Universitat de Girona

PhD Thesis

Computer-aided lesion detection and segmentation on breast ultrasound images

Gerard Pons Rodríguez

2013

Universitat de Girona

PhD Thesis

Computer-aided lesion detection and segmentation on breast ultrasound images

Gerard Pons Rodríguez

2013

Doctoral Programme in Technology

Supervised by:

Joan Martí and Robert Martí

Work submitted to the University of Girona in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Dr. Joan Martí and Dr. Robert Martí from Universitat de Girona.

DECLARE

That the work entitled *Computer-aided lesion detection and segmentation on breast ultrasound images* presented by *Gerard Pons Rodríguez* to obtain the degree in Doctor of Philosophy has been developed under our supervision and complies with the requirements needed to obtain the International Mention.

Therefore, in order to certify the aforesaid statement, we sign this document.

Girona, December 2013.

Agraïments

Doncs sí, al final he escrit una tesi doctoral. Si un llunyà dia de 2004 m'haguessin dit que m'acabaria passant gairebé 10 anys a la universitat i escriuria una tesi no m'ho hauria cregut. Recordo que en aquella època la meva idea era acabar l'enginyeria tècnica i posar-me a treballar. Però el teu camí a la vida difícilment succeeix com el tens planejat.

Aquesta tesi comprèn tres anys de feina però podria dir que és el resultat del llarg camí recorregut des d'aquell llunyà dia de 2004. Un camí on he coincidit i conegut a molta gent, la majoria de la qual ha contribuït, en major o menor part, en el fet que ara tinguis aquesta tesi entre les teves mans.

Per començar, vull donar les gràcies als meus pares, Esteve i Pilar, i a la meva germana Sandra per estimar-me tant i fer-me com sóc. Per guiar-me, aconsellar-me i renyar-me quan tocava. Per dir-me no quan era necessari, per tal que valorés el cost de les coses. Gràcies.

També vull donar les gràcies a l'Agustí, que tot i els seus rotllos i opinions a contracorrent és la millor persona que podria tenir com a pare dels meus nebots, l'Aina i l'Arnau, a qui m'estimo amb bogeria.

A l'Eli, element central de la meva vida. Per estimar-me tal com sóc. Per estar sempre al meu costat i recolzar-me en situacions difícils, com les estades a l'estranger. Per fer-me feliç.

Durant aquest camí he tingut la sort de coincidir amb un grup de persones fantàstiques. Hem compartit el mateix camí junts, ens hem ajudat uns als altres i ens ho hem passat molt bé. Hem fet pràctiques junts, hem discutit a crits a la biblioteca a hores intempestives, hem viatjat, hem compartit moltes nits de dijous, moltes cerveses i més cafès. Parlo d'en Ricard, en Pablo, en Mariano, en Gubern i en Pla. Estic segur que sense haver-los conegut aquest camí no hauria estat tan llarg.

Tampoc em vull oblidar d'en Massi i l'Enric, que tot i aparèixer més tard en aquest

camí que he recorregut, s'han convertit en una peça clau dins d'aquest grup d'amics. Ni de l'Eloy i en Pepe, amb qui he compartit llargues estones als cafès de les 10 i de les 3.

Evidentment, he de fer una especial menció als meus supervisors Joan i Robert. Gràcies per ajudar-me, guiar-me, animar-me i per tenir sempre la porta oberta del despatx per atendre els meus dubtes i problemes.

Una part important d'aquest camí també és ON ha transcorregut. VICOROB, un grup de recerca ple de persones entranyables, on no hi trobes companys de feina sinó amics. Gràcies a tots els que en formeu part per fer un entorn laboral tan agradable i inoblidable. En especial a en Freixenet, a en Lladó i a l'Arnau, per ser els primers que van introduir el concepte "recerca" dins del meu cap, en aquells llunyans projectes fi de carrera. A la Meritxell, a en Bertu i a en Sik, per solucionar-me molts dubtes i ajudar-me quan estava encallat. I a la Joseta i a la Mireia, per estar sempre disponibles i ajudar-me amb qualsevol dubte que he tingut relacionat amb la burocràcia.

També vull donar les gràcies als doctors Kai Vilanova i Melcior Sentís per proporcionar-me les dades necessàries per realitzar aquesta tesi. I en especial al Dr. Sergi Ganau, pel seu compromís i les hores que ha dedicat a segmentar lesions i barallar-se amb l'ecògraf per obtenir les imatges d'elastografia utilitzades en els meus experiments.

Finally, this PhD thesis was enriched thanks to two international research stays. I would like to thank Prof. Alison Noble from the University of Oxford and Dr. Moi Hoon Yap from the Manchester Metropolitan University to supervise me and especially for their guidance during the development the work performed in such stays.

Gràcies a tots.

Publications

Journals

- **[ULTRAS 2013]** G. Pons, M.H. Yap, J. Martí, S. Ganau, M. Sentís and R. Martí. "A Comparative Study on the Performance of Breast Ultrasound Lesion Detection Algorithms". Ultrasonics, submitted. Sep. 2013.

- **[UMB 2013]** G. Pons, R. Martí, S. Ganau, M. Sentís and J. Martí. "Computerized Detection of Breast Lesions using Deformable Part Models in Ultrasound Images". Ultrasound in Medicine and Biology, submitted. Oct. 2013.

- **[JUM 2013]** G. Pons, J. Martí, R. Martí, S. Ganau, J. C. Vilanova and J. A. Noble. "Evaluating Lesion Segmentation on Breast Sonography as Related to Lesion Type". Journal of Ultrasound in Medicine. 32(9):1659-1670, 2013.

Conferences

- **[IbPRIA 2013]** G. Pons, R. Martí, S. Ganau, M. Sentís and J. Martí. "Feasibility Study of Lesion Detection Using Deformable Part Models in Breast Ultrasound Images". Iberian Conference on Pattern Recognition and Image Analysis. LNCS 7887, pp. 269-276. Madeira, Portugal, June 2013.

- **[ECR 2013]** G. Pons, J. Massich, R. Martí, S. Ganau, M. Sentís, J. Fernandez-Bayo, R. Aguilar, J.C. Vilanova and J. Martí. "Computer-aid lesion detection and segmentation in breast ultrasound images". doi: 10.1594/ecr2013/C-1303. European Congress of Radiology (ECR). Vienna, Austria, March 2013.

- **[SPIE 2012]** G. Pons, J. Martí, R. Martí, M. Cabezas, A. Di Battista and J.A. Noble. "Lesion Segmentation and Bias Correction in Breast Ultrasound B-mode

Images Including Elastography Information". SPIE Conference on Medical Imaging, 8314, pp. 83141E1-1E6. San Diego, California, USA, February 2012.

- **[IbPRIA 2011]** G. Pons, R. Martí and J. Martí. "Simultaneous Lesion Segmentation and Bias Correction in Breast Ultrasound Images". Iberian Conference on Pattern Recognition and Image Analysis. LNCS 6669, pp. 692-699. Las Palmas de Gran Canaria, Spain, June 2011.

List of Acronyms

ABUS Automated Breast Ultrasound 87

ACM Active Contour Models 26

ADE Average Detection Error 20

AMED Average Minimum Euclidian Distance 31

AO Area Overlap 28

ARD Average Radial Derivative 32

ARE Average Radial Error 30

BNN Bayesian Neural Network 24

BIRADS Breast Imaging and Reporting Data System 117

BSE Breast Self-Examination 2

BUS Breast Ultrasound 9

LLIP Local Lesion Information at the Partial image 97

LLSI Local Lesion Surrounding Information 97

MAP Maximum A Posteriori 23

MD Minimum Distance 30

MLO Medio-Lateral Oblique 3

MRF Markov Random Fields 26

MRI Magnetic Resonance Image 4

NLF Non-lesion Location Fraction 59

NN Neural Network 36

NRV Normalized Residual Value 29

RGI Radial Gradient Index 20

PBT Probabilistic Boosting Tree 22

PD Proportional Distance 31

PDF Probability Density Function 94

SWE Shear-Wave Elastography 115

TN True Negative 26

TP True Positive 26

TPR True-Positive Rate 21

US Ultrasound 4

USE Compression Ultrasound Elastography 115

List of Figures

List of Tables

Contents

Abstract

This thesis deals with the detection, segmentation and classification of lesions on sonography. The presence of these lesions is a sign of breast cancer. Hence, the earlier these lesions are detected, the better and more effective the treatment will be. The contribution of the thesis is the development of a new Computer-Aided Diagnosis (CAD) framework capable of detecting, segmenting, and classifying breast abnormalities on sonography automatically. Firstly, an adaption of a generic object detection method, Deformable Part Models (DPM), to detect lesions in sonography is proposed. The method uses a machine learning technique to learn a model based on Histogram of Oriented Gradients (HOG). This method is also used to detect cancer lesions directly in a multi-class detector, simplifying the traditional cancer detection pipeline (candidate lesion detection, segmentation, feature extraction and classification between benign and malignant). Secondly, different initialization proposals by means of reducing the human interaction in a lesion segmentation algorithm based on Markov Random Field (MRF)-Maximum A Posteriori (MAP) framework is presented. Furthermore, an analysis of the influence of lesion type in the segmentation results is performed. Finally, the inclusion of elastography information in this segmentation framework is proposed, by means of modifying the algorithm to incorporate a bivariant formulation. The proposed methods in the different stages of the CAD framework are assessed using different datasets. The evaluation, carried out in a quantitative and qualitative manner, uses several metrics for detection and segmentation, is performed for each stage independently, and is compared with the most relevant methods in the state-of-the-art. The analysis of the results points out a better performance relative to state-of-the-art approaches for lesion detection, and a similar performance for cancer detection but with a considerable simplification of the traditional pipeline. In addition, a reduction of the user interaction of the MRF-MAP is achieved, obtaining results in line with the state-of-the-art approaches for lesion segmentation, and a more robust and accurate performance of the segmentation framework is reported with the inclusion of elastography.

Resum

Aquesta tesi es centra en la detecció, segmentació i classificació de lesions en imatges d'ecografia. La presència d'aquestes lesions és un indicatiu de càncer de mama, i per tant, com més ràpida sigui la seva detecció, millor i més efectiu serà el tractament proporcionat al pacient. La contribució d'aquesta tesi és el desenvolupament d'una nova eina de Diagnòstic Assistit per Ordinador (DAO) capaç de detectar, segmentar i classificar automàticament lesions en imatges d'ecografia de mama. Inicialment, s'ha proposat l'adaptació del mètode genèric de detecció d'objectes Deformable Part Models (DPM) per detectar lesions en imatges d'ecografia. Aquest mètode utilitza tècniques d'aprenentatge automàtic per generar un model basat en l'Histograma de Gradients Orientats. Aquest mètode també és utilitzat per detectar lesions malignes directament, fent servir un detector multi-classe, simplificant així l'estratègia tradicional (detecció de lesions candidates, segmentació, extracció de característiques i classificació entre benigne i maligne). A continuació, s'han realitzat diferents propostes d'inicialització en un mètode de segmentació basat en Markov Random Field (MRF)-Maximum A Posteriori (MAP) per tal de reduir la interacció amb l'usuari. Per avaluar aquesta proposta, s'ha realitzat un estudi sobre la influència del tipus de lesió en els resultats aconseguits. Finalment, s'ha proposat la inclusió d'elastografia en aquesta estratègia de segmentació, modificant l'algoritme amb una formulació bi-variant. Els mètodes proposats per a cada etapa de l'eina DAO han estat avaluats fent servir bases de dades diferents. L'avaluació, feta de manera qualitativa i quantitativa fent servir diferents mètriques, ha estat realitzada per cada etapa independentment, i ha estat comparada amb els resultats dels mètodes més importants de l'estat de l'art. En aquesta comparació, la nostra proposta ha obtingut els millors resultats per a detecció de lesions, i uns resultats similars per a detecció de lesions malignes tot i que simplifica l'estratègia tradicional. Finalment, millora la inicialització del mètode MRF-MAP i ha obtingut resultats més acurats i robustos incorporant elastografia.

Resumen

Esta tesis se centra en la detección, segmentación y clasificación de lesiones en imágenes de ecografía. La presencia de estas lesiones es un indicativo de cáncer de mama, y por consiguiente, cuanto más rápida sea su detección, mejor y más efectivo será el tratamiento proporcionado al paciente. La contribución de esta tesis es el desarrollo de una nueva herramienta de Diagnóstico Asistido por Ordenador (DAO) capaz de detectar, segmentar y clasificar automáticamente lesiones en imágenes de ecografía de mama. Inicialmente, se ha propuesto la adaptación de un método genérico de detección de objetos Deformable Part Models (DPM) para detectar lesiones en imágenes de ecografía. Este método utiliza técnicas de aprendizaje automático para generar un modelo basado en el Histograma de Gradientes Orientados. Este método también es utilizado para detectar lesiones malignas directamente, usando un detector multi-clase, simplificando así la estrategia tradicional (detección de lesiones candidatas, segmentación, extracción de características y clasificación entre benigno y maligno). A continuación, se han realizado diferentes propuestas de inicialización en un método de segmentación basado en Markov Random Field (MRF)-Maximum A Posteriori (MAP) para reducir la interacción con el usuario. Posteriormente, se ha realizado un estudio sobre la influencia del tipo de lesión en los resultados conseguidos. Finalmente, se ha propuesto la inclusión de elastografía en esta estrategia de segmentación, modificando el algoritmo con una formulación bi-variante. Los métodos propuestos para cada etapa de la herramienta DAO han sido evaluados usando bases de datos diferentes. La evaluación, hecha de manera cualitativa y cuantitativa usando diferentes métricas, ha sido realizada en cada etapa independientemente, y ha sido comparada con los resultados de los métodos más importantes del estado del arte. En esta comparación, nuestra propuesta ha obtenido los mejores resultados para la detección de lesiones, y unos resultados similares para la detección de lesiones malignas pero simplificando la estrategia tradicional. Finalmente, mejora la inicialización del método MRF-MAP y obtiene resultados más precisos y robustos incorporando elastografía.

1 Introduction

1.1 Breast cancer

Breast cancer is the most common cause of death from cancer in women worldwide. According to a study developed in 2012 by the American Cancer Society, the chance of a woman having invasive breast cancer some time during her life is about 1 in 8, and it is estimated that 296,980 new cases of breast cancer will be diagnosed, and approximately 39,620 deaths are expected in the United States alone in 2013 [7]. This proportion is reduced in our country, Catalonia, where it is estimated that 1 in 9 women will develop breast cancer during their lifetime [1], and 4,700 new cases will be diagnosed. This quantity represents 30% of all cancer diagnoses in women [90].

In the European Union, breast cancer is the leading cause of death from cancer among women in most of its regions. The standardized death rate from breast cancer in the EU was 23.1 deaths per 100,000 female inhabitants during the period 2008-10 [41]. Figure 1.1 shows the distribution of deaths from breast cancer in women by EU regions.

Breast cancer incidence has increased over the past decade due to the introduction of screening programmes in which more cancers are detected in their early stages [90].

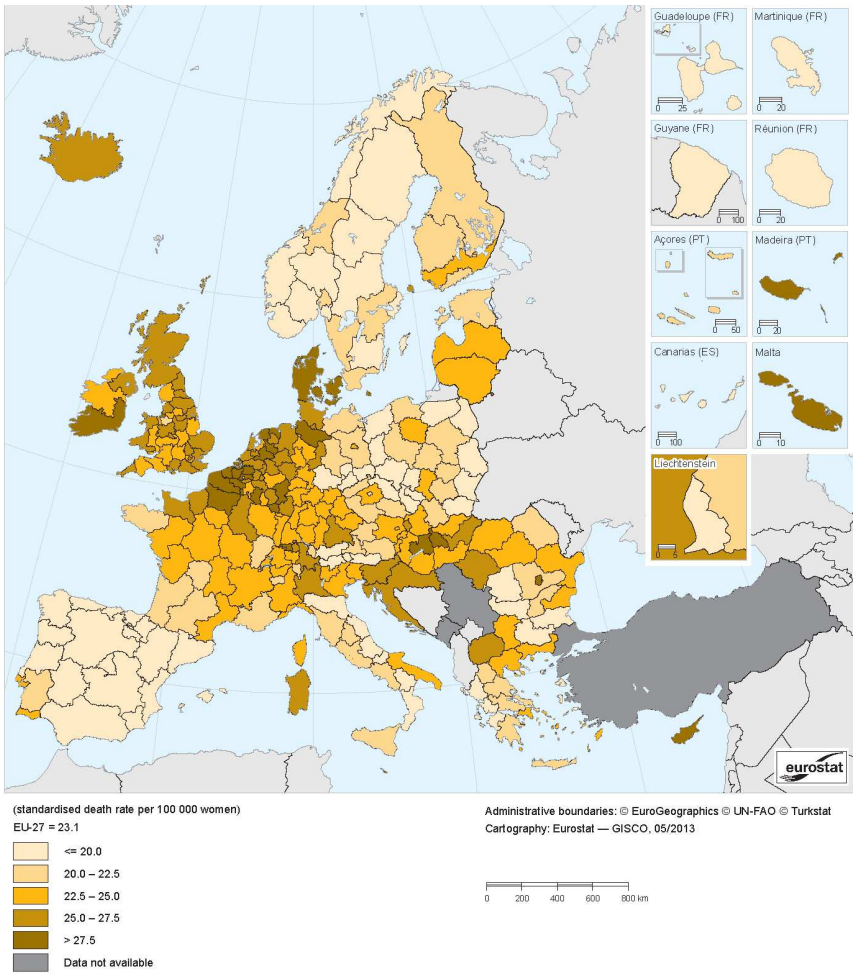


Figure 1.1: Map of deaths from breast cancer in women by EU regions (standardized death rate per 100,000 women) during 2008-10 [41].

However, breast cancer mortality has declined among women of all ages thanks to early detection [7]. The earlier breast cancer is detected, better and less aggressive therapeutic options are available.

1.2 Breast cancer diagnostic techniques

Since the causes of breast cancer still remain unknown, early detection is regarded as the best option to reduce the death rate [27].

Breast Self-Examination (BSE) or manual palpation is a very important part of every adult woman's personal health care. It is recommended that BSE should be performed

once each month beginning at age 20 and should continue throughout a woman's lifetime.

However there are some limitations to manual palpation. The major drawback is that the abnormality can only be felt after it has grown to a certain size (approx. 1.5 to 2 cm in diameter) [10]. Therefore, these techniques are not sensitive enough for early breast cancer detection.

Currently, manual examination is complemented with technology, such as imaging devices and signal processing algorithms. The development of these tools/systems improves the accuracy of human vision and judgment in the diagnosis. In the following sections, popular techniques for breast cancer diagnosis are introduced.

1.2.1 Digital Mammography (DM)

Mammograms capture low energy X-rays which pass through a compressed breast. Normally, two different viewpoints of the X-ray projection are obtained: the Cranio-Caudal (CC) view and the Medio-Lateral Oblique (MLO) view. In Figure 1.2, (a) shows the viewpoints' directions, while (b) and (c) show an example of the two mammogram views of the same patient.

(a) (b) (c)

Figure 1.2: Mammography viewpoints: (a) shows the direction of the two most used views, which produce images like (b) the CC view, and (c) the MLO view. The images illustrating this figure are taken from [98]

Nowadays, digital mammography is the most widely used and effective technique for breast cancer detection and diagnosis, and is accepted as the "gold standard" for breast imaging [106]. Digital Mammography (DM) can detect small cancers, known as lumps or masses, as well as calcifications and micro calcifications, which can indicate cancers in their

initial development.

Digital mammography can either be digitized Screen-Film Mammography (SFM) when the image is obtained as the digitization of an analogical film or Full-Field Digital Mammography (FFDM) when the image is directly generated in a digital sensor instead of a sensitive film.

However, there are limitations to mammography in breast cancer detection. Many unnecessary (65-85%) biopsies are due to the low specificity of mammography [26]. It is difficult for mammography to detect breast cancer in young women with dense breasts, where lesions have a similar attenuation compared to dense tissue [76], hence lesions are hidden by the surrounding tissue. In addition, the radiation of mammography can increase the health risk for patients and clinical staff.

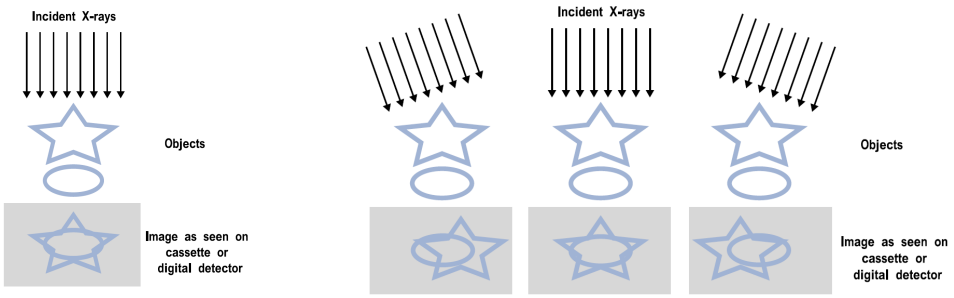
Therefore, it is advisable to use other image modalities like Magnetic Resonance Image (MRI) and Ultrasound (US) screening as complementary tools since they are more sensitive than mammography when dealing with dense breasts [16]. These techniques, in some cases, also offer higher specificity than mammography, allowing doctors and technicians to distinguish benignant and malignancy signs, which reduces the number of unnecessary biopsies [46, 114, 116].

