Figure 4.8 shows an example where the results obtained were unsatisfactory. These situations occur mainly because the gradient isn't strong enough to be considered as an edge. This may be due to normal breast tissue that is overlapping with the mass or poor contrast between the breast tissue and the mass which causes its borders to be fuzzy.

Table 4.2: Mean values of the distances and overlap - Region-growing algorithm

	Original	Enhanced	Texture
Hausdorff	0.120790	0.110806	0.107691
Average	0.036702	0.038660	0.042905
Overlap	0.321777	0.465372	0.323223

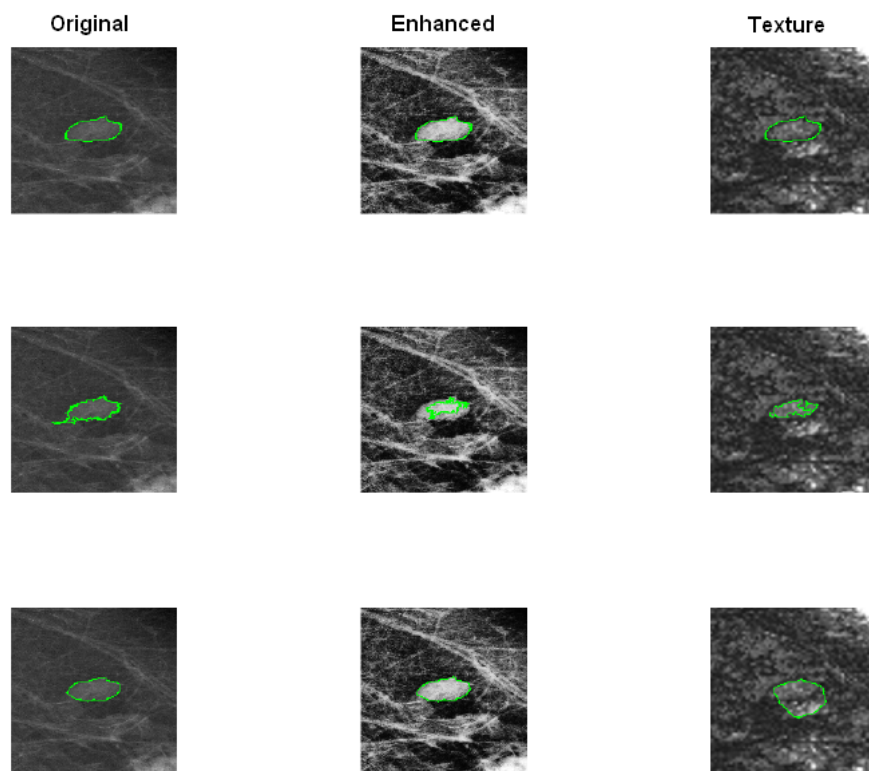


Figure 4.7: Example of a segmentation with good results. Top Row: Ground Truth Annotation; Middle Row: Region Growing Algorithm; Bottom Row: Shortest-Path Method

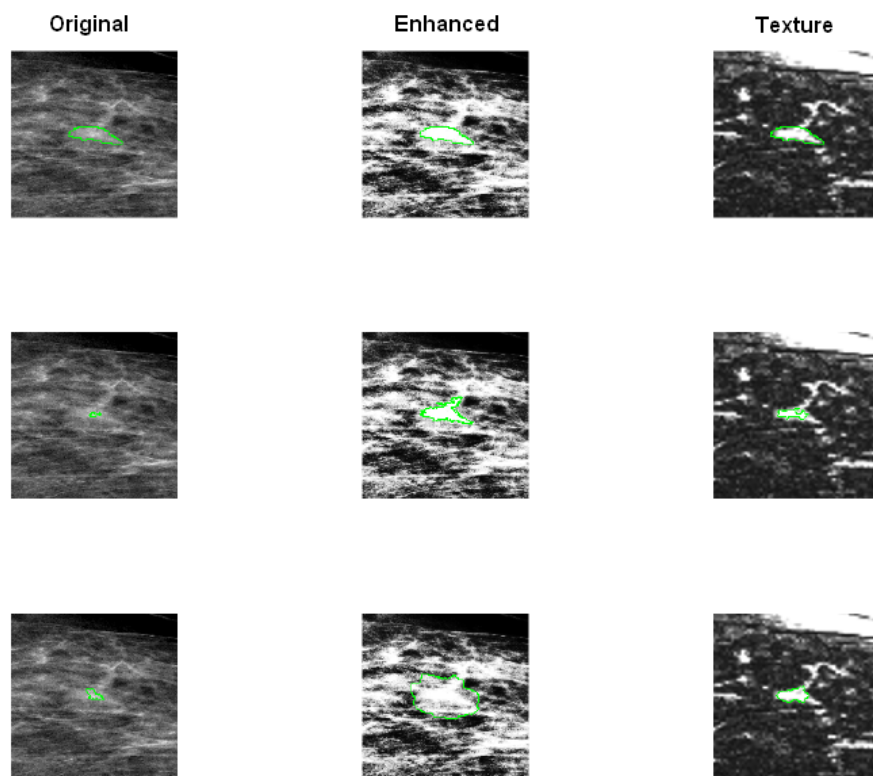


Figure 4.8: Example of a segmentation with poor results. Top Row: Ground Truth Annotation; Middle Row: Region Growing Algorithm; Bottom Row: Shortest-Path Method

Chapter 5

Microcalcifications Detection

This chapter deals with the implementation of a microcalcification detection method [40] in order to compare with the current SPACS microcalcification detection method. Section one briefly summarizes the SPACS method, section two will describe the implementation of a different microcalcification detection algorithm and finally section three describes the procedure and results obtained.

5.1 SPACS method

The current micro-calcification detection algorithm consists in three stages: pre-processing, detection, and classification.

Image enhancement is preceded by a cropping operation on the original image. The image enhancement method is based on the phase preserving de-noising algorithm proposed by Peter Kovesi [41]. This method differs from standard wavelet de-noising techniques in that it uses non-orthogonal log-Gabor wavelets and, unlike other existing techniques, ensures that phase information is preserved in the image.

In the second step, the method uses global thresholding to detect potential micro-calcifications. In the output of that stage, besides the true micro-calcifications, a high number of false positives and noise are also present. The last stage begins with the extraction of 94 features for each suspicious block of size 16x16 that contain a suspicious micro-calcification in the center. Of the original 94 features, 90 come from the textural domain (Haralick Texture features, Gray Level Run Lengths and Laws Texture Energy Measures) and 4 from the spatial domain (skewness, kurtosis, standard variation and mean gray level).

In the classification step, Support Vector Machines are being used. The training set consists in information from 136 mammography images (104 malignant and 32 normal).

5.2 Implemented method

The detection method that was implemented takes advantage of the fact that microcalcifications are most of the times brighter than their surroundings. It first applies a series of image enhancement methods to improve the contrast and then uses a thresholding technique to segment them. This method has comprised 4 stages: Noise removal; computation of a band-pass filtering and the Laplacian in parallel; multiplication of these images and finally thresholding.

Before the noise removal operation, the image was cropped and resized for reasons that will be explained further. The cropping operation just removed the part of the mammogram that did not contain any information and after the image was resized to 1024x1024. The noise removal was implemented by applying a low pass filter whose mask is given by 5.1

$$M = \frac{1}{15} \times \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 2 & 1 \end{bmatrix} \quad (5.1)$$

The values of the elements of the mask were assigned depending on the distance to the central pixel and are inversely proportional to the distance to the central pixel. This filter smooths image preserving at the same time many of the important detail. This was also taken into account when choosing the size of the mask since the bigger the mask, the bigger the smoothing.

The next step consists in normalizing the histogram. This operation increases the contrast of the image by spreading a designated range of the histogram along all intensity levels. That area right of the point A is computed automatically and corresponds to 0.5% of the relative frequency of the maximum histogram value. The normalized image I_N is given by equation 5.2:

$$I_N = \frac{I_O - A}{MAX - A} \times MAX \quad (5.2)$$

This normalized image is then filtered by a band-pass filter in the frequency domain. A butterworth filter of order five is applied to the image in order to cut the very low and high frequency and highlight the rest. The image is then transformed back to spatial domain through the inverse FFT. However, these operations are computationally expensive to be applied to the whole image reason for which the image was cropped and resized in the first step.