Breast tomosynthesis

One of the recent advances in X-ray screening is tomosynthesis, which creates a "pseudo" 3-dimensional picture of the breast. This technique tries to overcome the tissue overlap effect, present in regular mammograms. The acquisition technique is similar to mammography: the breast is compressed between two plates and X-ray attenuation is measured. The main difference is that tomosynthesis takes multiple images of the breast from different angles instead of from a single viewpoint. These images are further combined to reconstruct them into cross-sectional slices. Figure 1.3 illustrates the effect of taking images at different angles.

1.2.2 Magnetic Resonance Imaging (MRI)

Magnetic Resonance Image (MRI) is a radiological technique begun in the 1980s. MRI uses magnetic fields and the magnetic properties of the body to generate detailed images of body tissues. MRI acquires 3D data, where cross sections at arbitrary view angles can be calculated. Figure 1.4 shows the most common views: (a) the axial, (b) the sagittal



(a) (b)

Figure 1.3: Mammography and Tomosynthesis acquisition processes. (a) Mammography single image take illustrating the tissue overlap problem. (b) Multiple images take for tomosynthesis showing how the relative position between two targets vary depending on the X-ray illumination angle. The images illustrating this figure are taken from Smith et al. [113].

and (c) the coronal views.

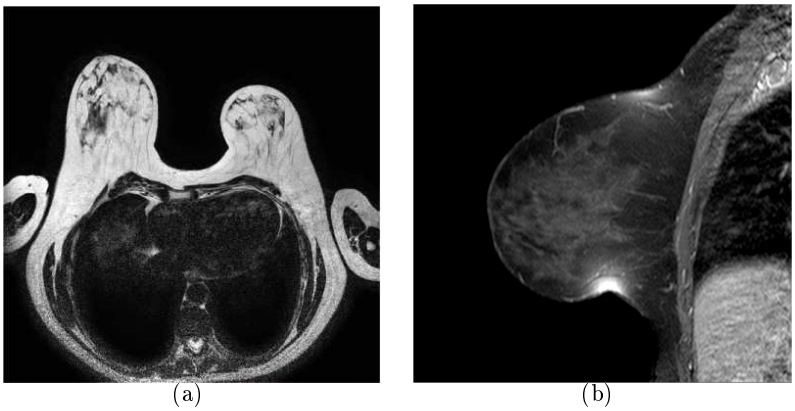
MRI is a widely used method for the detection of breast cancer with sensitivity as one of its principal strengths [32]. Some other advantages are the noninvasive nature and the ability to image in any plane. It is also useful in the detection of recurrences after surgery has been performed for mastectomy or lumpectomy cases.

The disadvantages of MRI are that it is expensive and time consuming. MRI is an impractical tool for routine screening, but plays a major role as an adjunct to mammography and US imaging.

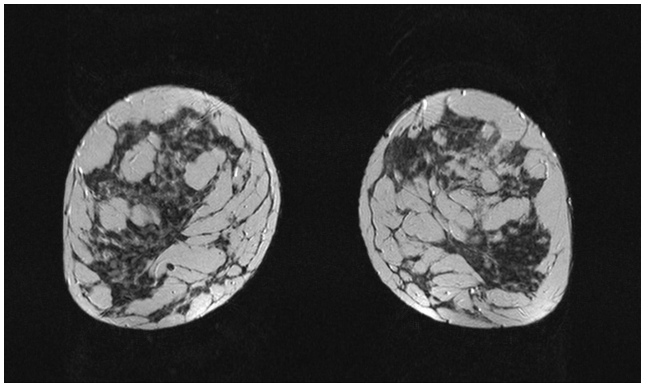
1.2.3 Breast ultrasound

Breast ultrasound imaging uses the transmission of high-frequency mechanical waves (sound waves typically within the $1 \sim 20$Mhz range) through the human body in order to capture the waves partially reflected at the boundaries between tissues with different acoustic properties [40]. US imaging or sonography is an important modality in the evaluation and treatment of breast masses. It is used as a complementary modality to answer specific questions raised in the findings from mammography or physical examinations [122].

B-mode (Brightness-mode) ultrasound is the most common form of ultrasound imaging nowadays. Sweeping an ultrasound beam through the area being examined while transmitting pulses and detecting echoes along closely spaced scan lines produces B-scan images. The vertical position of each bright dot is determined by the time delay from pulse trans-



(a) (b)



(c)

Figure 1.4: MRI viewpoints of the same patient: (a) shows axial view, (b) sagittal view, and, (c) coronal view.

mission to return of the echo, and the horizontal position is given by the location of the receiving transducer. To generate a rapid series of individual 2D images that show motion, the ultrasound beam is swept repeatedly. The returning sound pulses in B-mode have different shades of darkness depending on their intensities. The varying shades of gray reflect variations in the texture of internal organs. Interfaces between tissues with very different impedances are displayed with high brightness, and, when the impedances are the same or nearly the same, the position appears dark. This form of display is also called gray scale (see Figure 1.5).

Nowadays, researchers and practitioners are showing an increasing interest in the use of US images for breast cancer diagnosis [117]. Use of ultrasound can increase over all cancer detection by 17% [35], and reduce the number of unnecessary biopsies by 40% [9]. Ultrasound examination is safe for patients and radiologists in daily clinical practice [70] due to the lack of radiation in the acquisition procedure. Since it is a cheap and fast

Figure 1.5: B-mode ultrasound image of a breast.

technique, it is also convenient for low-resource countries [8]. Ultrasound imaging is more sensitive than other techniques, i.e. mammography, for detecting abnormalities in dense breasts (normally the case of women younger than 35 years of age [35]). It is also accurate at detecting some kinds of lesions, such as cysts, reducing the number of unnecessary biopsies. However, US imaging is operator-dependent compared to other common used techniques. Interpreting US images requires experienced and well-trained radiologists due to its complexity and the existence of artifacts.

1.3 Breast US artifacts

Due to the ultrasonic image formation process, the final visible image presents some artifacts. Some of these are useful in the diagnostic procedure while others degrade the visualization of the image. In this section, we explain the most influential artifacts in US imaging: speckle noise and the shadowing effect.

1.3.1 Speckle noise

Speckle is an inherent characteristic of ultrasound imaging. It is generated by scatterers located throughout the tissue, and appears as a granular structured noise [2, 3]. Speckle is

detrimental because it reduces the distinction between the structures and the background of the image. Although speckle seems to form an acoustic texture pattern, it is intrinsically dependent on the imaging system, and is not directly linked to any physical characteristic of the tissue. Figure 1.6 shows an example of speckle noise in a synthetic ultrasound image.

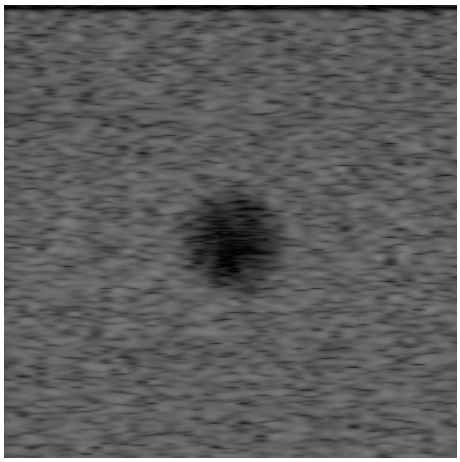


Figure 1.6: Example of speckle noise in a synthetic image.

1.3.2 Shadowing effect

The shadowing effect occurs when the ultrasonic beam is attenuated by a structure within the B-mode scan range. In B-mode US images, this effect can be observed as a dark shadow below the structures (see Figure 1.7). Shadowing may indicate the presence of tumors, which attenuate the ultrasonic beam due to their higher tissue density. On the other hand, the absence of shadowing may be a hint to distinguish cystic lesions as cysts are filled with liquid and do not absorb the ultrasonic beam. Hence, shadowing is an important feature in the diagnosis. Also, note that as long as the ultrasonic beam is propagated through the media, it becomes inherently attenuated, and therefore, the B-mode image becomes darker without meaning that there is any shadowing presence.

1.4 Lesion pathologies

A breast lesion is an area that is an abnormality or an alteration in the breast tissue's integrity. Breast lesions are mainly classified in two categories depending on whether they are composed of cancerous cells or not. We can further categorize lesions within cancerous or non-cancerous, such as cysts, fibroadenomas or carcinomas. Each type of lesion has

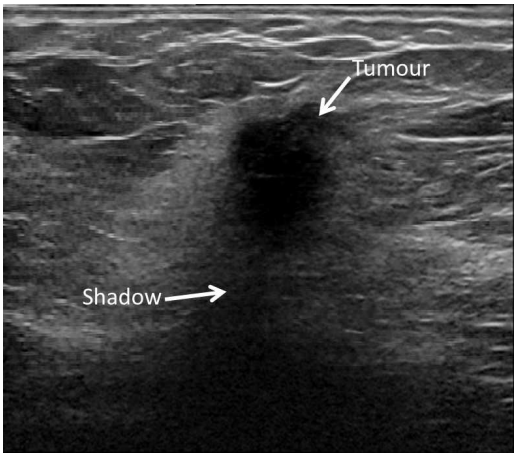


Figure 1.7: Shadowing effect. Breast B-mode image, where the tumor and the shadow produced by the beam attenuation are marked with arrows.

unique characteristics that are present when visualized in sonography.

1.4.1 Benign tumors

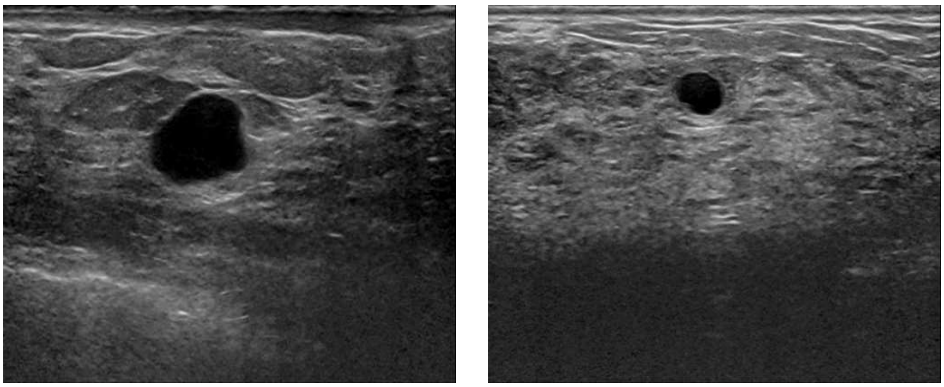
The majority of breast tumors detected by screening are benign [7]. Benign breast tumors are non-cancerous areas in which breast cells have grown abnormally and rapidly, often forming a lump [62].

Benign breast tumors are not life-threatening, and do not spread beyond the breast to other organs. Still, some benign breast conditions are important to diagnose as they are regarded as important risk factors in developing breast cancer.

In general, benign masses tend to be of low density, and vessels may be seen through the mass. Malignant masses, however, are often denser than the adjacent parenchyma, and may appear too dense for their size. The most common types of benign tumors are described below.

Cyst

Cysts are fluid-filled, anechoic, well-marginated, round or ovoid structures. The incidence of cysts is in as many as one third of women between 35 and 50 years old. Cysts cannot reliably be distinguished from solid masses by clinical breast examination or mammography; in these cases, sonography is used. Figure 1.8 shows examples of Breast Ultrasound (BUS) images with cysts.



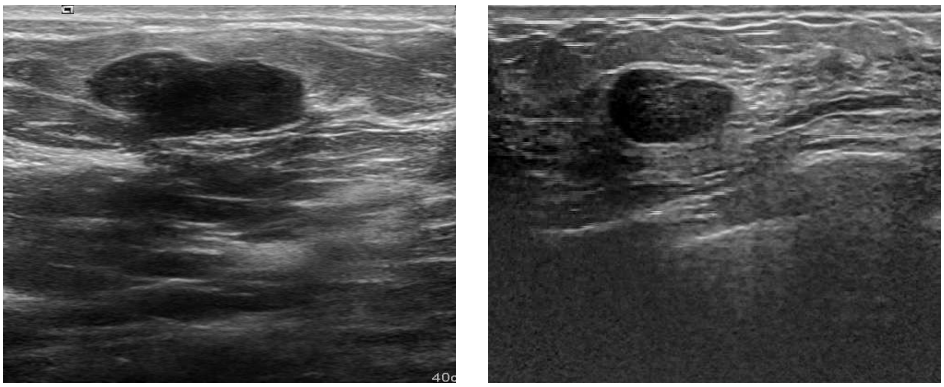
(a) (b)

Figure 1.8: Benign tumors: cysts.

Fibroadenoma

Fibroadenoma (FA) is the most common lesion in the breast, and it occurs in 25% of asymptomatic women [39]. It is usually a disease of early reproductive life. The peak incidence is between the ages of 15 and 35 years, but they may be found in women of any age. Women with fibroadenomas have a slightly increased risk of breast cancer (about 1.5 to 2 times the risk of women with no breast changes) [7].

Figure 1.9 shows examples of fibroadenoma tumors. In appearance, these masses are firm, smooth and oval shaped, with borders that are distinct from the surrounding breast tissue. They grow as spherical nodules that are usually sharply circumscribed, and movable in the surrounding breast tissue. Since they are not filled with fluid like cysts, the masses appear more heterogeneous.

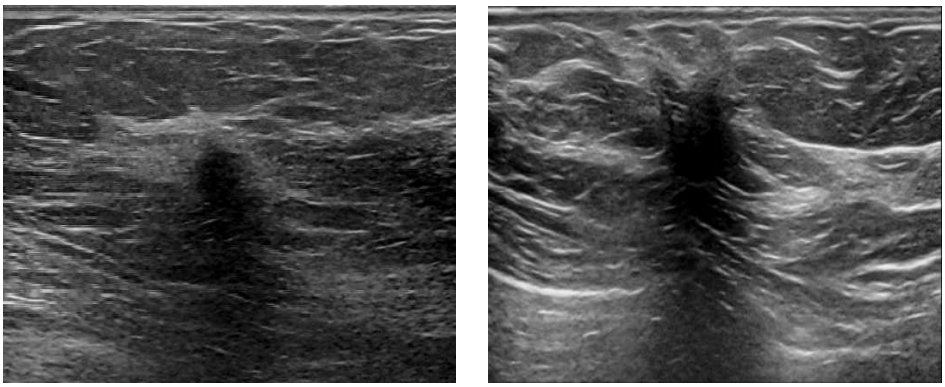


(a) (b)

Figure 1.9: Benign tumors: fibroadenomas.

1.4.2 Malignant tumors

Characteristics of malignant lesions include a hypo-echoic nature (heterogeneous internal echoes and acoustic shadowing), and irregular margins, as illustrated in Figure 1.10. Malignant tumors or carcinomas are the most common malignancy seen in the female breast. Breast carcinoma can be categorized into two main groups: ductal carcinoma, where the malignancy originates in ductal epithelial cells, and lobular carcinoma, where the malignancy originates in the more distal cells of the lobule [116].



(a) (b)

Figure 1.10: Malignant tumors: (a) invasive ductal carcinoma and (b) invasive lobular carcinoma.

Each category is often subdivided into 'in-situ' (non-invasive) and infiltrating (invasive) types.

Non-invasive breast cancer

Non-invasive breast cancer is also known as cancer or carcinoma in situ. This cancer is found in the breast ducts, and has not developed the ability to spread outside the breast. This form of cancer rarely shows as a lump in the breast, and is usually found in a mammogram. The most common type of non-invasive cancer is Ductal Carcinoma In Situ (DCIS).

Invasive breast cancer

Invasive Carcinoma (IC) has the ability to spread outside the breast, although this does not mean it necessarily has spread. The most common form of breast cancer is Invasive

Ductal Carcinoma (IDC), which develops in the cells that line the breast ducts. Invasive ductal breast cancer accounts for about 80% of all breast cancer cases [7].

1.5 Computer Aided Diagnosis

Since the introduction of screening programmes, Computer Aided Diagnosis (CAD) has become a part of the routine clinical work for diagnosis of breast cancer [34], mostly used in mammography. A CAD system is a set of automatic or semiautomatic tools developed to assist radiologists in the evaluation of medical images [19, 51]. With CAD, radiologists use the computer output as a "second opinion", and make the final decisions.

Nowadays, with the increasing use of complementary techniques such as sonography, researchers have been working on CAD systems capable of dealing with them. Hence, CAD systems are very important in sonography due to the complexity of US images as explained before. Reading and understanding US images requires well-trained and experienced radiologists. Furthermore, even well-trained experts may have a high inter-observer variation rate; therefore, CAD has been investigated to help radiologists in making accurate diagnoses. One advantage of a CAD system is that it can use some features which cannot be obtained visually and intuitively by radiologists when performing the diagnosis process. Another advantage is that CAD systems can reproduce the diagnosis process in different images, eliminating the operator-dependent nature inherent in sonography. Current research into the use of CAD systems is being done with an eye toward helping radiologists to increase diagnosis accuracy, reduce the biopsy rate, and save time and effort.

Generally, ultrasound CAD systems for breast cancer detection involve four stages, as shown in Figure 1.11.

1. Lesion detection: The task of lesion detection is to provide the spatial location of the lesion in the image. Usually, a point inside the actual lesion region is reported as a result.

2. Lesion segmentation: Image segmentation divides the image into non-overlapping regions and it separates the objects (lesions) from the background.

3. Feature extraction: This step finds features that can accurately distinguish between benign and malignant lesions.

4. Lesion classification: The task of lesion classification is to classify between benign and

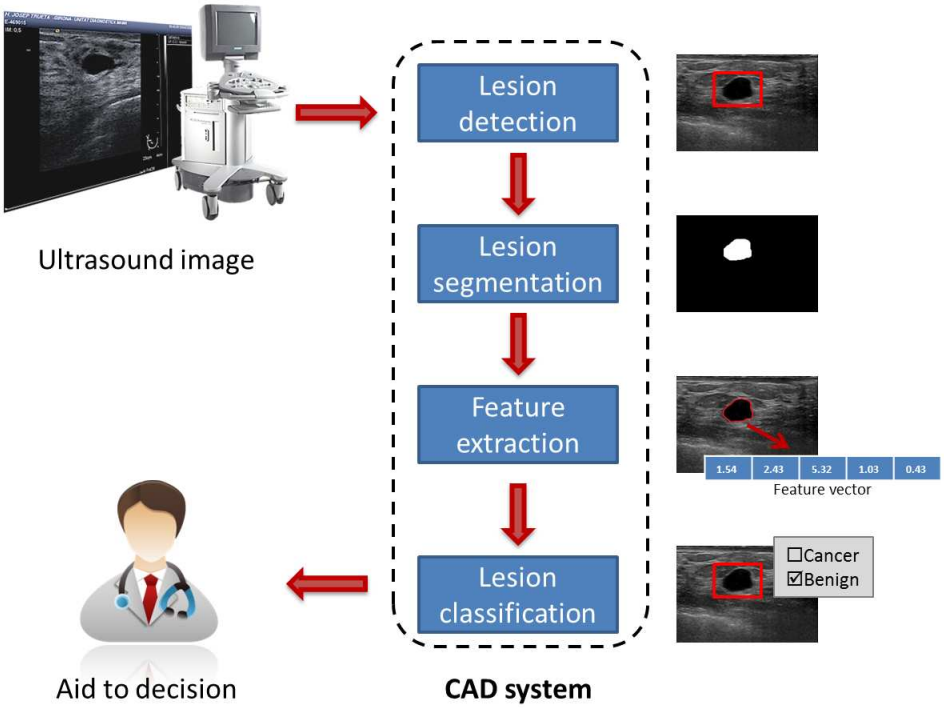


Figure 1.11: A CAD system for breast cancer diagnosis.

malignant lesions the suspicious regions found in the previous stages using machine learning techniques.

1.6 Aims and objectives

One of the main research efforts in early detection of breast cancer is to include the development of software tools to assist radiologists in the diagnosis procedure. Along this line, the main objective of this thesis is

to provide the basis of a new CAD system capable of automatically detecting, segmenting, and classifying breast abnormalities in ultrasound images.

The main objective can be divided as follows:

- Analyze the state-of-the-art sonography CAD methods. This focuses on the critical

and technical study of the literature to learn the main solutions previously applied to solve the detection, segmentation, and classification problem, and highlight advantages and drawbacks.

- Propose a novel methodology for automatic lesion detection in sonography.
- Propose a lesion segmentation algorithm suitable to be implemented in a real clinical environment.
- Study the segmentation results according to the type of lesions.
- Investigate the inclusion of additional diagnostic information (i.e. elastography) into the CAD proposal.

1.7 Thesis outline

This thesis is organized into three parts. The first part, Chapter 1 and 2, provides fundamental and background knowledge of the subject area and the state-of-the-art in technological development. The second part of the thesis, Chapter 3 to 5, presents the proposed framework. The third part, concludes the thesis with an insight to future directions of research and development, Chapter 6.

- Chapter 1 provides an overview of the thesis. It defines the problem domain, provides background knowledge of the subject area, and specifies the thesis' aims and objectives. Finally, it outlines the thesis' organization.
- Chapter 2 provides a technical review of state-of-the-art image analysis approaches used in ultrasound breast imaging, including different detection, segmentation, and classification techniques. It further highlights advantages and limitations of the existing algorithms.
- Chapter 3 proposes a novel automatic lesion detection approach in breast US images. The proposed approach adapts the Deformable Part Models methodology to be used in sonography. The best configuration of parameters for lesion detection as well as cancer detection is studied. Finally, the performance of the proposed method is compared to the most relevant methods in the literature.
- Chapter 4 presents a lesion segmentation framework whose user interaction process is reduced thus making it suitable for real clinical practise. The initialization of the

segmentation method is reduced to a seed point that can be marked by the user with a "one-click" interaction, or provided by a lesion detection method. It further analyzes the results depending on the lesion type. This study allows one to obtain conclusions regarding the best segmentation approach overall and depending on lesion type, as well as which pathologies present more complexity for the segmentation process.