In parallel with these operations, the Laplacian of the filtered image was computed. The Laplacian, being a second-order derivative operator, enhances fine detail of the image. Therefore, the microcalcifications will be highlighted as well as inevitably some fibres and glandular tissue of breast. The Laplacian image as well as the filtered image have in common the set of detected microcalcifications. In this step they are going to be multiplied in order to highlight what is common to both images and to attenuate the remaining structures.

Once the microcalcifications are highlighted, it is possible to segment them by thresholding the image.

Table 5.1: Detection results comparisson: mean values of the number of false positives per image, missed detections and missed rate

	False positives	Missed detections	Missed rate (%)
Implemented Algorithm	78.1028	9.7714	47.8269
SPACS	32.5721	6.4278	38.1198

The choice of this threshold is done empirically and when increased it detects less findings (low sensitivity); when decreased it increases it's sensitivity at the expense of a greater number of false positives.

5.3 Evaluation

5.3.1 Procedure

The detection algorithm was tested on 208 mammograms of the HSJ database containing microcalcifications only. It was then evaluated using the procedure described in chapter three. The detection threshold used corresponded to the width of the biggest microcacification present in the images which was determined to be 15. The results were only compared with the output of the SPACS detection stage since the output of the whole method is not comparable for it uses SVM's to reduce the number of false positives.

5.3.2 Results

The results obtained by the SPACS and implemented algorithm are summarized in Table 5.1. The main idea of both of these methods is to apply enhancement techniques and then a global threshold. Even though this kind of approach has several advantages like simplicity, ease of implementation and efficiency (which is desirable for real-time clinical applications), this also comes at the expensive of reduced effectiveness in some cases. Indeed the results obtained in this case are unsatisfactory since they have a high number of false positives per image as well as a high missed rate. This is however the result of the detection stage, most of the detection approaches use posterior machine learing methods to reduce the number of false positives. Figure 5.1 shows a case where the implemented method obtained good and poor results. The results obtained by the SPACS detection algorithm are shown in Figure 5.2. In cases such as those in Figure 5.3, these methods detect a high number of calcifications in the pectoral muscle (which are false positives) and as well as the calcifications in the vessels. The latter are not relevant since they are part of the normal ageing process and do not indicate the presence of cancer. These situations also help explaining the high rate of false positives obtained by the algorithms.

obtain a high number of false positives

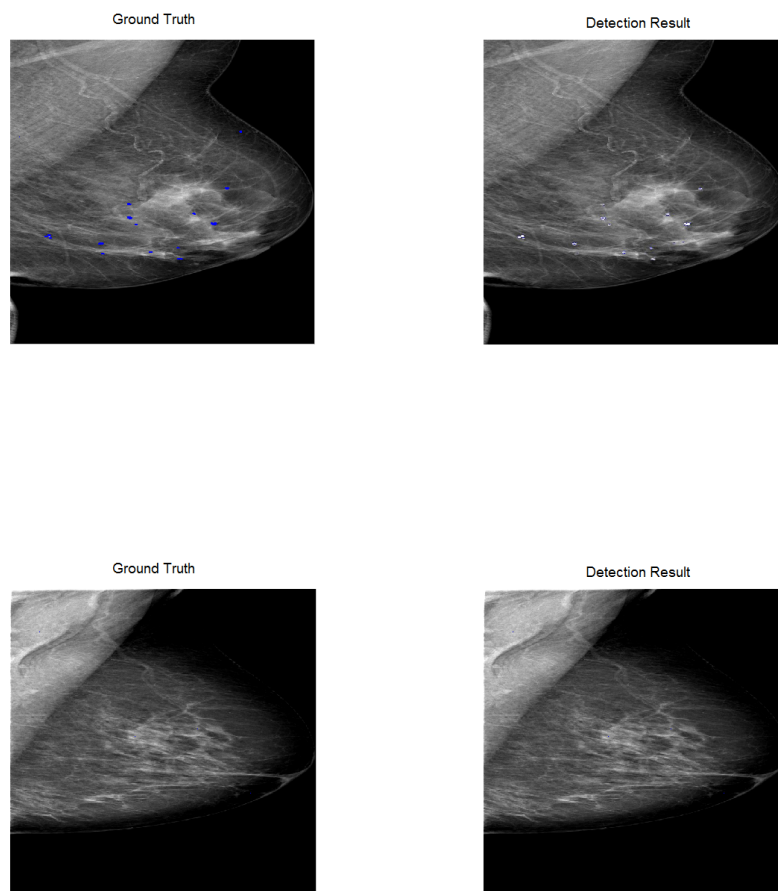


Figure 5.1: Example of a detection with good results (top) and poor results (bottom) - Implemented Algorithm

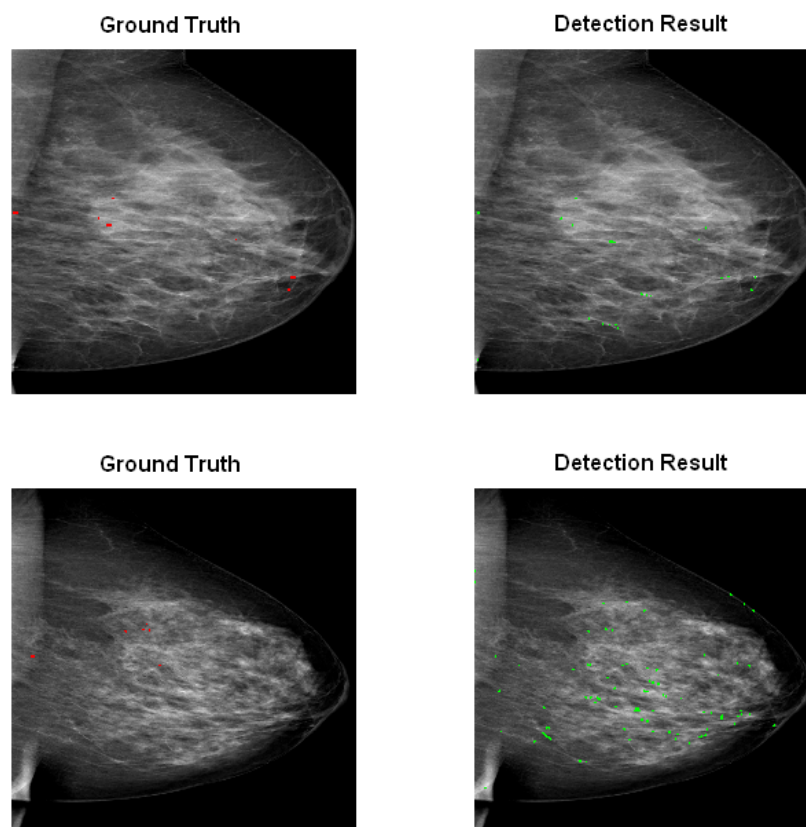


Figure 5.2: Example of a detection with good results (top) and poor results (bottom) - SPACS