- Chapter 5 proposes the inclusion of elastography information in the segmentation framework. This provides the necessary background knowledge of the elastography technique, as well as a technical review of state-of-the-art approaches that use elastography in CAD.

- Chapter 6 concludes the thesis with a summary of contributions made by the thesis, limitations of the present context of research and an insight into future directions of research.

2

Literature review

2.1 Introduction

When analyzing a new image from a patient, any CAD system needs to answer a set of questions to help the radiologists in the diagnosis process. These questions are related to the three main stages of all CAD systems: Detection (*Where is the lesion?*), Segmentation (*What are the lesion margins?*), and Classification (*Is the lesion a cancer?*).

In this chapter, the recent state-of-the-art of detection, segmentation and classification in breast sonography is reviewed. Notice that some works may propose one (i.e. only segmentation) or more of these stages (i.e. detection + segmentation). In such cases, each step is analyzed individually in its corresponding section.

2.2 A review of lesion detection methods on breast sonography

Due to the artifacts inherent in sonography, lesion detection is an important procedure for breast US CAD systems, in which the spatial location of the abnormality is automatically

provided. As a result, this location is commonly labeled with a point inside the actual lesion region or as a Region Of Interest (ROI).

The main goal of this section is to review the most important lesion detection algorithms found in the state-of-the-art, pointing out their advantages and disadvantages, as well as their evaluation results. These lesion detection methods can be classified into methods based on image processing or machine learning techniques.

First, all of the metrics used by the literature methods are commented in the following subsection.

2.2.1 Lesion detection evaluation criteria

When analyzing different detection methods of the state-of-the-art, different measures are commonly used. A list of the most common criteria for assessing lesion detection is presented below. Most of the measures take into consideration the elements of the confusion matrix in Table 2.1. In terms of lesion detection, two play an important role:

- True Positive (TP): the algorithm returns as a detection a region that is actually a lesion.
- False Positive (FP): the algorithm returns as a detection a region that is actually background.

Table 2.1: Confusion matrix.

		Reference	
		Positive	Negative
Prediction	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

The following measures have been proposed in the literature for lesion detection:

True Positive Rate (TPR), also called sensitivity [75] or accuracy [130] in other works, determines the relationship between the number of true positive detections and the total number of actual positives.

False Positive Rate (FPR), determines the fraction of False Positive (FP) out of the total actual negatives. Notice that in [109], False-Positive Rate (FPR) is considered as the

fraction of mis-detected images out of the total number of images.

Receiver Operating Characteristic (ROC) analysis is a graphic plot that illustrates the performance of a classifier system as its discrimination threshold is varied. It is created by plotting the fraction of true positives out of the total actual positives (TPR) vs. the fraction of false positives out of the total actual negatives (FPR) at various threshold settings. For quantitative results, the area under the curve (A_z) is used. Figure 2.1 shows an example of different ROC curves and their respective A_z value. The ROC analysis has been used to assess the following works: [35, 37, 38, 80, 94].

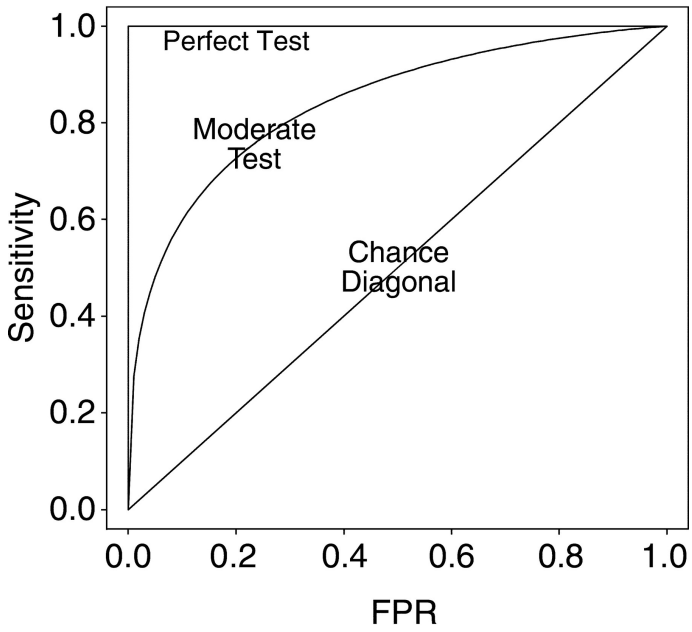


Figure 2.1: Example of different ROC curves. A perfect test has an area under the ROC curve A_z of 1.0. The chance diagonal has an A_z of 0.5. Tests with some discriminating ability have ROC areas between these two extremes.

Free-response ROC (FROC) analysis is similar to the ROC analysis, except that the false positive rate on the X-axis is replaced by the number of false positives per image. FROC seeks location information from the algorithm, rewarding it when the reported lesion is marked in the appropriate location, and penalizing it when it is not. For quantitative results, the sensitivity (or TPR) at a certain number of FP/image is reported. The FROC analysis has been used to assess the following works: [35, 36, 37, 38].

Average Detection Error (ADE) is used in [134] as the average Euclidean distance between the corners of the detected box (ROI) and the ground truth box (reference delineated for the radiologist).

Precision Ratio (PR) is defined in [85] as follows:

$$PR = \frac{N_{CR}}{N_C} \tag{2.1}$$

When detecting lesions, they divide the image into lattices which are then classified as lesion or normal tissue. Let the number of lattices classified as lesion and normal tissue be N_C, and the number of correctly classified lattices be N_{CR}.

Recall Ratio (RR) is defined in [85] as follows:

$$RR = \frac{N_{CR}}{N_R} \tag{2.2}$$

where N_R is the number of lattices in the actual lesion region.

2.2.2 Image processing detection

Image processing detection methods involve the use of common image analysis techniques such as thresholding, filtering, or modeling. The lesion detection procedure is performed directly to the target image without a previous training process.

Examples of these approaches are the work of Drukker et al. [35] who studied the use of Radial Gradient Index (RGI) filtering. Each point in the image is multiplied with a constraining function (bivariate Gaussian function). The resulting Gaussian constrained image is then thresholded at several levels in order to generate a set of contours. The RGI [79] is calculated for all the contours in every contour set. The maximum RGI value of every contour set is used to generate a RGI image, which is thresholded to determine a set of lesion candidates. A dataset of 757 images from 400 different cases was used in order to evaluate the methodology in terms of ROC and Free-response Receiver Operating Characteristic (FROC) analysis. An area under the ROC curve (A_z) value of 0.84, an overall performance by case of 94% sensitivity at 0.48 FP per image, and a sensitivity of 89% by image was reported in the original paper. Later, they tested their proposed method with different datasets. First, in [38], with two different datasets, one of 757

images from 400 patients, and one of 1740 from 458 patients, obtaining an Az value of 0.91, and a sensitivity of 90% at 0.45 FP/image. Notice that these results were reported by case instead of by image. Further in [37], they also used two different datasets, one consisting of 151 images all from different patients, and one with 1740 images from 458 patients, obtaining an A_z of 0.95, and a sensitivity of 80% with 0.6 FP per image.

To improve the detection of malignant tumors, Drukker and Giger [36] developed a non-linear filtering technique based on the skewness of the gray-level distribution to detect the posterior acoustic lesion shadowing rather than the lesion itself. They reported a sensitivity of 66% per image at a rate of 0.25 FP/image, detecting malignant lesions for a dataset consisting of complicated cysts, solid benign lesions, and malignant lesions, and a sensitivity of 30% at 0.25 FP/image for the overall dataset (194 images from 94 patients).

Kutay et al. [80] applied the Power-Law Shot Noise (PLSN) model for tissue characterization. The PLSN model is generated using features extracted directly from the ultrasound Radio Frequency (RF) signal. Parameters of their model are estimated from clinical ultrasound images, and then used in the detector. They obtained an A_z of 0.97 detecting lesions, and 0.81 for malignant lesion detection, using a dataset of 100 images from 25 cases.

Further, Yap et al. [130] proposed the use of hybrid filtering, multifractal processing, and thresholding segmentation for lesion detection. They first pre-processed the US images with histogram equalization and a speckle noise reduction process that uses a hybrid filtering approach that combines the nonlinear diffusion filtering and linear filtering (Gaussian blur). Then, multifractals [53] are used to further enhance the partially processed images. After pre-processing, the images are thresholded and a rule-based candidate selection (based on the size of the region and location) is used as a discriminative criterion. They evaluated the detection performance in a dataset of 360 images obtaining an accuracy of 86%.

Shan et al. [109] proposed a lesion detection methodology that considered both texture features and spatial characteristics. They first use the Speckle Reducing Anisotropic Diffusion (SRAD) [133] as a de-speckling method. Once the image is de-speckled, an iterative threshold selection algorithm is applied to segment the image. Only regions that intersect with an image center region (a window about half the size of the entire image located at the image center) are considered as lesion candidates. A True-Positive Rate (TPR) of 95.2% and a FPR of 4.76% were reported using a dataset of 105 images.

Summarizing, image processing detection methods are fast and avoid any kind of offline computation. In principle. they can also be used in a dataset of any size since they do

not need to split the dataset in training and testing. However, most of these methods are designed to obtain good results in their own datasets, taking advantage of unique features in their images such as lesion location, speckle noise influence or intensity appearance. This sometimes leads to a poor generalization capability of the methods, specially with different datasets.

2.2.3 Machine learning-based detection

Machine learning-based methods generate statistic models from a training dataset to detect lesions in a target image using any sort of machine learning techniques and features.

One of the first works to include a machine learning detection process was Madabhushi and Metaxas [88] in the proposal of a fully-automatic segmentation method. For detecting lesions in sonography, they proposed using the Stavros criteria [116] to determine which pixels are most likely to be part of a lesion. The Stavros criteria integrates the posterior probability of intensity and texture, constraining it with prior knowledge, taking the position of the pixel into account. To generate this posterior probability, a set of training images are needed. Although they performed a detection step for seed placement in a segmentation process, only the segmentation results were presented in the paper.

Mogatadakala et al. [94] proposed a nonparametric model based on different order statistics estimated from multiresolution decompositions (wavelets) of energy-normalized subregions. The features are then classified using a Linear Classifier (LC). They assessed the detection method using a dataset of 204 images from 84 cases, achieving an area under the ROC curve A_z of 0.91.

Liu et al. [85] also used texture features in their method, where the image is divided into lattices of the same size. The texture information (entropy, contrast, sum average and sum entropy) of each lattice is extracted and then a well-trained classifier based on Support Vector Machine (SVM) is employed for classifying the lattices. In the end, the ROI is generated according to the classification results and a set of background knowledge-based rules. They achieved an average PR of 82.33%, and an average RR of 83.81% using a dataset of 112 images.

Zhang et al. [134] proposed a machine learning framework using a Probabilistic Boosting Tree (PBT) classifier with Haar-like features [123] extracted from the image. They first collect positive patches determined by the bounding boxes of tumors and negative patches randomly cropped from background. Then Haar-like features are extracted and a PBT

classifier [121] is trained, which involves the recursive construction of a tree where each of its nodes is an Adaboost classifier [52]. When testing, given a new image, tumors are detected by exhaustively checking all possible bounding boxes within a search range. They assessed their method in a large dataset of 347 images from different patients, obtaining an ADE of 15.47 pixels.

Jiang et al. [75] also presented a machine learning approach for automatic tumor detection using Haar-like features extracted from the image. In contrast, an Adaboost classifier is used in their work to locate potential tumor locations. They assessed the detection performance in a dataset of 112 images in terms of accuracy (87.5%), sensitivity (88.8%), and specificity (84.4%).

Recently, Massich et al. [93] used a multifeature Bayesian machine learning framework to determine whether a particular pixel of the image is a lesion or not. From the learning step, a Maximum A Posteriori (MAP) probability plane of the target image is obtained and thresholded with certain confidence. Then the largest area is selected as the candidate region for further expansion. Their work evaluated the detection results in relationship with the segmentation results: the segmentation results are analyzed depending on the proximity of the seed point detected by the method to the actual lesion center.

Finally, Hao et al. [64] proposed a fully-automatic segmentation framework of breast lesions using an objective function combining Deformable Part Models (DPM) [48] detection with intensity histograms, texture descriptors, and position information using a graph-cut minimization tool and normalized cuts [112] as image segments. The deformable part-based detector produces a large number of detections for potential lesion areas. Then, instead of selecting the detection with maximum confidence and discarding all of the others, they propagate the confidence of any detection to the corresponding segment. As in Madabhushi and Metaxas' [88] work, only the segmentation results were reported.

In summary, machine learning-based methods are in general robust and adaptable. Since such methods are machine-learning-based, a particular model is constructed for each dataset used, which makes the method adaptable to the specific characteristics of any dataset. However, the extracted features used to train the classifier need to be sufficiently distinguishable to discriminate between lesions and non-lesions. Moreover, the size and variability of the dataset play an important role in the training and testing steps.

2.2.4 False-positive reduction

Lesion detection in breast sonography is a difficult task because of the nature of the images, which induces, in some cases, the detection of regions without the presence of abnormalities (known as FP). Some of the works reviewed above require an extra procedure to reduce the number of false positive detections and thus improve their results.

Drukker et al. [35, 37, 38] proposed the use of machine learning classification to reduce the number of lesion candidates found in the lesion detection procedure. From the rough contours generated by thresholding the lesion in the detection process, some features such as lesion shape, margin sharpness, texture, acoustic characteristics and shadowing are extracted. Then, a Bayesian Neural Network (BNN) is trained to discriminate between actual lesions and FP results. In [35], results without FP reduction were reported to assess the performance of this process. Specifically, they obtained a TPR of 87% with 0.76 FP/image before the FP reduction and TPR of 89% with 0.48 FP/image after the process.

Another methodology that proposed the use of an additional machine learning step in their framework to reduce the number of false-positive results was presented by Jiang et al. [75]. In their work, a SVM classifier is trained using quantized intensity features extracted from the lesion candidate regions to remove FPs.

Finally, Shan et al. [109] proposed a region ranking procedure to assign a value of likelihood for each lesion candidate taking into account the variance of the intensity values in the center of the region, the area of the region, and its position in the image.

Notice that the results of the methods that include a FP reduction step described in sections 2.2.2 and 2.2.3 were obtained after this process. It is also worth commenting that only Drukker et al. [35] published the intermediate detection results to evaluate the performance of the FP reduction process.

2.2.5 Summary

In this section, different lesion detection approaches in breast US images have been reviewed. We have described several algorithms, pointing out their main features. Table 2.2 summarizes the analysis done on the principal methods discussed in this section. The methods are first sorted by the detection criteria (image processing or machine learning-based), and then, by year of publication (ascending order). Column *FP reduction* specifies the process followed, if any. Column *Classifier* shows the machine learning technique used by the machine learning-based detection methods, or in the FP reduction process. Note

that in Jiang et al. [75], an AdaBoost was used for the initial detection, and a SVM classifier for the FP reduction step.

Table 2.2: Summary of the results presented in the articles analyzed for lesion detection.

Reference	Detection	Method	FP reduction	Classifier	Measure	Result	# cases (# images)
Drukker (2002) [35]	Image processing	RGI filtering	Feature extraction	BNN	A_z TPR FP/image	0.84 0.89 0.48	400(757)
Drukker (2003) [36]	Image processing	Skewness filtering	-	-	TPR FP/image	0.66 0.25	94(194)
Drukker (2004) [38]	Image processing	RGI filtering	Feature extraction	BNN	A_z TPR FP/image	0.91 0.9 0.45	400(757) & 458(1740)
Drukker (2005) [37]	Image processing	RGI filtering	Feature extraction	BNN	A_z TPR FP/image	0.95 0.8 0.6	400(757) & 151(151)
Yap (2008) [130]	Image processing	Multifractal filtering	-	-	TPR	86%	360(360)
Shan (2008) [109]	Image processing	Thresholding	Region ranking	-	TPR FPR	95.2% 4.76%	105(105)
Madabhushi and Metaxas (2003) [88]	Database trained	Stavros criteria	-	-	-	-	-
Mogatadakala (2006) [94]	Database trained	Wavelets	-	LC	A_z	0.91	84(204)
Liu (2010) [85]	Database trained	Feature extraction	-	SVM	PR RR	82.33% 83.81%	112(112)
Zhang (2010) [134]	Database trained	Feature extraction	-	PBT	ADE	15.47 pix.	347(347)
Jiang (2012) [75]	Database trained	Feature extraction	Feature extraction	AdaBoost/ SVM	TPR	88.8%	112(112)
Massich (2012) [93]	Database trained	Thresholding	-	-	-	-	-
Hao (2012) [64]	Database trained	DPM	-	LSVM	-	-	-

The analysis in Table 2.2 shows a trend for the newest algorithms to use machine learning techniques. We can see that the most recent work using image processing detection dates from five years ago. The adaptability of the machine learning-based methods to the specific characteristics of the target dataset makes them an attractive choice when designing a novel detection algorithm. They also do not need a false positive reduction step as most of the detection methods based on image processing, with the exception of the approach presented by Jiang et al. [75].

In addition, two of the traditional problems in machine learning works, computation time and the dataset size, are no longer a difficulty. Currently, when dealing with machine learning approaches, the offline computation is not an insurmountable problem. There is also no need of large datasets for training and testing, since training/testing strategies such as leave-one-out or k-fold cross validation may be used in reduced datasets to simulate the performance of the approach in large datasets. However, a representative number of cases is needed for a good performance.

2.3 A review of lesion segmentation methods on breast sonography

Segmentation of lesions or abnormalities is a fundamental procedure for breast CAD systems. The aim of segmentation is to divide the image into two different regions: foreground (lesion) and background (other tissues). Subsequently, statistics of the suspicious area can be computed to assist radiologists in the diagnoses, therefore, it is important to obtain accurate boundaries of the masses.

The main goal of this section is to review the most important mass segmentation algorithms found in the current literature, pointing out their advantages and disadvantages. We also describe the evaluation of the methods reported in their papers indicating the measures, results, and size of dataset used. All of the metrics are described in subsection 2.3.1.

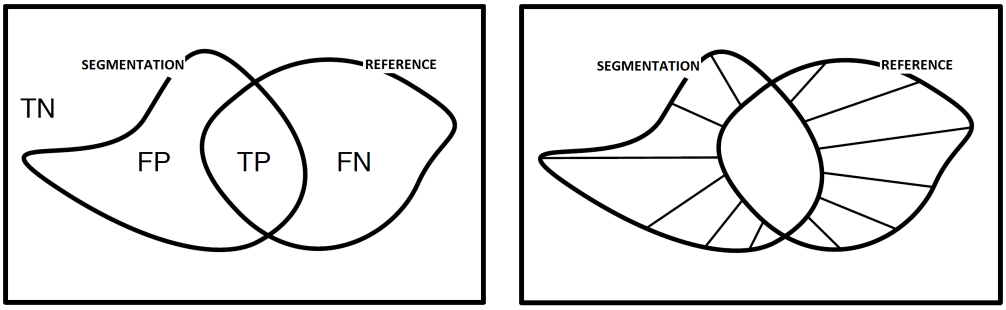
Classification criteria of the methods is based on the methodology used for segmenting the lesions, hence, the segmentation techniques are classified into histogram thresholding, Active Contour Models (ACM), Markov Random Fields (MRF), machine learning, watersheds, and graph-based methods. Note that some works combine more than one methodology; in such cases they are assigned to the methodology with the most important role in the segmentation process.

2.3.1 Lesion segmentation evaluation criteria

Multiple criteria arise in the literature when assessing different segmentation methods. However, this criteria can be grouped into two categories depending if they are area or distance based metrics as illustrated in Figure 2.2. Area based metrics evaluate the amount of area shared between the segmentation obtained and the ground truth. On the other hand, distance based metrics quantify the displacement between the obtained and the desired lesion boundary.

Area based metrics

When analyzing the areas described in Figure 2.2(a), as a result of assessing the segmentation result with the reference delineation, four areas become evident: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN); corresponding to the regions of the confusion matrix in Table 2.1. In terms of object segmentation, these four



(a) (b)

Figure 2.2: Methodology evaluation. Graphic representation of (a) the evaluation in terms of area and (b) boundary distance measures.

values mean the following,

- True Positives (TP): the number of pixels segmented as foreground that are actually foreground.
- True Negatives (TN): the number of pixels segmented as background that are actually background.
- False Positives (FP): the number of pixels that are incorrectly classified as foreground, when they actually belong to the background.
- False Negatives (FN): the number of pixels that are incorrectly classified as background, when they actually belong to the foreground.

Expressing the results in terms of how many pixels belong to each of these classes is not clear enough to determine how good the results are. For that reason, different area metrics relating the four regions are commonly used. Most of the indexes are defined within the interval [0, 1], where 1 indicates perfect overlap and 0 means no overlap at all, although some works report their results as a percentage.

Dice Similarity Coefficient (DSC) [33] is a well-known measure, and is the most commonly used. This measure penalizes the over-segmentation due to the FP being included in the denominator. However, the TP is multiplied by two to give more relevance to the well classified pixels. This measure is expressed as

$$DSC = \frac{2 * TP}{2 * TP + FP + FN} \tag{2.3}$$

It is typically considered that DSC values equal to or higher than 0.7 suggest good agreement between two segmentations [13]. The DSC metric has been used to assess the following works: [54, 69, 71, 102].