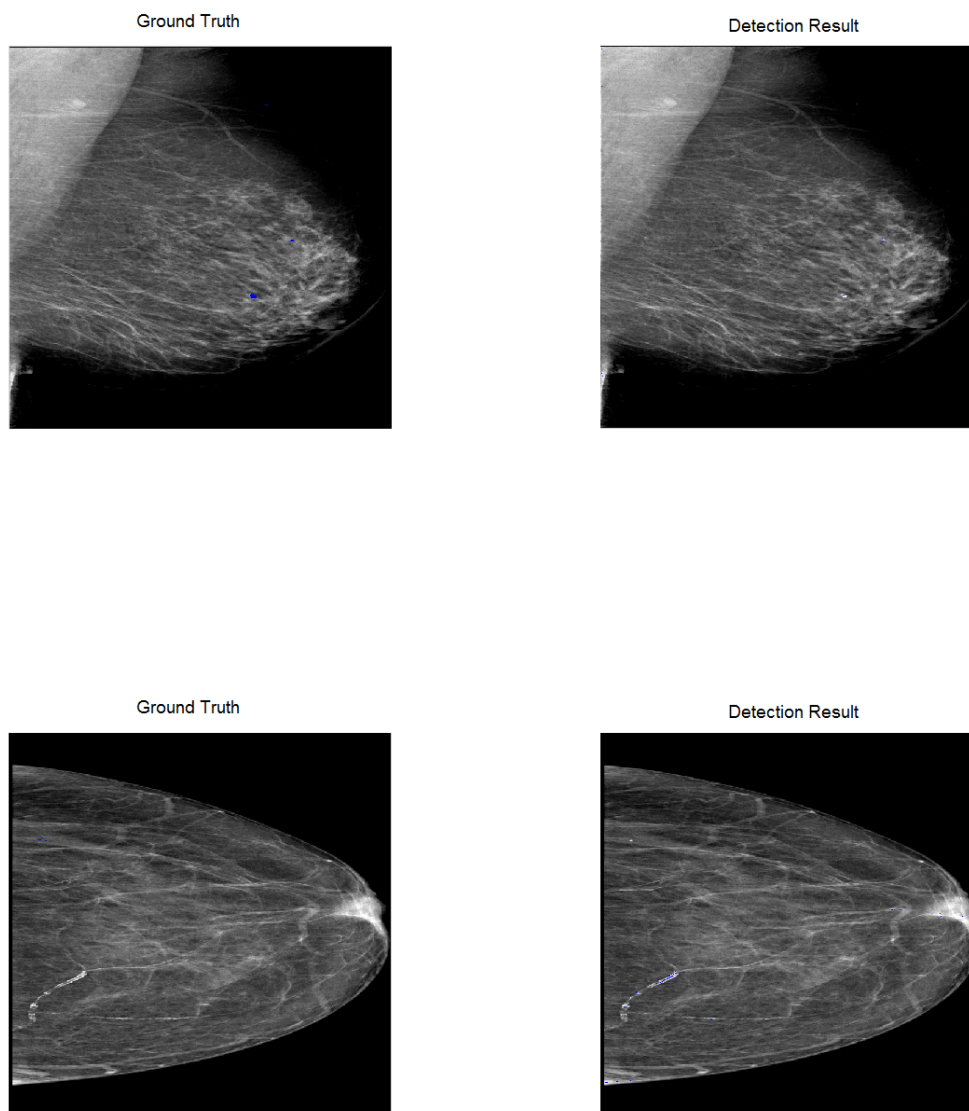


Figure 5.3: Example of a detection with false positives - Implemented Algorithm

Chapter 6

Conclusion and Future Work

This chapter presents the contributions and suggests some future work to be done as a result of this thesis.

6.1 Contributions

This work has successfully introduced a new framework to evaluate the performance of microcalcifications or mass detection algorithms that are robust rigorous and predictable. These metrics permitted a better evaluation of the work developed within the SPACS project and hopefully will contribute to a better uniformisation of the performance evaluation metrics used within the medical image processing community.

The proposed mass segmentation method obtained some encouraging results (70% overlap with the enhanced version), since according to the literature in [16], there are few contour-based methods due to the difficulty of extracting the boundary between masses and normal tissue. Also, within the SPACS project, this method was a significant improvement relative to the region growing algorithm since it obtained an additional 20% of overlap area in the enhanced version of the ROI.

As to the microcalcifications, the implemented approach was compared to the current detection algorithm which highlighted some limitations of both approaches and allowed to devise some directions in the future work.

6.2 Future Work

Some future work can be done regarding the segmentation of masses by computing different types of gradient, use texture features or any other features that better emphasizes the border between mass and regular tissue. The algorithm to find the shortest path may also need to be

improved namely in the assignment of the weights to the pixels. This method, however is semi-automatic since the ROI has to be selected manually. The future work can also consist in devising a fully automated method.

The future work regarding microcalcification detection scheme consists in improving the enhancement algorithms as well as the thresholding technique by using a adaptive thresholding technique rather than global.