Area Overlap (AO), also known as the Jaccard Similarity Coefficient (JSC) in [54] or Similarity Index (SI) in [110], is another common similarity index presented by Jaccard [74]. It is defined as the ratio of the area of the mass automatically segmented to the area of the mass segmented manually by an experienced radiologist (Ground Truth (GT)). This measure is defined as

$$AO = \frac{TP}{TP + FP + FN} \tag{2.4}$$

and is related to the DSC as

$$DSC = \frac{2 \cdot AO}{1 + AO} \tag{2.5}$$

Notice that the DSC is expected to be greater than the AO. The AO metric has been used to assess the following works: [4, 28, 30, 54, 59, 64, 66, 86, 92, 102, 110, 134].

True-Positive Rate (TPR), also known as sensitivity [102] or overlap fraction (OF) [71], measures the number of pixels correctly labeled as lesion with respect to the area of the lesion reference

$$TPR = \frac{TP}{TP + FN} \tag{2.6}$$

The TPR metric has been used to assess the following works: [69, 71, 67, 84, 86, 88, 110, 131].

False-Positive Rate (FPR) corresponds to the number of pixels wrongly labeled as lesion with respect to the area of the lesion reference as expressed in

$$FPR = \frac{FP}{TP + FN} \tag{2.7}$$

The FPR metric has been used to assess the following works: [67, 84, 86, 88, 110, 131]. Notice that the FPR calculated as in equation 2.7 differs from the classic FPR obtained in

Table 2.1, which corresponds to the ratio between FP and the total number of negatives (FP + TN).

False-Negative Rate (FNR)) is the number of pixels belonging to the actual lesion delineation wrongly labeled as background as expressed in

$$FNR = \frac{FN}{TP + FN} \tag{2.8}$$

The FNR metric has been used to assess the following works: [67, 84, 88, 131].

Specificity measures the proportion of negatives correctly identified. Specificity is described as

$$SPEC = \frac{TN}{TN + FP} \tag{2.9}$$

and is usually given as a complementary information on the sensitivity (TPR). The specificity index is also used to assess the work in [102].

Positive Predictive Value (PPV) is the probability that the pixel is well classified when restricted to those pixels that test positive. It is computed as in equation 2.10. The PPV metric has been used in [102].

$$PPV = \frac{TP}{TP + FP} \tag{2.10}$$

Normalized Residual Value (NRV), also found as the Precision Ratio (PR) [68], corresponds to the residual area between the segmentation (S) and the reference (R) regularized by the size of the reference delineation as described in

$$NRV = \frac{Area(S \oplus R)}{Area(S)} \tag{2.11}$$

where $\oplus$ represents an exclusive OR operation. The NRV metric has been used to assess the following works: [59, 68].

Distance based metrics

Distance based metrics assess the displacement between the segmentation and the reference instead of comparing the resulting overlap. As in area metrics, multiple distance criteria arise in the literature. Most of the boundary metrics are physical quantitative error measures that are assumed to be reported in pixels.

Most of the works base their similitude indexes on the analysis of the Minimum Distance (MD) coefficients. The MD corresponds to the minimum distance between a particular point s_i of the segmentation boundary S and any other point r_j within the reference delineation R, and is defined as

$$\mathrm{MD}(s_i, R) = \min_{r_j \in R} \|s_i - r_j\| \tag{2.12}$$

Hausdorff Distance (HD) measures the maximum distance between the two delineations S and R as defined in

$$\mathrm{HD}(S, R) = \max \left\{ \max_{s_i \in S} \mathrm{MD}(s_i, R), \max_{r_i \in R} \mathrm{MD}(r_i, S) \right\} \tag{2.13}$$

This definition of HD is used in [54, 55]. However, in other works, such as Madabhushi and Metaxas [88] and Shan et al. [110], only the error between the assessed delineation S and the reference delineation R is considered. Here it is denoted as HD'.

$$\mathrm{HD'}(S, R) = \max_{s_i \in S} \mathrm{MD}(s_i, R) \tag{2.14}$$

Average Radial Error (ARE) was proposed by Huang et al. [67], and is defined as follows

$$ARE = \frac{1}{n} \sum_{i=1}^{n} \frac{|S(i) - R(i)|}{|R(i) - R_0|} \tag{2.15}$$

A set of n radial rays are generated from the center of the reference delineation R_0 intersecting both delineations. The ARE index consists of averaging the ratio between the distance of the two outlines $|S(i) - R(i)|$, and the distance between the reference outline and its center $|R(i) - R_0|$.

Average Minimum Euclidian Distance (AMED) is the average MD between the two outlines [54, 55], and it is defined as

$$\text{AMED}(S,R) = \frac{1}{2}\left[\frac{\sum_{s_i \in S} \text{MD}(s_i, R)}{|S|} + \frac{\sum_{r_i \in R} \text{MD}(r_i, S)}{|R|}\right] \tag{2.16}$$

where $|S|$ and $|R|$ are the number of points in the segmentation and reference boundary, repectively. As in the definition of Hausdorff Distance (HD), in Madabhushi and Metaxas [88] and Shan et al. [110] only the error between the assessed delineation S and reference delineation R is considered. Here it is denoted as AMED'

$$\text{AMED'}(S,R) = \frac{\sum_{s_i \in S} \text{MD}(s_i R)}{|S|} \tag{2.17}$$

The AMED measure also appears in some works under the names of Mean Error [88] and Mean absolute Distance [110].

Proportional Distance (PD) regularizes the AMED distance with the area of the reference delineation according to equation 2.18. It is used in [4, 59].

$$\text{PD}(S,R) = \frac{1}{2\sqrt{\frac{Area(R)}{\pi}}} \cdot \left[\frac{\sum_{s_i \in S} \text{MD}(s_i, R)}{|S|} + \frac{\sum_{r_i \in R} \text{MD}(r_i, S)}{|R|}\right] * 100 \tag{2.18}$$

Other metrics. Zhang et al. [134] proposed using average contour-to-contour distance (Ecc) for assessing their work, and Chiang et al. [29] the use of Computer-to-Observer Distance (COD). However, no definition or reference can be found in their work. Cheng et al. [28] used a distance metric called the Williams Index to deal with multiple reference segmentations as defined in

$$\text{WI}(S,O) = \frac{\frac{1}{n}\sum_{j=1}^{n} 1/D_{S,O_j}}{\frac{2}{n(n-2)}\sum_{i=1}^{n}\sum_{j=1(j\neq i)}^{n} 1/D_{O_i,O_j}} \tag{2.19}$$

where O_j denotes the jth set of manual delineations, n is the number of manual delineations for each lesion, and $D_{A,B}$ is the average distance between the corresponding pair of boundaries in sets A and B.

2.3.2 Histogram thresholding

Histogram thresholding is one of the most widely used techniques for gray level image segmentation. In a histogram thresholding method, an intensity threshold is chosen in the valley of the image histogram to separate the image into background and foreground.

Horsch et al. [66] proposed the use of a segmentation algorithm based on maximizing a Gaussian constraint function over partition margins defined through gray-value thresholding of a preprocessed image, in which the visualization of the mass structures were enhanced. The final segmentation consists of finding the margin that maximizes the Average Radial Derivative (ARD) measure. However, the center, width, and height of the lesion need to be selected manually. The segmentation performance was tested on a 400-image dataset achieving a mean AO of 0.73, comparing the segmentation results with delineations of an expert radiologist. Massich et al. [92] proposed a similar methodology, where the final margin is chosen using a disparity measure instead of maximizing the ARD function. They evaluated the segmentation performance in a small dataset of 25 images obtaining a mean AO of 0.64. However, they reduced the user interaction to a single click.

Yeh et al. [131] proposed an iterative disk expansion methodology. The authors showed that contour extraction of breast lesions in ultrasound images can be achieved by removing the speckle noise after image thresholding. This methodology can be divided into three main parts: (1) an adaptive thresholding to convert an ultrasound B-mode image into a binary image, (2) a disk expansion to extract the significant objects, and finally, (3) a refinement of the the extracted object to obtain the more accurate lesion boundary. The method was evaluated on only 4 clinical cases, reporting a mean normalized TPR over 90%. However, the method needs human interaction to provide the region of interest.

Histogram thresholding methods are fast and easy to implement, but do not perform well for very noisy images or images with a high influence of inherent artifacts in sonography. However, histogram thresholding methods have been widely used to obtain a rough segmentation to initialize more complex techniques such as active contour models or graph-based methods [71, 88].

2.3.3 Active Contour Models (ACM)

Active Contour Models (ACM), also known as snakes, is a framework for delineating an object boundary from a 2D image, and has been massively used as an edge-based segmentation method [4, 30, 54, 69, 71, 84, 86, 88] in BUS B-mode images. This approach

attempts to minimize the energy associated with the current contour as the sum of the external and internal energies. During the deformation process, the forces are calculated from the internal and external energy. The external energy is used to adapt the contour of the desired object boundary, and the internal energy is used to control the shape and regularity of the contour.

Combining intensity, texture, and directional gradient, a deformable shape-based model was presented by Madabhushi and Metaxas [88] to find lesion boundaries automatically. In their method, a set of empirical rules used by radiologists in ultrasonic breast lesion detection is employed to determine a seed point in the image automatically, indicating the lesion location. An initial segmentation of the lesion is obtained classifying the image pixels according to the intensity and texture, followed by region growing. Boundary points are found using the directional gradient of the image. These boundary points are supplied as the initial shape of a deformable model. This method does not need any manual initialization of the contour. A dataset of 42 images was used in order to evaluate the methodology in terms of boundary error (6.6 pixels), TPR (75%), FPR (20.9%), and FNR (25%).

A watershed transform was used by Huang et al. [69] to generate an over-segmentation of the image automatically. The regions are then merged depending on the region intensities and texture features to obtain a rough tumor shape and boundary delineation. Next, ACM automatically determines the contours of the tumor. They evaluated the methodology using a dataset of 20 images achieving a SI (called DSC in other works) of 0.88 and a TPR of 81%. Later in [71], they proposed to preprocess the image using a modification of the anisotropic diffusion method. Then, histogram thresholding was used to generate an initial contour for the level set procedure. A DSC of 0.88 and a TPR of 85.7% were reported using a dataset of 118 images.

Alemán-Flores et al. [4] used a geodesic active contour method. They first preprocess the image using an anisotropic diffusion method based on a texture description provided by a set of Gabor filters. Then, a manual seed initializes a snake to produce an initial segmentation, which is used to initialize a geodesic snake ACM using intensity information of the inner and outer parts of the lesion. A dataset of 32 images with 4 ground truth delineations provided by 2 radiologists was used to evaluate their results, achieving an AO of 88.3% and a PD of 4.96 pixels. In a similar way, Cui et al. [30] presented a two-stage active contour method. The initial contour was based on a manually identified point approximately at the lesion center. The two-stage active contour method iteratively refines the initial contour and performs a correction on the segmentation result. A mean AO of

0.74 was reported on a large dataset of 488 images from 250 patients.

Liu et al. [84] proposed an automatic lesion segmentation algorithm using a level set-based method, combining both global statistical information and local edge information. The global information is used to model the statistical information of speckle patterns to handle the noise and undefined boundaries, while the local information is extracted from the edges of the lesion. A dataset of 103 images was used in order to evaluate the methodology in terms of TPR (91.31%), FPR (7.26%), and FNR (8.69%). Later, in [86], they proposed a level set-based active contour model obtained by fitting a Rayleigh distribution to training lesion samples. The level set evolves then to fit the model into the target image. The level set initialization corresponds to a centered rectangle of one third the size of the target image. Despite its naive initialization, the reported average AO using a dataset of 76 images is 0.88.

Recently, Gao et al. [54] combined an edge stopping term taking into account phase congruency texture and a modified Gradient Vector Flow (GVF) in a level set-based framework. They evaluated their proposal in a small dataset of 20 images achieving a mean AO of 0.863.

Although accurate segmentation results are reported using ACM, they are largely linked to the active contour initialization. Some methodologies work directly with ROIs to avoid this problem, while others use simpler segmentation methods to obtain an initial rough contour. Furthermore, the snake-deformation procedure is very time-consuming.

2.3.4 Markov Random Fields (MRF)

The segmentation problem can be seen as a labeling problem consisting of assigning a set of labels to pixels. This is a natural representation for MRF [125]. This methodology alternatively approximates the maximization of the posterior estimation of the class labels and estimates the class parameters. A MRF model deals with the spatial relations between the labels obtained in an iterative segmentation process. The process of assigning pixel labels iteratively can be achieved by maximizing either a posteriori estimation or a posterior marginal estimation.

For instance, Xiao et al. [129] defines the ultrasound image as a multiplicative model in which one of the components is a distortion field. It uses a combination of the Maximum A Posteriori (MAP) and MRF to estimate that distortion field while labeling image regions based on the corrected intensity statistics. The MAP is used to estimate the intensity

model parameters while the MRF provides a way of incorporating the distributions of tissue classes with a spatial smoothness constraint. The method is implemented iteratively in an Expectation-Maximization (EM) approach, and the distributions of tissue classes are defined as a Gaussian distributions. The method is only qualitatively evaluated in a reduced set of synthetic and real data, and one of the main drawbacks is the user interaction. The user is required to determine different ROIs placed inside and outside the lesion in order to extract the intensity distribution in both regions.

MRF is also used by Boukerroui et al. [20] to model the segmentation process, and to focus on the adaptive characteristics of the algorithm. They introduced a new function to control the properties of the segmentation process, taking local and global statistics into account. The performance of the algorithm is demonstrated on synthetic data, a BUS image and on echocardiographic sequences. No quantitative results were reported regarding breast sonography.

The merit of MRF modeling is that it takes advantage of the pixel neighboring correlations. However, its iterative process is in general complex (easy to get stuck in local minima states) and time-consuming.

2.3.5 Machine learning

When addressing the lesion segmentation problem, machine learning is also a useful and reliable alternative. Machine learning takes advantage of ground truth data to build up a model for predicting or inferring the nature of elements with no such GT provided. Thus, these models generated from a training procedure can be used to drive a segmentation procedure.

Zhang et al. [134] proposed using a two-step machine learning procedure. First, a supervised machine learning for lesion detection is performed using a PBT. Detected regions with high confidence of being lesion or non-lesion are further used to learn the appearance model of the lesion within the target image. The second step consists of a supervised machine learning segmentation procedure trained on the target image using the previous detected regions. They evaluated their proposal in a dataset of 347 images, achieving an AO value of 84%, and an average contour-to-contour distance of 3.75 pixels. However, they only used 90% of the segmentations to perform the segmentation assessment, arguing that the remaining segmentations suffered poor detection, and that segmentation results assessment should not be subject to wrong initializations.

Later, Shan et al. [110] proposed using the lesion detection to initialize a region growing procedure to obtain a rough segmentation and, hence, generate a ROI automatically. Then a database driven supervised machine learning segmentation procedure is carried out in the ROI to determine a lesion/non-lesion label for all the pixels. The segmentation stage takes advantage of intensity, texture, energy-based phase information, and distance to the initially detected contour as features for the training process using an artificial Neural Network (NN). They assessed the segmentation performance using a dataset of 120 images, obtaining a SI (DSC in other works) of 0.83, a TPR of 92.8% and a FPR of 12%.

Hao et al. [64] proposed segmenting automatically breast lesions using an objective function combining a machine learning technique (Deformable Part Models (DPM)) [48] with intensity histograms, texture descriptors derived from grey-level co-occurrence matrix and position information using a graph-cut minimization tool and normalized cuts [112] as image segments. They evaluated the segmentation performance in a large dataset of 480 images achieving, an AO of 75%.

For machine learning methods, feature selection and training processes are two important steps highly influence the segmentation results. If the extracted features are sufficiently distinguishable and the method is well trained, machine learning methods can generate satisfactory lesion segmentations. However, over-training or insufficient training may severely affect the segmentation performance on new data. Note that the training process is usually time-consuming, but is often performed offline.

2.3.6 Watershed transformation

The intuitive idea underlying this technique comes from topography. Any gray-level image can be considered as a topographic surface. If we flood this surface from its minima and prevent the merging of the waters coming from different sources, the image is partitioned into two different sets: the catchment basins and the watershed lines. If we apply the watershed transformation to the image gradient, the catchment basins should theoretically correspond to the homogeneous gray level regions of the image.

Huang et al. [68] integrates a Self Organizing Map (SOM) NN classification with morphological watershed segmentation. Texture features are employed to yield inputs to the SOM NN. Finally, the watershed transformation automatically determines the contours of the tumor. The method was evaluated in a dataset of 20 images achieving a PR (NRV in other works) of 0.82 and a match rate (TPR in other works) of 95%.

Another work based on watershed is the work proposed by Gómez et al. [59] where a watershed transform is used to condition a Gaussian constraining function. As in the Horsch et al. [66] proposal, ARD maximization is used in order to find the adequate threshold that leads to the final segmentation. The method deals with ROIs instead of the entire image. A mean AO of 0.85 is reported using a 50 image dataset.

Recently, Cheng et al. [28] presented a two-step perceptual organization process to segment a breast lesion in sonography. Initially, a two-pass watershed transformation is performed to tessellate a ROI into a cell structure. Then, a cell competition process is carried out by allowing merge and split operations of cells. The method then performs a contour grouping to organize the edge entities found in the first step. Their method was evaluated in a 324 image dataset obtaining an AO of 92.4%.

In general, watershed-based methodologies are used for fast segmentation, but they face the problem of oversegmentation, and are sensitive to the initialization of the method.

2.3.7 Graph-based methods

Graph partitioning methods can effectively be used for image segmentation. In these methods, the image is modeled as a weighted, undirected graph. Usually a pixel or a group of pixels are associated with nodes, and edge weights define the (dis)similarity between the neighborhood pixels. The graph (image) is then partitioned according to a criterion designed to model "good" clusters. Each partition of the nodes (pixels) output from these algorithms are considered to be an object segment in the image.

Chiang et al. [29] proposed a graph-cut segmentation method that transforms prominent regions within a given ROI into graph nodes instead of pixels. These prominent regions are obtained by using a two-pass watershed transformation. The method was evaluated in a small dataset of 16 images obtaining a value of COD of 2.71 pixels.

Later, Jiang et al. [75] proposed the use of a graph-based segmentation algorithm (random walks) [61] to retrieve the boundaries of previously detected tumor regions. The detection process is performed in a two-stage classification. First, the method uses an AdaBoost classifier on Haar-like features to detect a preliminary set of tumor regions. These regions are further classified with a SVM using quantized intensity features. However, the segmentation step of their proposed method was not evaluated quantitatively.

Huang et al. [67] presented a segmentation algorithm that constructs a graph using improved neighborhood models. They also proposed a pairwise region comparison predicate

to determine the mergence of adjacent subregions. This predicate takes into account the local statistics and the measures of Signal to Noise Ratio (SNR) in US images to make the segmentation insensitive to noise. They evaluated the method in a dataset of 10 benign and 10 cancerous images. For the benign lesions, they achieved a TPR of 87.4%, a FPR of 1.6%, and an ARE of 9.2 pixels. For the cancerous lesions, a TPR of 87.6%, a FPR of 1.8%, and an ARE of 8.1 pixels.

Contemporarily, Gao et al. [55] proposed the use of a normalized cuts framework [112] with textural features extracted from homogeneous patches of the image. The algorithm was evaluated in a dataset of 100 breast US images (50 benign and 50 malignant). The mean HD measure, the AMED measure and the AO measure achieved 7.1 pixels, 1.58 pixels, and 86.67%, respectively, for benign tumors whereas 10.57 pixels, 1.98 pixels, and 84.41%, respectively was achieved, for malignant tumors. Their algorithm requires manual specification of a ROI.

The use of graph-based methodologies has increased over the last few years. By analyzing the graph-based current methodologies we can see that it obtains good results, although some kind of user interaction is needed in most of the algorithms.

2.3.8 The role of user interaction in lesion segmentation methods

Since segmentation algorithms are designed to be integrated with a CAD system, the degree of user interaction involved in the segmentation procedure is an important factor to take into consideration. Here, we analyze the segmentation methods regarding their degree of automatization: semi-automatic or fully-automatic.

Semi-automatic segmentation

Semi-automatic segmentation methods require some degree of user interaction. Commonly, the user constrains or initializes the segmentation procedure by indicating parts or elements belonging to each object to be segmented (i.e. foreground/background). This is the case of Xiao et al. [129], where the user defines two windows (one for the lesion, one for the background) to analyze the histograms of these regions. Other works such as [28, 29, 54, 55, 59, 68, 131], require the specification of a ROI that envelops the lesion. Finally, the semi-automatic methodologies which need less interaction are those that only require a seed point inside the lesion [4, 30, 66, 92, 102]. Finally, other works such as Huang et al. [67] proposed a segmentation method that needs the user to tune some parameters to

achieve a good final result.

Nowadays, semi-automatic methods play an important role in CAD systems since the fully-automatic segmentation problem needs to achieve a better performance. It is clear that manual delineations are unacceptable in a clinical environment due to the amount of time and effort needed, as well as the huge inter- and intra-user variability. Hence, semi-automatic methods have become a trade-off that reduces the inherent problems in manual segmentation by taking advantage of the user to assist the segmentation procedure.

Fully-automatic segmentation

Fully-automatic segmentation methods perform the segmentation procedure without any user interaction. In this category, two distinct strategies have been adopted. First, there are methodologies that automatize semi-automatic procedures so that no user interaction is required, commonly including a lesion detection procedure [71, 75, 88, 110, 134]. And second, there are ad-hoc methodologies designed in a manner that can not be supplied by user information [20, 64, 68, 69, 67, 84, 86, 131].

Although automatic segmentation of breast lesions in sonography needs to achieve a better performance, huge efforts to obtain lesion delineations with no user interaction have been made in the last few years.

2.3.9 Summary

We have presented and reviewed different approaches to the automatic and semi-automatic segmentation of lesions in ultrasound images. We have described several algorithms, pointing out their main features. Special emphasis has been put on the different strategies and a classification of these techniques has been proposed. In each method description, we have outlined the main advantages and disadvantages.

Table 2.3 summarizes the analysis of the principal methods discussed in this section. The methods are first sorted by the method used (histogram thresholding, active contours, MRF, machine learning, watersheds, and graph-based), and then by the year of publication (ascending order).

In order to compare the different segmentation methods more clearly, Table 2.4 provides descriptions, advantages and disadvantages of the different methods at a glance.

Analyzing Table 2.3, we notice that, as in lesion detection, the introduction of the

Table 2.3: Summary of the results presented in the articles analyzed for lesion segmentation.

Reference	Method	User interaction	Measure	Result	# cases (# images)
Horsch (2001) [66]	Histogram thresh.	Seed point	AO	94%	400(757)
Yeh (2009) [131]	Histogram thresh.	ROI specified	TPR	90%	4(4)
			FNR	15%	
			FPR	16%	
Massich (2010) [92]	Histogram thresh.	Seed point	AO	64%	25(25)
Madabhushi and Metaxas (2003) [88]	Active contours	-	TPR	75.1%	42(42)
			FNR	25%	
			FPR	20%	
			HD	19.72 pix.	
			AMED	6.6 pix.	
Huang (2005) [69]	Active contours	-	DSC	0.88	20(20)
			TPR	81%	
Huang (2007) [71]	Active contours	-	DSC	0.87	118(118)
			TPR	85.7%	
Alemán Flores (2007) [4]	Active contours	Seed point	AO	88.3%	32(32)
			PD	4.96 pix.	
Cui (2009) [30]	Active contours	Seed point	AO	74.5%	250(488)
Liu (2009) [84]	Active contours	-	TPR	91.3%	103(103)
			FNR	8.7%	
			FPR	7.3%	
Liu (2010) [86]	Active contours	-	DSC	0.88	46(46)
			TPR	93.9%	
			FPR	6.9%	
Gao (2012) [54]	Active contours	ROI specified	DSC	0.93	20(20)
			OA	86%	
			HD	7 pix.	
			AMED	2 pix.	
Xiao (2002) [129]	MRF	Histogram analysis	-	-	-
Boukerroui (2006) [20]	MRF	-	-	-	-
Zhang (2010) [134]	Machine learning	-	AO	84%	347(347)
			Ecc	3.75 pix.	
Hao (2012) [64]	Machine learning	-	AO	75%	480(480)
Shan (2012) [110]	Machine learning	-	DSC	0.83	120(120)
			TPR	92.8%	
			FPR	12%	
Huang (2004) [68]	Watersheds	ROI specified	TPR	94.6%	20(20)
			NRV	81.7%	
Gómez (2010) [59]	Watersheds	ROI specified	AO	86%	50(50)
			NRV	16%	
			PD	6.58 pix.	
Cheng (2010) [28]	Watersheds	ROI specified	AO	92.4%	324(324)
			WI	1.07 pix.	
Chiang (2010) [29]	Graph-based	ROI specified	COD	2.71	16(16)
Jiang (2012) [75]	Graph-based	-	-	-	-
Huang (2012) [67]	Graph-based	Parameter select.	TPR	87%	10(10) benign
			FPR	2%	
			FNR	13%	
			ARE	9.2 pix.	
			TPR	88%	10(10) cancerous
			FPR	2%	
			FNR	13%	
			ARE	8.1 pix.	
Gao (2012) [55]	Graph-based	ROI specified	AO	86.6%	50(50) benign
			HD	7.1 pix.	
			AMED	1.58 pix.	
			AO	84.4%	50(50) cancerous
			HD	10.57 pix.	
			AMED	1.98 pix.	

Table 2.4: Advantages and disadvantages of the segmentation techniques reviewed.

Method	Description	Advantages	Disadvantages
Histogram thresholding	Threshold value is selected to segment the image.	Simple and fast.	No good results for images with non-bimodal histograms.
Active contour model	Snake-deformation model is utilized.	It ensures closed region boundaries.	Slow in the iteration process. Pre-labeled ROI or initial contour is required. Complex definition of internal and external energies.
Markov random fields	Probabilistic model which captures pixel neighborhood constraints.	Smooth and accurate segmentation.	Complex (easy to get stuck in local minima states) and time-consuming for many iterations.
Machine learning	Segmentation is regarded as a classification task.	Different lesion characteristics can be incorporated by feature extraction.	How to select the training set is problematic, and training is time-consuming and depends on the image database.
Watershed	Considers image as a topographic surface where the grey level of a pixel is interpreted as its altitude.	It ensures closed region boundaries.	Over-segmentation problem is not completely solved. Sensitive to the initialization.
Graph-based	The image is modeled as a weighted, undirected graph and is then partitioned according to a criterion designed to model good clusters.	Good approximation of the optimal segmentation can be computed very efficiently.	Requires user interaction.

machine learning segmentation in the state-of-the-art has been done in the last few years. The same trend can be observed with the graph-based methods, in which the oldest method

analyzed dates from 2010. Also, recent works using watershed-based segmentation methods were found in the literature, but only the work of Chen et al. [28] assessed the method in a large dataset. Active contour based methods have been widely used for the past ten years, and still are. They reported good results, although most have been tested in small datasets, and the correctness of the results are highly subject to the initialization of the contour. Histogram thresholding is not a feasible option nowadays due to the complexity of current US images, in which the tissue structures are displayed with higher definition and detail. Finally, it is not possible to extract any trend from MRF segmentation methods due to the fact that any of these methods presented quantitative results.

We have also analyzed the segmentation methods regarding the user interaction involved. We consider this as a complementary information that needs to be taken into account in the comparison. For example, it is clear that a method that segments a given specified ROI would have to obtain a better performance than methods dealing with the entire image because in a ROI most of the artifacts are avoided. Analysis of Table 2.3 shows that, among the most contemporary works, machine learning-based methods are fully automatic and obtained good results, while graph-based methods require user interaction.

2.4 A review of lesion classification methods on breast sonography

Finally, an important stage in a CAD system is the classification of the detected lesions. Needless to say, the relevance in any diagnosis to determine if a located lesion is benign or malignant is uppermost. Hence, the CAD system extracts features from the previously segmented region to train a classifier to distinguish between cancerous and non-cancerous abnormalities with a certain likelihood and, thus, helps radiologists in their diagnosis.

The main goal of this section is to review the most relevant works in lesion classification found in the literature. These lesion classification methods can be grouped regarding the classifier used to distinguish between cancerous and benign lesions.

2.4.1 Linear classifiers (LC)

Linear Classifier (LC) are easy and fast to implement due to the use of a discriminative lineal function. Horsch et al. [65] used the segmentation results [66] to extract features such as lesion shape, margin sharpness, echogenic texture, posterior acoustic enhancement,

and shadowing, and then trained a LC to distinguish cancerous lesions from benign ones. They achieved an A_z value of 0.87 using a dataset of 757 images from 400 different cases, in which 94 were malignant.

Some works do not specify the segmentation method used, and focus the work on the features used to classify the lesions. This is the case of Seghal et al. [107], who used the margin sharpness, margin echogenicity, and angular variation in the margin to classify lesions in a dataset of 56 images, in which 36 were benign and 20 were malignant, obtaining an A_z value of 0.87.

Shen et al. [111] also focused on the extracted features. The segmentation of the lesions was performed consensually by two expert radiologists. Subsequently, shape, orientation, margin, lesion boundary, echo pattern, and posterior acoustic features are used to discriminate between malignant and benign lesions. They achieved an A_z value of 0.97 with a dataset of 265 images, including 180 benign and 85 malignant masses.

Finally, within the methods that used a LC, the recent work by Alvarenga et al. [6] should be mentioned. They investigated the combination of morphological and texture parameters to distinguish between malignant and benign breast tumors. First, radiologists determine the ROI in the image, and then segmentation is achieved [115]. An exhaustive analysis for 27 different features and their combinations is then performed. The best-performance was obtained with the combination of two morphological and three texture features, resulting in an A_z value of 0.87 in a dataset of 246 images from 197 patients, containing 177 malignant and 69 benign tumors.

2.4.2 Neural networks (NN)

Neural Network (NN) are inspired by the way biological nervous systems process information. They are usually presented as systems of interconnected neurons that can compute values from inputs by feeding information through the network. Chen, D.R. et al. [25] presented a lesion classification algorithm where initially an expert radiologist locates the ROI in the image, and then, a naive segmentation algorithm is performed. Cooperating with the segmentation algorithm, three feasible features, including variance contrast, autocorrelation contrast, and the distribution distortion of wavelet coefficients are extracted from the ROI images for further classification. A multilayered perceptron NN trained using an error back-propagation algorithm with momentum is then used for the classification of breast lesion. In the experiment, 242 cases including benign breast tumors from 161 patients and carcinomas from 82 patients were sampled to evaluate the performance, resulting in an A_z

value of 0.93.

Further, Chen, C.M. et al. [24] proposed a classification work using a multilayer feed-forward NN on the basis of novel morphologic features such as the number of substantial protuberances and depressions, lobulation index, elliptic-normalized circumference, elliptic-normalized skeleton, and long axis to short axis ratio. In order to obtain these features, the delineation of the tumors was performed by expert radiologists. The results were assessed in a dataset of 271 images (140 with malignant lesions) reaching an A_z value of 0.96.

Joo et al. [77] determined whether a breast mass is benign or malignant by extracting features with a naive segmentation algorithm applied to a manually preselected ROI. A NN then distinguished malignant lesions based on five morphological features representing the shape, edge characteristics, and darkness of the mass. They assessed the classification in a dataset of 584 images containing 300 benign and 284 malignant breast lesions, obtaining an A_z value of 0.95.

Finally, the work by Drukker et al. [37], which has already been presented in section 2.2.2, proposed the classification of the segmented regions using a Bayesian Neural Network (BNN) on the basis of the extracted image features. The features used in their work are the same as those used in Horsch et al. [65]. They used two different datasets, one for training and one for testing. The first dataset was composed of 1740 images from 458 patients (23 with cancerous lesions, 204 benign ones and 231 patients without abnormalities), and the second by 151 images, all from different patients (45 with cancerous lesions, 100 benign ones and 6 patients without abnormalities). They reported A_z values of 0.81 and 0.86, distinguishing between benign and malign lesions for the two datasets.

2.4.3 Support vector machines (SVM)

Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane in a multidimensional space. Given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane that categorizes new examples. Huang et al. [69] proposed the use of a SVM to classify features extracted from a ROI specified by an expert radiologist. They used the correlation between neighboring pixels in the images as features to classify breast tumors. Two different datasets were used to assess the proposal. First, a dataset of 140 images (88 with benign and 52 with malignant tumors) achieved an A_z value of 0.96. Second, a dataset of 250 images (215 with benign and 35 with malignant tumors) achieved an A_z value of 0.95.

Later in [72], they outperformed the results using different texture features, such as block difference of inverse probabilities, block variation of local correlation coefficients, and auto-covariance matrix, to train the SVM classifier. They obtained an A_z value of 0.96 using a dataset of 250 images.

2.4.4 Summary

A summary of the results reported by the principal methods discussed in this section is presented in Table 2.5. The methods are first sorted by the classifier used (linear classifier, neural networks, or support vector machines), and then by year of publication (in ascending order).

Table 2.5: Summary of the results presented in the articles analyzed for lesion classification.

Reference	Classifier	Training	User interaction	Features	Measure	Result	# cases (cancer/benign)
Horsch (2002) [65]	LC	Split dataset	-	Morphologic/ texture	A_z	0.87	400 (94/306)
Seghal (2004) [107]	LC	Leave-one-out	N/A	Morphologic/ texture	A_z	0.87	56 (20/36)
Shen (2007) [111]	LC	10-fold cross validation	Manual delineation	Morphologic/ texture	A_z	0.97	265 (85/180)
Alvarenga (2012) [6]	LC	Split dataset	ROI specified	Morphologic/ texture	A_z	0.87	197 (177/69)
Chen, D.R. (2002) [25]	NN	10-fold cross validation	ROI specified	Texture	A_z	0.93	242 (82/161)
Chen, C.M. (2003) [24]	NN	Split dataset	Manual delineation	Morphologic	A_z	0.96	271 (140/131)
Joo (2004) [77]	NN	10-fold cross validation	ROI specified	Morphologic	A_z	0.95	584 (284/300)
Drukker (2005) [37]	NN	Leave-one-out	-	Morphologic/ texture	A_z	0.81 0.86	458 (23/204) 151 (45/100)
Huang (2005) [69]	SVM	10-fold cross validation	ROI specified	Texture	A_z	0.96 0.95	140 (52/88) 250 (35/215)
Huang (2006) [72]	SVM	10-fold cross validation	ROI specified	Texture	A_z	0.96	250 (35/215)

Regarding the classification methods reviewed, it is not possible to conclude which classifiers work better than others as well as which are the best features to distinguish between malignant and benign tumors. However, both play a key role in the classification results. Analysis of Table 2.5 shows that half of the methods (5 out of 10) chose the use of combined texture and morphological features, while 3 used texture information alone, and only 2 proposals chose the use of morphological features alone. In most of the works, these features are extracted after a segmentation process requiring user interaction (definition of a ROI or manual delineation of the lesion).

In regard to the training and testing process, only 3 methods divide their dataset into training and evaluation subsets. The k-fold cross validation is the most used procedure (understanding the leave-one-out as a specific case where k is the number of images). Analyzing the composition of the datasets used, only 3 works trained and assessed their classification methods using a balanced dataset (in which the number of malignant and benign cases are approximately the same). Hence, we notice a trend to use datasets in relationship to the screening population (the number of benign findings is higher than malignant ones), except in the case of Alvarenga et al. [6], who used a dataset with more malignant than benign lesions.

Although the results are easily comparable due to the majority use of ROC analysis in the assessment, the lack of a common dataset makes the comparison impossible. For example, a dataset formed only of carcinomas and cysts would obtain better results (both type of lesions are highly distinguishable) than a dataset formed of carcinomas and solid benign masses (the lesions may look similar in some cases).

2.5 Conclusions

In this chapter we have performed a survey of the most important works in CAD sonography. However, an accurate and fair comparison of these works is not feasible. The major inconveniences are the lack of common assessing datasets and the inhomogeneity of the assessing criteria. Despite these inconveniences, some trends can be extracted from this survey.

In regard to the lesion detection methods listed in Table 2.2, a trend for the newest algorithms to use machine learning techniques should be noticed. One can justify this trend because of the adaptability of the machine learning-based methods to the specific characteristics of the target dataset, as well as their avoidance of the use of an extra false positive reduction step (necessary in most detection methods based on image processing).

In Table 2.3 we have reviewed the most relevant works in lesion segmentation. The newest algorithms tends to use machine learning or graph-based techniques. Machine learning methods obtained good results in large datasets and are fully automatic. On the other hand, all the graph-based methods studied were tested in small datasets and require user interaction. Watershed methods obtained good results, but they also require user interaction. We consider the intervention of a user in the method as complementary information that needs to be taken into account. For example, it is clear that a method that

segments a given specified ROI would have to obtain a better performance than methods dealing with the entire image because, in a ROI, most of the artifacts are avoided. Active contour based methods have been widely used for ten years, and still are because they report accurate results, even though their correctness is highly subject to the initialization process. Nowadays, histogram thresholding is mostly used for initial rough segmentation generation. The complexity of the current sonography, in which the tissue structures are displayed with higher definition and detail, makes it infeasible for an accurate segmentation. Finally, it is not possible to extract any trend from MRF segmentation methods due to the fact that any of these methods presented quantitative results.

Regarding the classification methods studied in this chapter (see Table 2.5), some factors play an important role in obtaining a better classification: the features, the classifier, the train/test procedure and the dataset. In terms of features, we can observe a trend for the classification methods to use combined texture and morphological features. In most of the works, these features are extracted after a segmentation process that requires user interaction (definition of a ROI or manual delineation of the lesion). In regard to the training and testing process, most of the methods used the k-fold cross validation (understanding the leave-one-out as a specific case where k is the number of images). Finally, analyzing the composition of the datasets, a trend can be noticed in the use of datasets in relationship with the screening population (the number of benign findings is higher than malignant ones).

3

Automatic breast lesion detection

3.1 Introduction

Detection of lesions in their early stages is a key factor in reducing the death rate. The earlier the disease is detected, a better and more effective treatment can be provided. Thus, one of the most important steps in CAD systems is the detection of the lesion. This process should ideally be unsupervised to help the radiologist make a diagnosis efficiently. For this reason, high sensitivity and specificity are required.

After the analysis performed in Chapter 2, we noticed a trend for the newest algorithms to use machine learning techniques due to the proved adaptability to the specific characteristics of a target dataset, as well as the good results obtained. Therefore, in this chapter we propose the use of a machine learning technique to detect lesions in sonography.

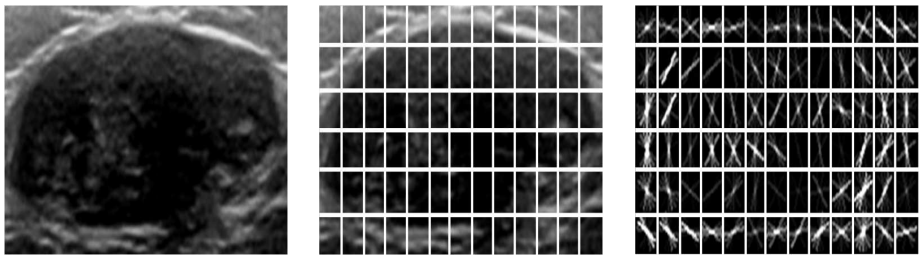
Lesion detection can be seen as a specification of the well-known object detection problem in computer vision. Thus, we propose to adapt a generic object detection technique to locate lesions in breast US images. Specifically, we propose the use of Deformable Part Models (DPM) presented by Felzenszwalb et al. [48], which was awarded the PASCAL VOC "Lifetime Achievement" Prize in 2010 [45]. The aim of the work presented in this

chapter is to adapt and evaluate the DPM approach to detect breast abnormalities in BUS B-mode images and compare the performance of this method with the most relevant proposals in the literature.

3.2 Deformable Part Models overview

The DPM approach [48] models the appearance of objects in terms of a root filter that approximately envelops the whole object, a set of part filters that cover smaller and representative parts of the object and deformation parameters penalizing the deviation of the parts from their default locations relative to the root using a latent SVM classifier.

The method uses a scanning window approach that searches a model over a Histogram of Oriented Gradients (HOG) pyramid [31] to detect objects in different scales. HOG are feature descriptors used in image processing for the purpose of object detection. The image is divided into small spatial rectangular regions (cells) and, in each cell, a weighted local 1-D histogram of gradient directions over the pixels of the cell is computed. Figure 3.1 shows an example of generating a HOG feature vector.



(a) (b) (c)

Figure 3.1: Example of generating a Histogram of Oriented Gradients (HOG) feature vector. (a) Original image of a breast lesion in a US image, (b) image divided into cells (c) resulting HOG descriptor for the image showing the gradient orientation histograms in each cell.

The HOG pyramid is defined by computing the HOG features at each level of an image pyramid. Hence, features at the top level capture coarse gradients, while at the bottom level, finer gradients are captured. The scale sampling in a feature pyramid is determined by a parameter λ defining the number of levels we need to go down in the pyramid to get a feature map computed at twice the resolution of another one. Figure 3.2 shows a representation of the feature pyramid detecting lesions in a breast US image. The feature pyramid is built via repeated smoothing and subsampling, and then computing a feature

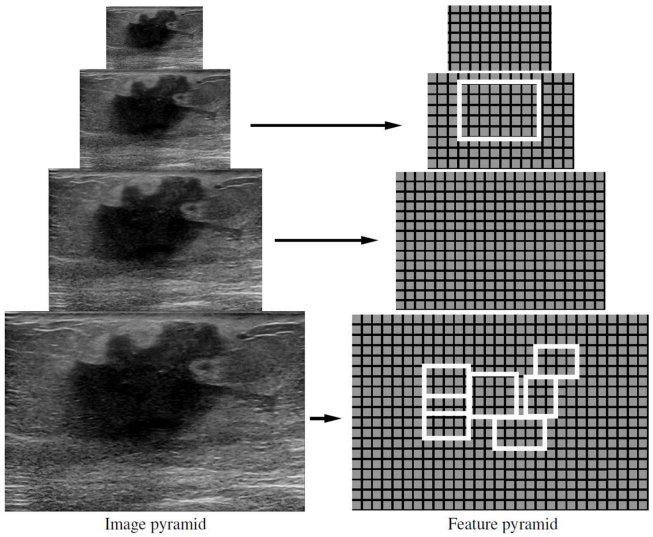


Figure 3.2: Representation of a feature pyramid detecting a lesion in a breast US image. The root filter is located at the top of the pyramid while the part filters are located at twice resolution of the placement of the root, with $\lambda = 2$.

map from each level of the image pyramid. In this example, the filters are displayed as white rectangles. Note that the part filters are placed at twice the spatial resolution of the placement of the root, with $\lambda = 2$.

A model is defined by a root filter that approximately envelops the whole object and part filters at twice the resolution of the root filter that cover smaller and more representative parts of the object. For instance, when creating a model for a bicycle detection, the root filter could capture coarse edges of the entire bicycle, while the part filters could capture details such as the wheels. Similarly, in breast US images, the root filter could capture coarse boundaries while part filters could be identified as salient regions of the lesion and internal structures. Figure 3.3 shows an example of root and part filters applied to a breast US lesion.