Bibliography

- [1] W. H. Organization, “Fact sheet no. 297: Cancer,” February 2009. Available online at <http://www.who.int/mediacentre/factsheets/fs297/en/>.
- [2] Eurostat, “Health statistics – atlas on mortality in the european union,” June 2009. Available online at http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-30-08-357/EN/KS-30-08-357-EN.PDF.
- [3] I. A. for Research on Cancer, “World cancer report 2008,” December 2009. Available online at <http://www.iarc.fr/en/publications/pdfs-online/wcr/2008/index.php>.
- [4] D. G, V. H. S, L. AD, and E. M. Murray CJ, “Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors.,” November 2005.
- [5] N. M. Gøtzsche PC, “Screening for breast cancer with mammography.,” October 2006.
- [6] U. N. C. Institute, “Mammogram fact sheet,” August 2009. <http://www.cancer.gov/cancertopics/factsheet/Detection/mammograms>.
- [7] J. G. Elmore, C. K. Wells, C. H. Lee, D. H. Howard, and A. R. Feinstein, “Variability in radiologists’ interpretations of mammograms,” *New Eng. J. Med.*, vol. 331, pp. 1493–1499, 1994., 1994.
- [8] D. B. Kopans, “The positive predictive value of mammography,” *AJR. Am. J. Roentgenol.* 158, 521–526 (1992)., 1992.
- [9] T. Freer and M. Ulissey, “Screening mammography with computer-aided detection: prospective study of 12860 patients in a community breast center,” *Radiology* 220 (2001), pp. 781–786., 2004.
- [10] P. Skaane and A. Skjennald, “Screen-film mammography versus full-field digital mammography with soft-copy reading: Randomized trial in a population-based screening program—the oslo ii study,” *Radiology*, 232, 197-204. (2004), 2004.
- [11] C. Kuzmiak, G. Millnamow, B. Qaqish, E. Pisano, E. Cole, and M. Brown, “Comparison of full-field digital mammography to screen-film mammography with respect to diagnostic accuracy of lesion characterization in breast tissue biopsy specimens,” *Acad. Radiol.* 9 (2002), pp. 1378–1382., 2002.
- [12] “Ddsm: Digital database for screening mammography home page.” <http://marathon.csee.usf.edu/Mammography/Database.html>.

- [13] "The mini-mias database of mammograms home page." <http://peipa.essex.ac.uk/info/mias.html>.
- [14] A. C. of Radiology (ACR), "Standards for the performance of screening mammography [adopted by the acr council 1990, revised 1994]," 1994.
- [15] U. Fischer, F. Baum, and S. Luftner-Nagel, "Breast imaging, direct diagnosis in radiology," 2009.
- [16] A. Olivera, J. Freixeneta, J. Martí, E. Pérezb, J. Pontb, E. R.E. Dentonc, and R. Zwiggelaar, "A review of automatic mass detection and segmentation in mammographic images," *Medical Image Analysis* 14 (2010) 87–110, 2009.
- [17] A. C. of Radiology, *Breast Imaging Reporting and Data System (BI-RADS®)*. Reston, VA: American College of Radiology, 2003, Fourth ed., 2003.
- [18] R. Rangayyan, F. Ayres, and J. Desautels, "A review of computer-aided diagnosis of breast cancer: toward the detection of subtle signs," *Journal of the Franklin Institute Volume 344, Issues 3-4, May-July 2007*, 2007.
- [19] T. Fawcett, "An introduction to roc analysis," *Pattern Recognition Letters* 27 (2006) 861–874, 2006.
- [20] S. Feig, B. Galkin, and H. Muir, "Evaluation of breast microcalcifications by means of optically magnified tissue specimen radiographs," *S. Brunner, B. Langfeldt (Eds.), Recent Results in Cancer Berlin, Germany, 1987*, 1987.
- [21] E. Sickles, "Breast calcifications: mammographic evaluation," *Radiology* 160 (1986) 289–293, 1986.
- [22] L. Shen, R. Rangayyan, and J. Desautels, "Detection and classification of mammographic calcifications," *Int. J. Pattern Recogn. Artif. Intell.* 7 (1993), 1993.
- [23] S. Yu and L. Guan, "A cad system for the automatic detection of clustered microcalcifications in digitized mammogram films," *IEEE Trans. Med. Imaging* 19 (2) (2000) 115–126, 2000.
- [24] W. Qian, M. Kallergi, L. P. Clarke, H. D. Li, P. Venugopal, D. Song, and R. A. Clark, "Tree structured wavelet transform segmentation of micro-calcifications in digital mammography," *Med. Phys.* 228, 1247–1254 (1995), 1995.
- [25] L. Wei, Y. Yang, R. Nishikawa, M. Vernick, and A. Edwards, "Relevance vector machine for automatic detection of clustered microcalcifications," *IEEE Trans. Med. Imaging* 24 (10) (2005) 1278–1285, 2005.
- [26] S.-N. Yua, K.-Y. Lib, and Y.-K. Huangac, "Detection of microcalcifications in digital mammograms using wavelet filter and markov random field model," *Comput. Med. Imaging Graphics* 30 (2006) 163–173, 2006.
- [27] N. Mudigonda, R. Rangayyan, and J. Desautels, "Detection of breast masses in mammograms by density slicing and texture flow-field analysis," *IEEE Trans. Med. Imaging* 20 (12) (2001) 1215–1227, 2001.
- [28] S. Liu, C. Babbs, and E. Delp, "Multiresolution detection of spiculated lesions in digital mammograms," *IEEE Trans. Image Process.* 10 (6) (2001) 874–884, 2001.