3.2.1 Formal definition of DPM

Both root and part filters are rectangular templates F of size $w \times h$ specifying weights[1] for subwindows of a HOG pyramid. Let H be a HOG pyramid and $p = (x, y, l)$ a location in the l-th level of that pyramid. The vector obtained by concatenating the HOG features in the $w \times h$ subwindow of H in p is defined as $\phi(H, p)$ and the score of F in this detection

[1]The filter weights are part of the model and will be obtained after a learning process.

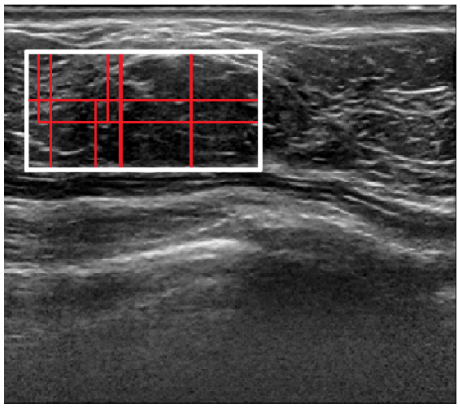
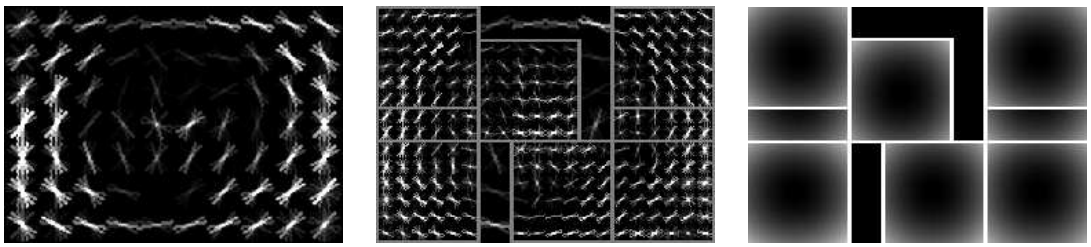


Figure 3.3: Example of DPM applied to a breast lesion. The white box represents the location of the root filter and the red boxes the location of the part filters.

window is $F \cdot \phi(H, p)$.

The model for an object with n parts is defined by a root filter F_0 and a set of parts $P_i = (F_i, v_i, d_i)$, where F_i is a filter for the i-th part, v_i is a two-dimensional vector specifying possible locations relative to the root, and d_i is a four-dimensional vector specifying coefficients of a quadratic function that defines a deformation cost for each possible placement of the part. A graphical representation of a lesion model is shown in Figure 3.4.



(a) (b) (c)

Figure 3.4: Lesion model. (a) is the root filter, (b) the part filters with twice the resolution and (c) shows the spatial deformation model. The filters visualization shows the positive weights at different orientations for the histogram of oriented gradients features in (a) and (b). The visualization of the spatial deformation model reflects the cost of placing the center of a part filter at different locations relative to the root filter in (c), where brighter areas represent high penalized placements.

The placement of the model is given by $z = (p_0, ..., p_n)$, where $p_i = (x_i, y_i, l_i)$ specifies the level and the position of the i-th filter. Note that the location of the root filter is defined when $i = 0$. The final score of a detection is the score of the root filter plus the

score of the best location of the parts, placed at twice the resolution in the pyramid, minus a deformation cost that penalizes undesired placements of the parts,

$$score(p_0, ..., p_n) = \sum_{i=0}^{n} F_i \cdot \phi(H, p_i) - \sum_{i=1}^{n} d_i \cdot \phi_d(dx_i, dy_i) + b \tag{3.1}$$

where

$$(dx_i, dy_i) = (x_i, y_i) - (2(x_0, y_0) + v_i \tag{3.2}$$

gives the displacement of the i-th part relative to the root location and

$$\phi_d(dx_i, dy_i) = (dx, dy, dx^2, dy^2) \tag{3.3}$$

are the deformation costs. Note that when $d_i = (0, 0, 1, 1)$ the deformation cost for the i-th part is the squared distance between its actual position and its anchor position relative to the root. The term b is the bias that makes the scores comparable in models with different components as we will further explain.

The score of a placement z can be expressed in terms of the dot product $\beta \cdot \psi(H, z)$, between a vector of model parameters β and a vector $\psi(H, z)$:

$$\begin{aligned} &\beta = (F_0, ..., F_n, d_1, ..., d_n, b) \\ &\psi(H, z) = (\phi(H, p_0), ..., \phi(H, p_n), \\ &-\phi_d(dx_1, dy_1), ..., -\phi_d(dx_n, dy_n), 1) \end{aligned} \tag{3.4}$$

Thus, each image example x is scored by a function of the form

$$f_\beta(x) = \max_{z \in Z(x)} \beta \cdot \Phi(x, z) \tag{3.5}$$

where $Z(x)$ is a range of valid placements for the root and part filters and $\Phi(x, z) = \psi(H, p_0)$. This representation is used for learning the model parameters using a latent SVM classifier. Formally, the classifier is trained by minimizing the following objective function [48],

$$\beta^*(D) = \frac{1}{2}||\beta||^2 + C \sum_{i=1}^{n} \max(0, 1 - y_i f_\beta(x_i)) \tag{3.6}$$

where D is a set of labeled examples $D = (\langle x_1, y_1\rangle, ..., \langle x_n, y_n\rangle)$, where $y_i \in \{-1, 1\}$ and x_i specifies a HOG pyramid $H(x_i)$ together with a range $Z(x_i)$ of valid placements for the root and part filters. $Z(x_i)$ is defined so the root filter must be placed to overlap the bounding box by at least a specified threshold t. The term $\max(0, 1 - y_i f_\beta(x_i))$ is the standard hinge loss and C a constant that controls the relative weight of the regularization term.

By restricting the latent domains $Z(x_i)$ to a single choice, f_β is linear in β, and a linear SVM classifier can be used. In practice, classical SVM training is applied iteratively to triples $(\langle x_1, z_1, y_1\rangle, ..., \langle x_n, z_n, y_n\rangle)$, where z_i is selected to be the best scoring latent label for x_i under the model trained in the previous iteration.

An object can be represented by a mixture model with m components defined by a m-tuple $M = (M_1, ..., M_m)$, where M_i is the model for the i-th component. A component is defined by a $(n + 2)$ tuple $(F_0, P_1, ..., P_n, b)$ where F_0 is a root filter, P_i the i-th part filter and b a bias term to make the scores of the different components comparable. A mixture model can capture different points of view of the same object making the detection process more robust. This is clear in regular object detection processes such as a bicycle, where one component could be a lateral view and another the front view. In the case of lesion detection, the angle of the transducer during acquisition could be perceived as the same problem. In addition, since masses have no standard shapes, the mixture of models will help the detector to model the most common shapes.

3.2.2 Implementation details

As stated above, in practice, when training a latent SVM, a classical SVM is applied iteratively training triples $(\langle x_1, z_1, y_1\rangle, ..., \langle x_n, z_n, y_n\rangle)$, where z_i is the best scoring latent label for the example x_i under the model trained in the previous iteration. Each of these triples leads to an example $\langle \Phi(x_i, z_i), y_i\rangle$ for training a linear SVM classifier.

Root filter initialization: The dimensions of the root filter are automatically selected by looking at the statistics of the bounding boxes in the training data. An initial root filter F_0 is trained with no latent variables. The positive examples are scaled to size and aspect ratio of the filter. Negative examples are generated from random subwindows in negative images.

Root filter update: Given the initial trained filter, the best-scoring placement for the filter (that overlaps with the bounding box more than a given threshold t) is found for each bounding box in the training set. Then, F_0 is retrained with the new positive set, iterating twice.

Part initialization: First, the area a of the part is defined so that $6a = 80\%$ of the F_0 area. Then, the rectangular region of area a of F_0 with the highest positive energy is selected, and the part is initialized from the subwindow values, but filled to handle the higher spatial resolution of the part. The initial deformation cost measures the squared norm of a displacement with $d_i = (0, 0, 1, 1)$. This procedure is repeated until the N parts are initialized.

Model update: To update the model, new training data triples are constructed. For each positive bounding box in the training data, the existing detector is applied at all positions and scales with an overlap of at least t with the given bounding box. Among these, the placements with the highest score are selected as positive examples. Negative examples are selected by finding high scoring detections in images not containing the target object. These negative examples are added to a cache of a limited size for computational reasons. A new model is then trained by running SVM with the new set of positive and negative examples. The model is iteratively updated, where in each iteration the correctly classified negative examples are removed from the cache and new negative examples are added. Towards the final iteration, the cache contains a set of hard negatives. This procedure is performed to deal with the vast number of negative examples in a training set.

3.3 Experimental results in BUS lesion detection

Various experiments were performed to evaluate the DPM approach for lesion detection. First, a default configuration of the DPM method was defined and evaluated. Posteriorly, this default configuration was changed to tune the parameters involved. The results of these parameters were analyzed to determine the best configuration of the DPM to detect lesions in breast US images. Subsequently, a post-processing step of false-positive reduction was analyzed using the best parameter configuration. In addition, an experiment to analyze the ability of the DPM to distinguish cancerous from benign lesions was also performed. Finally, the DPM was compared with the most relevant state-of-the-art methods using the

same dataset.

All the tests were performed on a PC (Intel® Core™ 2 Quad 2.83GHz 8GB RAM) using the implementation voc-release5 [57] of the DPM algorithm. All the experiments were refined using a non-maximum suppression post-processing to remove similar detections in a local area. Subsequently, the maximum number of detections per image was limited to the ten best detections in terms of score to avoid a large number of FPs per image but also to ensure that all the lesions could be detected. Only in Section 3.3.3, a different number of detections is chosen.

For the number of available images, we have configured the training and testing processes as a k-fold cross validation with $k = 10$. This methodology increases the computation costs heavily, but allows a more accurate assessment of the method.

3.3.1 Image acquisition

Two different datasets of BUS images, namely Dataset D1 and Dataset D2, were collected in this study. Both datasets were obtained from different US system specifications and time periods.

Dataset D1 was collected in 2001 from a professionally didactic media file for breast imaging specialists [103]. The images were obtained with B&K Medical Panther 2002 and B&K Medical Hawk 2102 US systems with an 8-12 MHz linear array transducer. The dataset consists of 406 images from different cases. From the 406 images, 306 images contain one or more lesions and 100 were acquired from healthy breasts. Within the lesion images, 60 images presented malignant masses and 246 were benign lesions. From the malignant images, 27 were diagnosed as Invasive Ductal Carcinoma (IDC), 4 were Ductal Carcinoma In Situ (DCIS), 6 were malignant phyllodes tumors and 23 were other unspecified malignant lesions. Of the benign images, 74 were complex cysts, 89 were simple cysts, 55 were Fibroadenoma (FA) and 28 were other benign lesions. The average size of the images is 377x396 pixels with a nominal pixel size of 0.098 mm.

Dataset D2 was collected more recently (2012/13) from the UDIAT Diagnostic Centre of the Parc Taulí Corporation, Sabadell (Spain) with a Siemens ACUSON Sequoia C512 system and a 17L5 HD linear array transducer (8.5 MHz). The dataset consists of 326 images from different patients, where 163 images present one or more lesions and 163 were acquired from normal breasts. In the 163 lesion images, 53 were images with cancerous masses and 110 with benign lesions. Of the malignant images, 40 were IDC, 4 were DCIS,

2 were Invasive Lobular Carcinoma (ILC) and 7 were other unspecified malignant lesions. Of the benign images, 65 were unspecified cysts, 39 were FA and 6 were other types of benign lesions. The average size of the images is 760x570 pixels with a nominal pixel size of 0.084 mm.

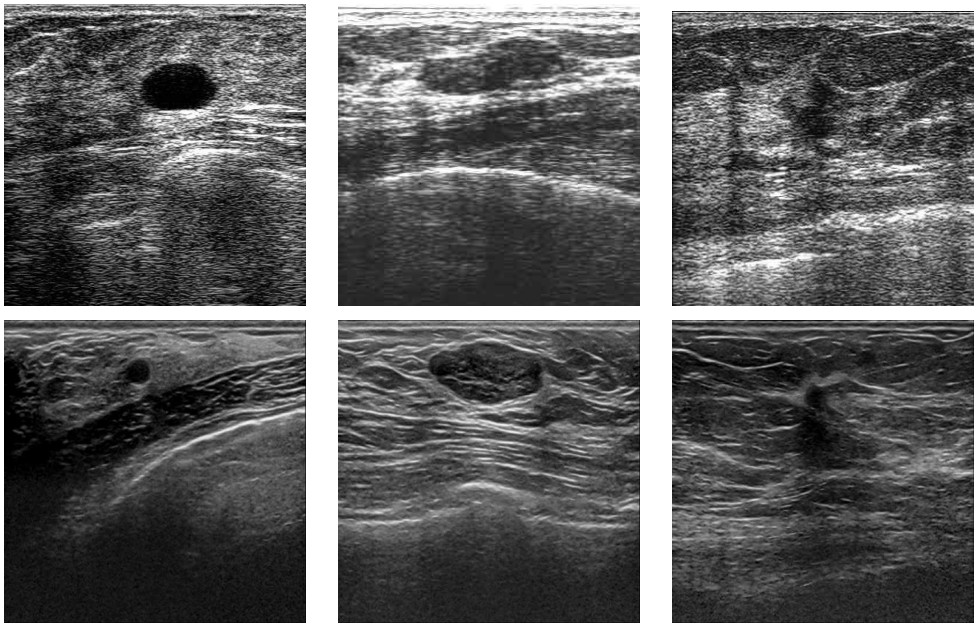
In both datasets the diagnoses were supported by a posterior biopsy/pathological examination after the acquisition. In all the images containing lesions, the lesions were delineated by an experienced radiologist. All of the images involved in this work were previously made anonymous to accomplish the rules issued by the Ethical Committee of both hospitals concerning the data confidentiality.

Figure 3.5 displays three images from the two datasets to represent the differences in three aspects: speckle noise, image quality and lesion appearance. In terms of speckle noise, images in Dataset D1 show a significant presence of this artefact but it is less obvious in images in Dataset D2, where the speckle noise was partly reduced by the US acquisition system. The image quality also varies in both datasets due to different resolutions. Note that the resolution for the recent US device used to produce Dataset D2 is better than in the old US device (Dataset D1). Consequently, the defined structures, such as ribs, pectoral muscle or parenchymal tissue, are more visible in Dataset D2. The lesion appearance also differs in the two datasets. In Dataset D2, the appearance of the tissues is better defined than in Dataset D1, as illustrated in Figure 3.5(b) where even the inner structures of the fibroadenoma lesion are visible.

3.3.2 Parameter analysis of DPM

Lesion detection and classification in breast B-mode US images is still a challenging problem in medical imaging due to the high variability of shapes among lesions of the same type (i.e. cysts, fibroadenomas, etc.). One could argue that the use of DPM does not properly fit the lesion detection problem due to the fact that breast lesion shapes present a large variability in comparison with detecting more structured objects such as bicycles or horses. In this work, we argue the opposite, that the deformability of the part filters in DPM and the possibility of defining a mixture of models allows us to cope with this variability of shapes.

Detecting objects with DPM involves different parameters which need to be set up empirically to improve the detection results. In this work, an exhaustive test of the main parameters of the DPM method has been performed in order to successfully detect lesions on BUS images. Specifically, the following parameters have been tested: the number of



(a) (b) (c)

Figure 3.5: Examples of images in Dataset D1 (first row) and Dataset D2 (second row). (a) shows an example of cyst images, (b) images with a fibroadenoma lesion and (c) examples of invasive ductal carcinoma.

components in a mixture model, the number of part filters, the size of the part filters, the overlap threshold for detecting positive examples in the learning stage, the cell size when computing the HOG features and the size of the HOG pyramid. Due to the vast number of possible combinations of these parameters, we have defined a default configuration, and only the value of one parameter at the time has been changed. The default configuration has been set up with: 3-component mixture model with 8 parts of 6×6 pixels, with a training threshold $t = 0.7$, a cell size of 8×8 pixels and a parameter $\lambda = 5$ that defines the size of the HOG pyramid. Due to computational reasons, only Dataset D2 has been used for assessment of the parameter estimation.

Evaluation criteria

The performance of each configuration was assessed using the ROC analysis [47]. For the lesion detection process, actual lesions correctly detected were counted as TP results while all the detections that did not locate an actual lesion were counted as FP results, and similarly for cancerous and non-cancerous lesions. The ROC analysis was performed on a "by region" basis [37], where each ROI is considered as an entity instead of the whole image,

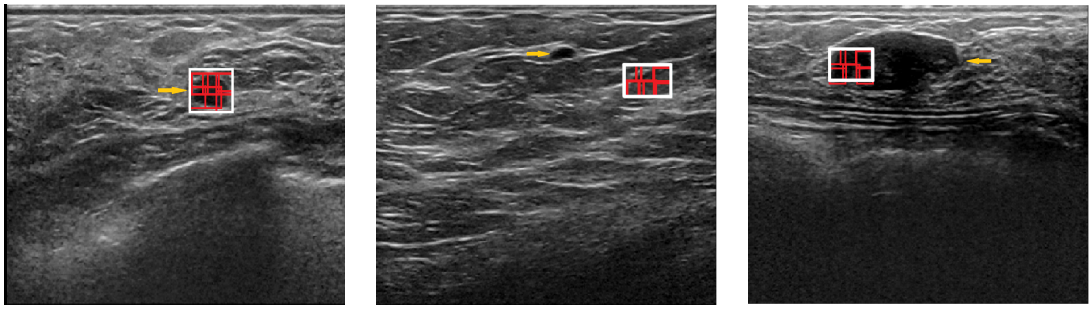
as in a typical ROC analysis. ROC analysis deals only with the classification process (i.e. how well the classifier is able to categorize a given detection) and we also want to analyze the performance of the detection process. For that reason, we also performed a FROC analysis in terms of Lesion Location Fraction (LLF) (also defined as sensitivity or TPR in other works) and Non-lesion Location Fraction (NLF) (number of FP detections per image) which assesses the overall process, including the detection of lesion candidates [22]. In the FROC analysis, a detection is considered as a TP result if the center of the resulting bounding box of the detection process is located within the delineation performed by a radiologist. Other detections are considered as a FP result. Note that in both analyses, only one TP per image is counted, while FP results are accumulative. ROC and FROC analysis were performed for each fold in the cross-validation procedure and mean values were computed. When analyzing FROC curves, the maximum number of sensitivity and the number of FPs per image at that point were reported.

To determine if there are significant differences in the performance between the different values of the parameters in comparison with the default configuration, a hypothesis test using the A_z, LLF and NLF values obtained for each configuration was performed. Initially, the Kolmogorov-Smirnov test [91] was used to confirm that the values we compared were normally distributed. Subsequently, a paired two-sample Student's t test [99] was then applied. The null hypothesis specifies that there are no significant differences between the mean values: $H_0 : \mu_1 \neq \mu_2$. Test results were provided in terms of p values, where a p value smaller than 0.05 indicates that the null hypothesis would be rejected.

Figure 3.6 shows a qualitative example of the results obtained. In Figure 3.6,(a) shows an image where the lesion was detected correctly, while an example of unsuccessful lesion identification is shown in (b). These FP detections generally occur when the lesion region is small. Finally, (c) shows a lesion that is successfully detected but the size of bounding box is not correctly estimated. We noticed that these partial detections occur when dealing with large lesions.

Model components

The first parameter we have evaluated is the number of components in the mixture model. Since mixture models deal with different views of the modeled object, it is necessary to set up the best number of components to obtain the best results. For instance, it is fair to assume that when detecting cars, the best number of configurations should be a model of two components: one for the lateral view and one for the frontal view. In the specific



(a) (b) (c)

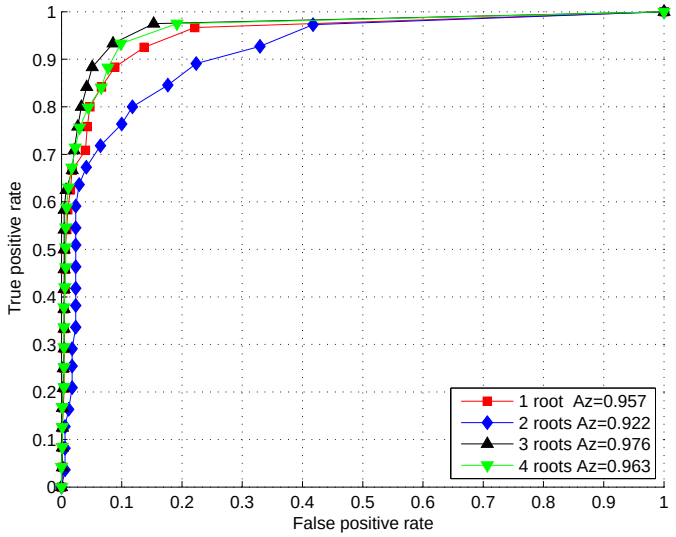
Figure 3.6: Examples of qualitative results: (a) correct lesion detection, (b) mis-detection and (c) partial detection where the lesion is detected but the estimation of the bounding box size is not correct. The root detections are depicted with a white rectangle, while parts are in red. The location of the lesions are depicted with yellow arrows.

case of lesions in BUS images, this assumption can not be made. Hence, we have tested models with different configurations. Specifically, we have tested models composed of 1, 2, 3 and 4 components. Figure 3.7 and Table 3.1 show the ROC and FROC analysis for these components. Note that in Figure 3.7 the LLF value at a given NLF point (in this case NLF=0.5) is depicted in the chart legend, while in Table 3.1, the maximum LLF value and its corresponding NLF value are reported. In terms of A_z, the models composed of 1, 3 and 4 components obtained similar results while the model of 2 components obtained the worst results. Of FROC analysis, the models composed of 1, 3 and 4 components obtained similar results in terms of LLF but the 3-components model obtained fewer FPs per image. However, there are no significant differences between the performance of the different number of components, as shown in Table 3.2.

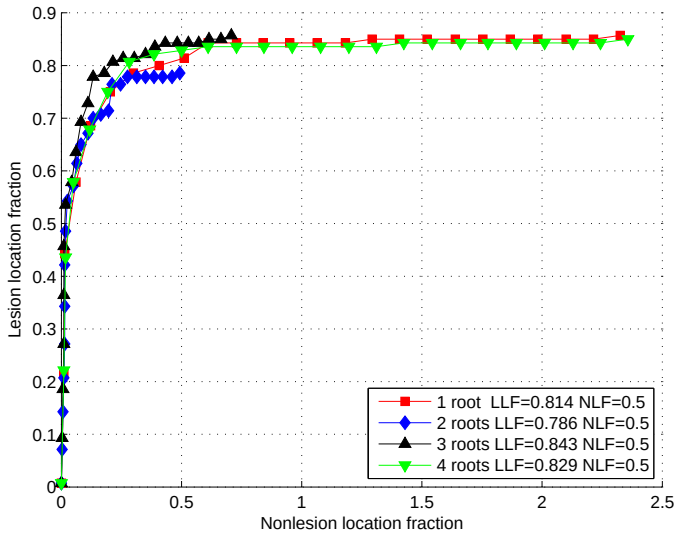
Number of parts

In Felzenszwalb et al. [48], the behavior of the model was tested using different numbers of components with and without part filters, concluding that the use of part filters improved object detection accuracy. However, all the experiments were performed using the same number of part filters. In this study, we followed a similar experiment but changed the number of parts in each model to analyze the effect of this parameter in detection accuracy.

Analyzing both ROC and FROC curves in Figure 3.8 and results in Table 3.1, we can assume that in this case, the ROC analysis does not provide essential information. A significance test, where any p value was higher than 0.05, confirmed this assumption (see

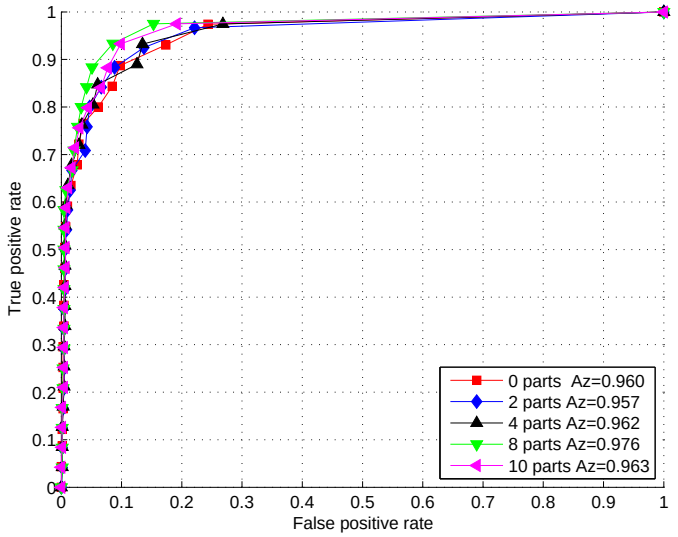


(a)

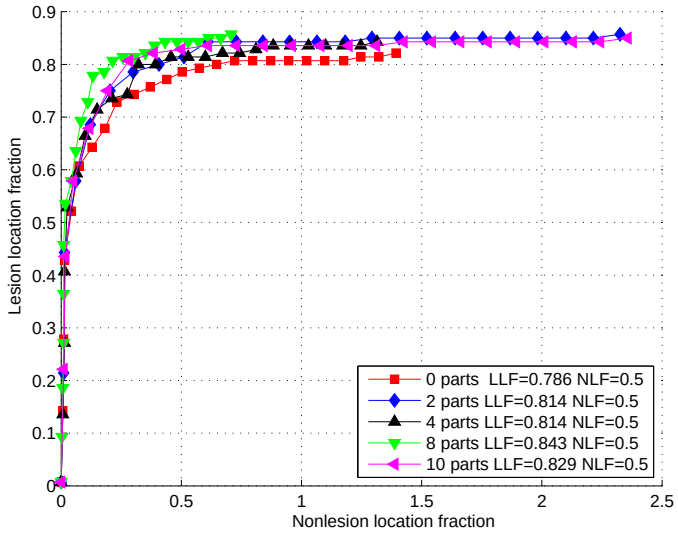


(b)

Figure 3.7: (a) ROC and (b) FROC curves for different numbers of model components.



(a)



(b)

Figure 3.8: (a) ROC and (b) FROC curves for different numbers of part filters.

Table 3.2). However, in the FROC analysis, we can see that models using parts obtained slightly better results. A model without parts obtained a LLF value of 0.82 with 0.60 FPs per image, while the best performance is obtained by the default configuration (with eight parts), giving a LLF value of 0.86 with 0.28 FPs per image.

Size of parts

Regarding part filters, it is also important to define the size of these parts in order to properly detect the finer structures of the lesions. Figure 3.9 depicts the ROC and FROC curves analyzing this parameter, and Table 3.1 shows the quantitative results of these analyses. The 6x6 and 10x10 part filters obtained the best results with A_z values of 0.975 and LLF values of 0.86 and 0.83 respectively. However, 10x10 part filters obtained the best results in terms of NLF, 0.23 compared with 0.28 using the 6x6 part filters. From Table 3.2, we can conclude that the size of the part filters does not have a significant influence in the performance of the DPM in breast US images.

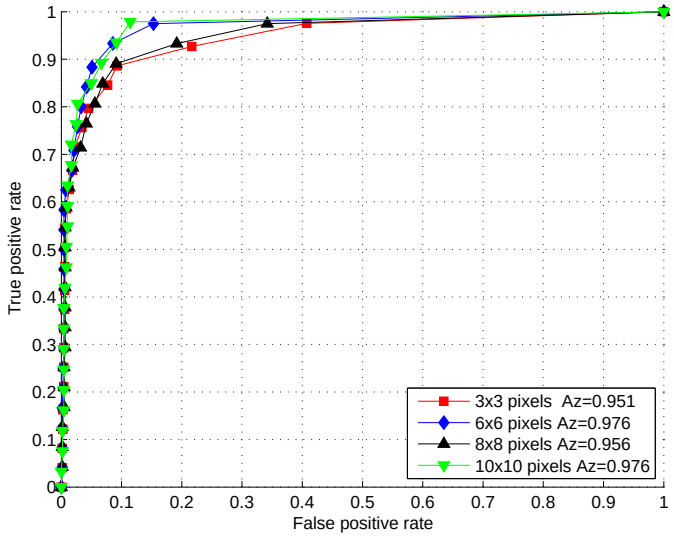
Training threshold

During the training step, a detection process is performed to obtain latent positive results. This parameter defines the minimum area overlap needed to consider a detection as a positive result. If the training threshold is too high, the resulting detector is too restrictive and it does not detect all the lesions in the evaluation process. If the threshold is too low, there are too many latent positive detections and the detector will not be able to distinguish between lesions and normal tissue while evaluating a new image. Our experiments show that all thresholds lower than 0.5 and higher than 0.8 dramatically decrease the performance of the detector. In the specific case of thresholds higher than 0.8, the training process fails because it does not detect any positive regions.

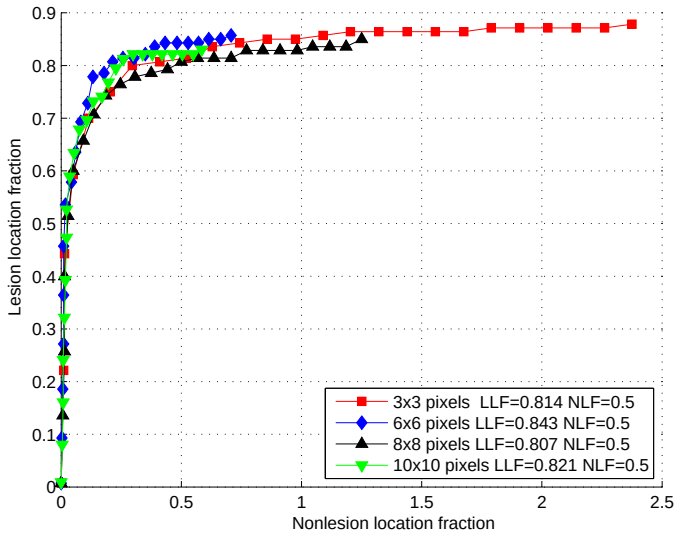
The ROC and FROC analysis for different training thresholds for lesion detection in the 0.5-0.8 range are shown in Figure 3.10 and Table 3.1. Analyzing both curves and Tables 3.1 and 3.2, we can assume that, in this case, the ROC and FROC analysis does not provide conclusive differences. The performance of the system with the different chosen thresholds is almost the same, reaching an A_z value of 0.975. In terms of LLF values, the training threshold selection is not a significant step while the threshold is between the range $0.5 - 0.8$, but threshold 0.7 obtained fewer FPs per image. From Table 3.2, we can conclude that there are no significant differences between the different threshold values.

Cell size

In the DPM approach, the image features are computed by dividing the image into small cells. In each cell, the histogram of gradient orientations is computed. Hence, the size of the cell defines the features involved in the whole detection process. In this experiment, we

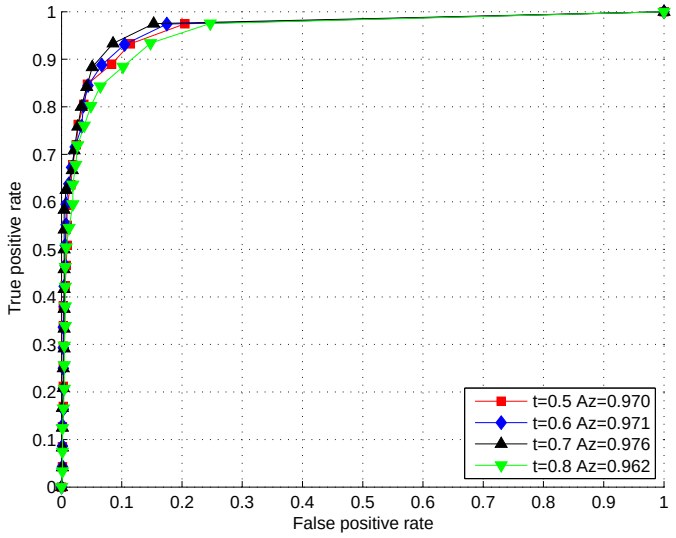


(a)

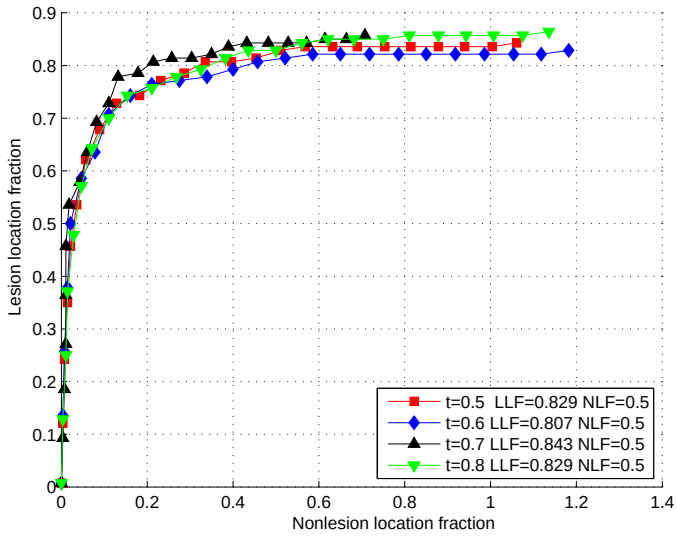


(b)

Figure 3.9: (a) ROC and (b) FROC curves for different sizes of part filters.

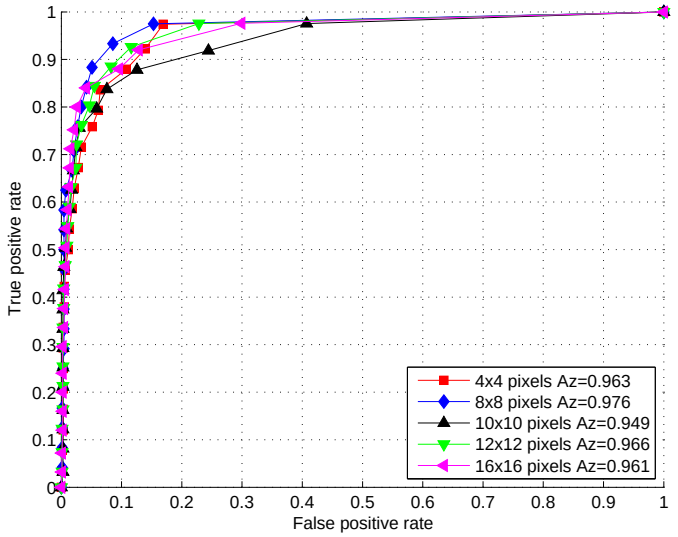


(a)

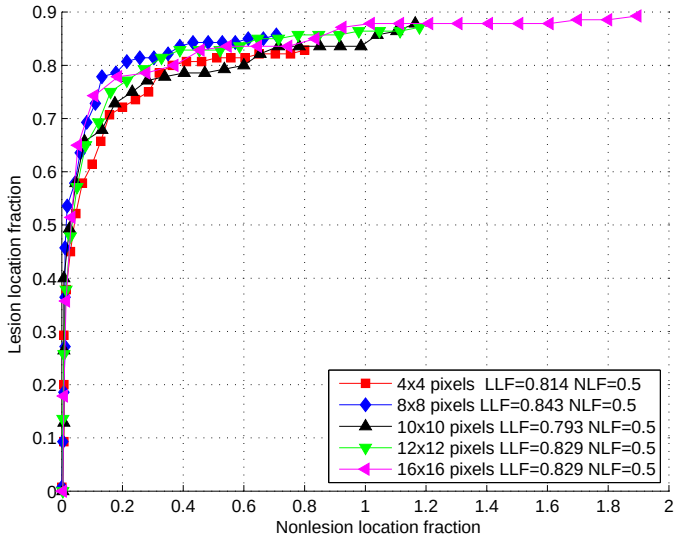


(b)

Figure 3.10: (a) ROC and (b) FROC curves for different training thresholds.



(a)



(b)

Figure 3.11: (a) ROC and (b) FROC curves for different cell size.

have evaluated different cell sizes to analyze the effect of this parameter in the detection process. Specifically, we have evaluated cells of 4x4, 8x8, 10x10, 12x12 and 16x16 pixels. Figure 3.11 and Table 3.1 show the ROC and FROC analysis where an 8x8 cell size obtained the best results in terms of A_z, while cells of 10x10, 12x12 and 16x16 pixels reached a LLF value higher than 0.87. However, Table 3.2 shows that 10x10 and 16x16 cell sizes are significantly worse in terms of NLF. Hence, bigger cells do not capture the features of the image properly, considerably increasing the number of false-positives per image.

HOG pyramid size

The last parameter to analyze is the size of the HOG pyramid. Specifically, parameter λ defines the number of levels necessary to go down in the pyramid to get a feature map computed at twice the resolution. Thus, the higher its value, the more levels of the pyramid in the root and part filters.

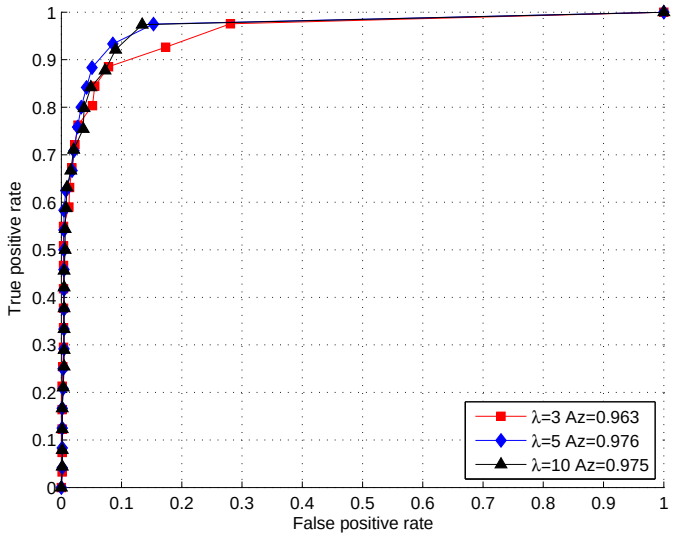
The ROC and FROC analysis in Figure 3.12 and Table 3.1 show that the model created with $\lambda = 10$ obtained the worst results. Hence, if the number of levels in the pyramid between the root and the part filters is too high, the method decreases its sensitivity when detecting lesions due to the low resolution of the root filter. Nevertheless, Table 3.2 states that these differences are not significant.

Summary of the parameter analysis

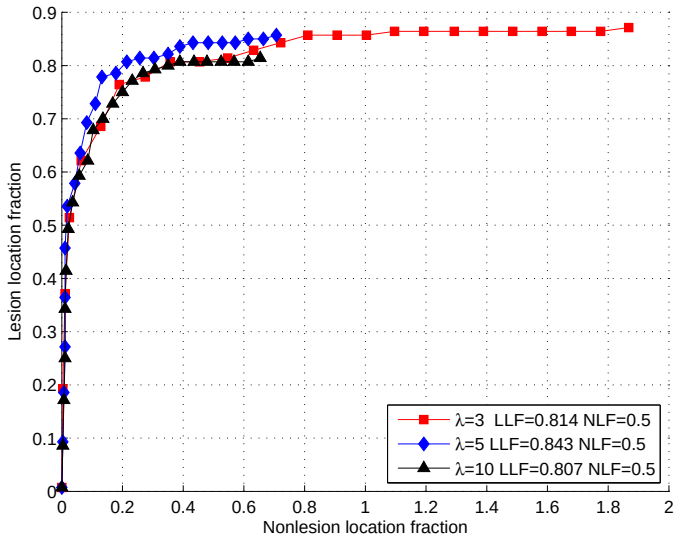
The results obtained in all the previously described experiments concerning the parameter analysis are summarized in Table 3.1. Table 3.2 summarizes the statistical significance results for parameter analysis. From all the experiments, only the cell size parameter has a significant influence on the DPM's performance.

In the selection of the number of components, it is also necessary to consider the computational cost of the method. Table 3.7 summarizes the computation times of the method for training and testing. Note that all training times are the average time for one fold and testing times are for a single image.

The training times vary greatly between 1, 2, 3 or 4 components and the different number of parts. Note that using 8 parts in the 3 component case increases the computation time 12-fold. It is worth recalling that this procedure is performed offline, thus the selection of the component number relies on the computational possibilities and the time available for this computation. In terms of detection, the time also increases with the number of parts and components (from 1.89 to 8.53 seconds for 3 components without parts and 10 parts, respectively). Hence, if the offline computation time is not an issue, we would recommend using a 3-component mixture model with 8 parts of 6×6 pixels, with a training threshold $t = 0.7$, a cell size of 8×8 pixels and a parameter $\lambda = 5$, since results indicate that this configuration obtained the highest sensitivity detecting lesions with fewer FP results per image.



(a)



(b)

Figure 3.12: (a) ROC and (b) FROC curves for different HOG pyramid sizes.

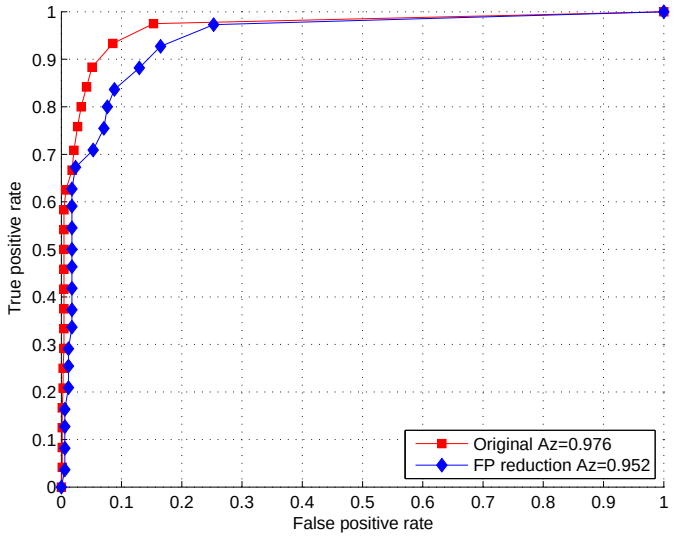
Table 3.1: Summary results for parameter analysis. Default configuration is a 3-component mixture model with 8 parts of 6×6 pixels, with a training threshold $t = 0.7$, a cell size of 8×8 pixels and a parameter $\lambda = 5$.