- [29] L. Zheng and A. Chan, "An artificial intelligent algorithm for tumor detection in screening mammogram," *IEEE Trans. Med. Imaging* 20 (7) (2001) 559–567, 2001.
- [30] R. Zwiggelaar, T. Parr, I. Hutt, C. Taylor, S. Astley, and C. Boggis, "Model-based detection of spiculated lesions in mammograms," *Med. Image Anal.* 3 (1999) (1), 1999.
- [31] N. El-Faramawy, J. Desautels, O. Alim, and R. Rangayyan, "Measures of acutance and shape for classification of breast tumors," *IEEE Trans. Med. Imaging* 16 (6) (1997) 799–810, 1997.
- [32] H. Zhang, S. W. Fool, S. Krishnan, and C. H. Thng, "Automated breast masses segmentation in digitized mammograms," *2004 IEEE International Workshop on Biomedical Circuits & Systems*, 1997.
- [33] T. O. Gulsrud, K. Engan, and T. Hanstveit, "Watershed segmentation of detected masses in digital mammograms," *Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference Shanghai, China, September 1-4, 2005*, 2005.
- [34] Y. Yuan, M. L. Giger, H. Li, K. Suzuki, and C. Sennett, "A dual stage method for lesion segmentation on digital mammograms," *IEEE Trans. Med. Imag.* 26 (6) (2007), 2007.
- [35] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Prentice Hall, Second ed., 2002.
- [36] G. Liu and R. M. Haralick, "Optimal matching problem in detection and recognition of performance evaluation," *Pattern Recognition vol 35, Issue 10, October 2002*, pp. 2125-2139, 2002.
- [37] J. S. Cardoso, I. Domingues, I. Amaral, I. Moreira, P. Passarinho, J. S. Comba, R. Correia, and M. J. Cardoso, "Pectoral muscle detection in mammograms based on polar coordinates and the shortest path," *32nd Annual International Conference of the IEEE EMBS Buenos Aires, Argentina, August 31 - September 4, 2010*, 2010.
- [38] J. E. Ball and L. M. Bruce, "Digital mammographic computer aided diagnosis (cad) using adaptive level set segmentation," *Proceedings*, pp. 4973-4978, 2007, 2007.
- [39] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol.3, n.6, pp.610-621, 1973, 1973.
- [40] M. J. Q. M. dos Santos Abreu, "Implementação e teste de métodos de segmentação de lesões em imagens de mamografia digital," Master's thesis, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, 2010.
- [41] P. Kovesi, "Phase preserving denoising of images," *Australian Pattern Recognition Society Conference (DICTA), Perth WA*, pp.212-217, 1999, 1999.