Parameter		A_z (95% CI)	LLF (95% CI)	NLF (95% CI)
Default		0.975 (0.958-0.989)	0.860 (0.794-0.919)	0.282 (0.107-0.457)
Components	1 root	0.962 (0.944-0.981)	0.878 (0.806-0.951)	0.346 (0.188-0.504)
	2 root	0.971 (0.958-0.983)	0.835 (0.749-0.921)	0.292 (0.179-0.406)
	4 root	0.950 (0.920-0.980)	0.885 (0.822-0.949)	0.675 (0.153-1.196)
Number parts	0 parts	0.957 (0.930-0.983)	0.821 (0.734-0.908)	0.603 (0.226-0.980)
	2 parts	0.955 (0.926-0.983)	0.857 (0.791-0.923)	0.646 (0.095-1.197)
	4 parts	0.953 (0.925-0.981)	0.843 (0.784-0.901)	0.550 (0.177-0.922)
	10 parts	0.957 (0.936-0.979)	0.850 (0.748-0.951)	0.675 (0.149-1.200)
Size parts	3x3 pixels	0.948 (0.916-0.980)	0.879 (0.815-0.941)	0.743 (0.232-1.253)
	8x8 pixels	0.951 (0.921-0.980)	0.850 (0.782-0.917)	0.614 (0.230-0.998)
	10x10 pixels	0.975 (0.963-0.987)	0.830 (0.742-0.917)	0.232 (0.103-0.361)
Threshold	0.5	0.964 (0.946-0.983)	0.843 (0.762-0.923)	0.446 (0.104-0.788)
	0.6	0.967 (0.948-0.986)	0.829 (0.747-0.909)	0.410 (0.075-0.746)
	0.8	0.959 (0.936-0.983)	0.864 (0.796-0.931)	0.457 (0.152-0.761)
Cell size	4x4 pixels	0.959 (0.942-0.977)	0.829 (0.747-0.909)	0.321 (0.162-0.480)
	10x10 pixels	0.945 (0.918-0.971)	0.879 (0.823-0.933)	0.667 (0.365-0.969)
	12x12 pixels	0.964 (0.951-0.978)	0.871 (0.799-0.943)	0.450 (0.319-0.580)
	16x16 pixels	0.956 (0.935-0.976)	0.893 (0.836-0.949)	0.775 (0.443-1.106)
Pyramid size	$\lambda = 3$	0.967 (0.952-0.983)	0.883 (0.796-0.971)	0.383 (0.184-0.583)
	$\lambda = 10$	0.967 (0.949-0.986)	0.821 (0.733-0.909)	0.303 (0.147-0.459)

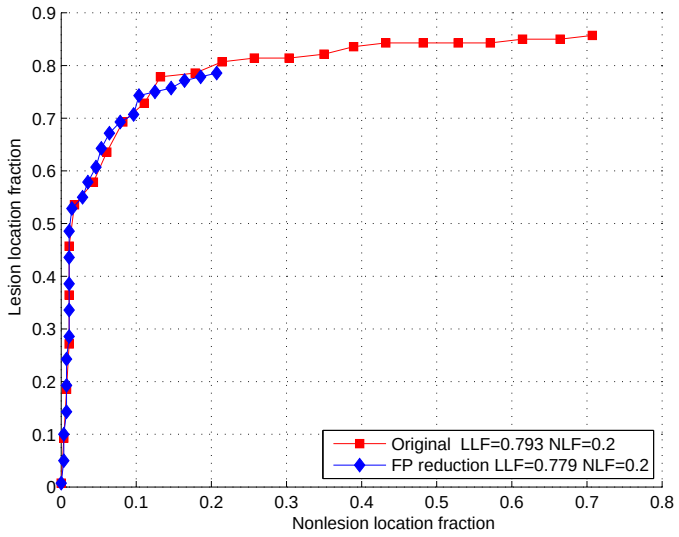
3.3.3 False-positive reduction

In the previous experiments, results were reported selecting a maximum number of ten detections per image. In terms of the FROC analysis, the mean sensitivity reaches 0.86 with 0.28 FP detections per image in the best performance. To reduce this number of FP detections, we performed an experiment where only the best detection per image is selected. Table 3.4 presents the quantitative results from which we can notice the fact that the reduction in terms of sensitivity and false-positives per image is not significant. This indicates that the result with the highest score successfully detects the lesion in most cases.

In Figure 3.13,(a) shows the ROC analysis and (b) the FROC analysis for a model with the default configuration.



(a)



(b)

Figure 3.13: (a) ROC and (b) FROC curves for false-positive reduction comparison.

Table 3.2: Statistical significance for parameter analysis (p value)

Parameter		A_z	LLF	NLF
Components	1 vs. 3 comp.	0.37	0.66	0.59
	2 vs. 3 comp.	0.77	0.69	0.92
	4 vs. 3 comp.	0.18	0.53	0.17
Number parts	0 vs. 8 parts	0.28	0.52	0.14
	2 vs. 8 parts	0.26	1	0.23
	4 vs. 8 parts	0.22	0.74	0.21
	10 vs. 8 parts	0.24	0.90	0.18
Size parts	3x3 vs. 6x6	0.17	0.64	0.11
	8x8 vs. 6x6	0.19	0.88	0.14
	10x10 vs. 6x6	0.92	0.62	0.67
Threshold	0.5 vs. 0.7	0.46	0.78	0.41
	0.6 vs. 0.7	0.59	0.59	0.33
	0.8 vs. 0.7	0.33	0.88	0.34
Cell size	4x4 vs. 8x8	0.24	0.59	0.74
	10x10 vs. 8x8	0.08	0.62	**0.04**
	12x12 vs. 8x8	0.38	0.77	0.14
	16x16 vs. 8x8	0.19	0.41	**0.02**
Pyramid size	$\lambda = 3$ vs. $\lambda = 5$	0.56	0.62	0.46
	$\lambda = 10$ vs. $\lambda = 5$	0.61	0.51	0.86

3.3.4 Bounding box estimation

All the experiments in this paper were assessed by determining if the center of the detected bounding box lies within the delineation of an expert. In the experiments performed with the default configuration, the mean area overlap between the TP detections and the actual lesions is 63%, which indicates that detections considered as TP results properly overlap the actual lesion. However, a complementary experiment was performed to evaluate the estimation of the bounding box using the DPM methodology. Figure 3.14 depicts the percentage of correctly segmented lesions as a function of the threshold overlap value. The overlap value between the detected bounding box and the bounding box of the actual lesion is computed using the AO [74].

Analysis of Figure 3.14 shows that restricting the TP detections to those with an overlap higher to 0.4 the method achieve a sensitivity close to 0.8. This indicates that most of the detections also estimates correctly the bounding box.

Table 3.3: Computation times for model component selection

Components	Training Time (min)	Testing Time (s)
1 root + 8 parts	52.44	2.81
2 root + 8 parts	94.81	3.93
3 root + 8 parts	196.71	6.25
4 root + 8 parts	540.05	7.62
3 root	17.4	1.89
3 root + 2 parts	56.2	3.16
3 root + 4 parts	74.3	4.23
3 root + 10 parts	208.36	8.53

Table 3.4: Comparison using false-positive reduction

	A_z (95% CI)	LLF (95% CI)	NLF (95% CI)
Original	0.975 (0.958-0.989)	0.860 (0.794-0.919)	0.282 (0.107-0.457)
FP reduction	0.956 (0.927-0.986)	0.785 (0.687-0.883)	0.114 (0.034-0.193)
p value	0.32	0.24	0.11

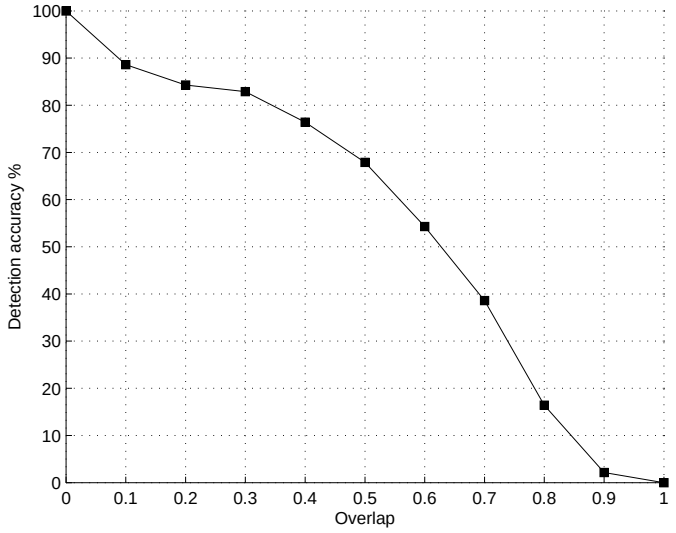


Figure 3.14: Percentage of correctly detected lesions as a function of the threshold overlap value.

3.3.5 Cancer vs benign lesions results

As previously stated, the traditional framework for cancer detection consists of (I) detecting the suspicious lesion, (II) segmenting the lesion, (III) extracting features and (IV) determining if the lesion is malignant with a trained classifier. However, since DPM is a

Table 3.5: Comparison cancer detection

	A_z (95% CI)	LLF (95% CI)	NLF (95% CI)
1 root+parts	0.926 (0.914-0.938)	0.71 (0.647-0.772)	1.54 (1.23-1.84)
2 roots+parts	0.929 (0.907-0.951)	0.78 (0.714-0.845)	1.15 (0.77-1.52)
3 roots+parts	0.953 (0.935-0.971)	0.71 (0.651-0.768)	0.79 (0.27-1.31)
4 roots+parts	0.866 (0.843-0.889)	0.70 (0.618-0.781)	1.23 (1.11-1.34)

multiclass detector, we can build models directly to detect cancerous lesions, simplifying the framework for cancer detection considerably.

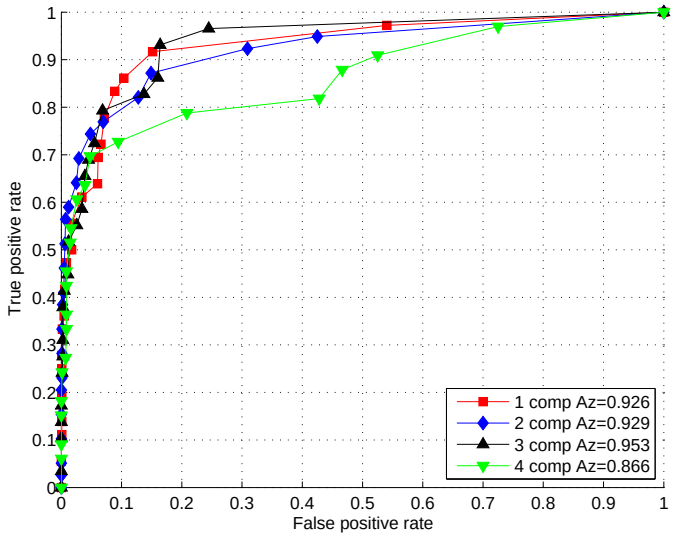
In this subsection, we assess the behavior of DPM for cancer detection. Since DPM is a multiclass detector, we trained the system with 3 different classes: cancerous lesions, benign lesions and normal tissue. Due to the large variability of the shapes of cancerous lesions, the number of components becomes a relevant parameter in terms of cancer detection. Hence, in this experiment, we compare different configurations of components.

Figure 3.15 shows the ROC and FROC results and Table 3.5 summarizes all the results.

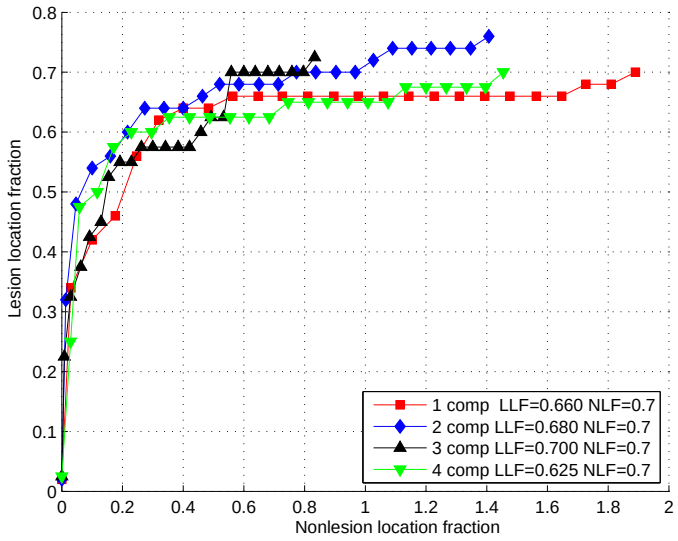
When detecting cancerous lesions, the difference in performance between the different number of components is higher than when detecting lesions. In this case, although A_z values were similar for 1, 2 and 3 components in the ROC analysis, the performance of two roots with parts in the FROC analysis obtained a better sensitivity (or LLF) but the three-component configuration obtained lower FP detections per image (or NLF) than the other options. This can be explained by the fact that cancer detection is a more complex process compared to lesion detection, hence the classifier needs more information to distinguish malignant from benign lesions than when detecting lesions from normal tissue.

3.3.6 Comparison with the most representative methods

In this subsection, the best configuration of the DPM algorithm for lesion detection is compared with some of the more important works in the current literature. One of the main drawbacks when comparing different methods is the heterogeneity in the datasets used. Thus, in this section, the RGI filtering proposed by Drukker et al. [35], the Multi-fractal filtering proposed by Yap et al. [130] and the Rule-based region ranking presented by Shan et al. [109] have been implemented to compare the results obtained by the proposed approach using the same images. Specifically, both dataset D1 and D2 have been used for assessment of these methods. For a better comparison, all the methods in this subsection have been assessed in terms of TPR and FP/image. The methods involved in



(a)



(b)

Figure 3.15: (a) ROC and (b) FROC curves for cancer detection.

this comparison are outlined next.

RGI filtering technique

Drukker et al. [35] proposed a lesion detection and classification method as a two-stage process. The first stage is the detection of lesion candidates using a RGI filtering technique. The second stage is the classification of these candidates, segmenting them by maximizing an average radial gradient ARD index for regions grown from the detected points and classifying them with a Bayesian NN into false positive results and actual lesions. This work focuses on the performance evaluation of the initial lesion detection stage, thus only the location of lesion candidates is evaluated.

Lesion candidates are identified using a filtering technique based on the calculation of the RGI of contours throughout the image [79]. Lesion like shapes for a given point (x, y) in the image are obtained by multiplying the image with a 2D isotropic Gaussian function centered at (x, y) to construct a constrained image. Contour candidates for a given point are obtained by gray-level thresholding on the constrained image. All possible lesion contours within a specified size range are determined, and the RGI value is calculated for each contour as a measure of the likelihood that a given contour represents a lesion.

$$RGI_i(x,y) = \frac{1}{\sum_{(x',y')\in C_i} |\vec{g}(x',y')|} \sum_{(x',y')\in C_i} \vec{g}(x',y') \cdot \hat{r}(x',y') \tag{3.7}$$

where C_i is the i-th possible lesion contour, $\vec{g}(x', y')$ is the maximum gray-value gradient vector of length $|\vec{g}(x', y')|$ and $\hat{r}(x', y')$ the unit radial vector pointing from (x, y) to (x', y').

By definition, due to normalization, the RGI values are between 1 and -1, which indicate that, along the contour, all gradients point radially outwards and inwards respectively. For a given image point (x, y), the contour with the maximum absolute RGI value is selected, and this value is assigned to the (x,y) coordinate in the RGI-filtered image. The RGI-filtered image subsequently undergoes thresholding to determine lesion candidates. The threshold is varied iteratively until either at least one area of interest is detected, indicating a lesion candidate, or the a minimum specified RGI threshold value is reached.

Multifractal filtering technique

The main contribution of the multifractal filtering technique lays in the implementation of multifractal analysis in breast US. In 2008, Yap et al. [130] presented a novel lesion

detection method based on a set of image processing operations, namely histogram equalisation, hybrid filtering, multifractal filtering, thresholding segmentation, and rule-based ROI selection. To ensure the homogeneity of the US images, histogram equalization is first implemented. Then the speckle noise is reduced using a hybrid filtering approach to combine the strength of the nonlinear diffusion filtering [127] to produce edge-sensitive speckle reduction, followed by linear filtering (Gaussian blur) to smooth the edges and to eliminate over-segmentation. Subsequent to hybrid filtering, multifractals [53] are used to further enhance the partially processed images. Multifractal analysis refers to the analysis of an image using multiple fractals (i.e. not just one as in fractal analysis). The generalized formulation for Multifractal Dimensions (D) of order q can be represented as:

$$D_q = \begin{cases} \frac{1}{q-1} \lim\limits_{\epsilon \to 0} \frac{\log(x_q(\epsilon))}{\log(\epsilon)} & \text{for } q \in \mathbb{R} \text{ and } q \neq 1 \\ \lim\limits_{\epsilon \to 0} \frac{\sum\limits_i \mu_i \log \mu_i}{\log(\epsilon)} & \text{for } q = 1 \end{cases} \tag{3.8}$$

where ϵ is the linear size of the cells, q is the order for cell size ϵ and μ is the measure defined as the probability of the grayscale level in the images, where all the gray levels fall in the range of $[0, 1]$. From Yap et al.'s experiment, multifractal analysis enables improved separability of tumour regions from normal regions. Further discussion on multifractal analysis can be found in Yap et al. [130].

After pre-processing, images were segmented by using a gray-value thresholding segmentation method [101]. This thresholding segmentation often leads to the identification of multiple regions of interest, of which, generally only one or two would be of diagnostic importance. To identify these important regions, a rule-based ROI selection based on the size of the region and location is used as a discriminative criterion. Based on the knowledge provided by expert radiologists [103], it is assumed that most of the lesions are located in the upper part of the images. Hence, a reference point (x, y) where

$$x = \frac{\text{image height}}{3}, \text{ and } y = \frac{\text{image width}}{2} \tag{3.9}$$

is chosen. Thus, the suspicious lesion is selected as the largest segment is closest to the (x, y) location.

Rule-based region ranking

Shan et al. [109] proposed a lesion detection methodology that considered both texture and spatial features. They first used the Speckle Reducing Anisotropic Diffusion (SRAD) [133]

as a de-speckling method. The SRAD method iteratively processes the image with adaptive weighted filters to reduce noise and preserve edges. The diffusion coefficient is determined by

$$c(q) = \frac{1}{1 + [q^2(x, y; t) - q_0^2(t)]/[q_0^2(t)(1 + q_0^2(t))]} \tag{3.10}$$

where $q(x, y; t)$ is the instantaneous coefficient of variation depending on the gradient ∇I and the Laplacian $\nabla^2 I$ and determined by

$$q(x, y; t) = \sqrt{\frac{(1/2)(|\nabla I|/I)^2 - (1/4)(\nabla^2 I/I)^2}{[1 + (1/4)(\nabla^2 I/I)]^2}} \tag{3.11}$$

The initialized $q_0(t)$ is given by

$$q_0(t) = \frac{\sqrt{var[z(t)]}}{z(t)} \tag{3.12}$$

where t is the iteration time and $z(t)$ is the most homogeneous area in the image at iteration t and $var[z(t)]$ is its variance.

Once the image is de-speckled, an iterative threshold selection algorithm is applied to segment the image. First, all local minima of the image histogram are calculated and the de-speckled image is binarized using the smallest local minimum as a threshold. Then, if the ratio between the number of foreground pixels and the number of background pixels is less than 0.1, the next local minimum value is set as a threshold. The process continues iteratively until the ratio is larger than 0.1, which is chosen experimentally in the original paper. After that, morphological operators (dilate and erode) are performed to remove noisy regions. Subsequently, if none of the regions intersect with an image center region (a window about half the size of the entire image located on the image center) the threshold becomes the next local minimum and the process is repeated. Once a region intersects with the central window, regions connected with the boundary that do not intersect with the central window are removed. Then, the remaining lesion region candidates are ranked following the scoring formula

$$S_n = \frac{\sqrt{Area_n}}{dis(C_n, C_0) \cdot var(C_n)}, n = 1, ..., k \tag{3.13}$$

where k is the number of candidate regions, $Area_n$ is the number of pixels in the region, C_n is the center of the region, C_0 is the center of the image, $dis(a, b)$ is the Euclidean

distance between points a and b and $var(C_n)$ is the variance of a small region (5×5 pixels) centered on C_n.

Finally, the seed point is located in the center of the winning region. Thus, $((x_{min} + x_{max})/2, (y_{min} + y_{max})/2)$ is considered as a seed point, where $[x_{min}, y_{min}, x_{max}, y_{max}]$ defines the minimum rectangle that contains the lesion.

Comparison results

Since we used the best configuration for DPM, we have also tuned several parameters involved in the other methods to obtain the best results in our datasets. In Shan et al. [109], since most of the lesions appear in the top region of the image, the central window was initialized in the center-top part of the image. In addition, the iteration time t was set to 50 in the SRAD process. For the Yap et al. [130] approach, in the Multifractal analysis the fractal order was set to $q = 1$ for the cell size $\epsilon = 3$.

Example results from both datasets D1 and D2 are presented in Figures 3.16, 3.17 and 3.18. Figure 3.16 shows two examples from different datasets where all the detection methods detected the lesion correctly. Both examples present lesions with well defined boundaries and with different appearances from the normal tissue (intensity values and texture).

Figure 3.17 depicts difficult examples in which none of the methodologies compared were able to detect the lesion. The example from Dataset D1 presents a complex shadow that induces all the methods to a wrong result. In the example from Dataset D2, there is a large area of the image corresponding to lung air with a similar intensity appearance to the lesion. In addition, the lesion's size is small which also made the detection difficult.

Finally, Figure 3.18 shows difficult examples in which only the DPM approach detected the lesion correctly. The example from Dataset D1 presents a complex shadow with similar intensity to the lesion region, which made all the methods except DPM fail in detecting the actual lesion. The same methods fail in the example from Dataset D2 that presents a lesion with a similar appearance to normal tissue. Furthermore, the lesion is located near the top border of the image and therefore missed by all the methodologies that assumed a given lesion location.

Quantitative results are presented in Table 3.6. They are provided in terms of TPR and FPs per image. Overall, the DPM approach out-performed the other methods for the lesion detection in both datasets, with 0.8 TPR and 0.28 FP/image in Dataset D1, and

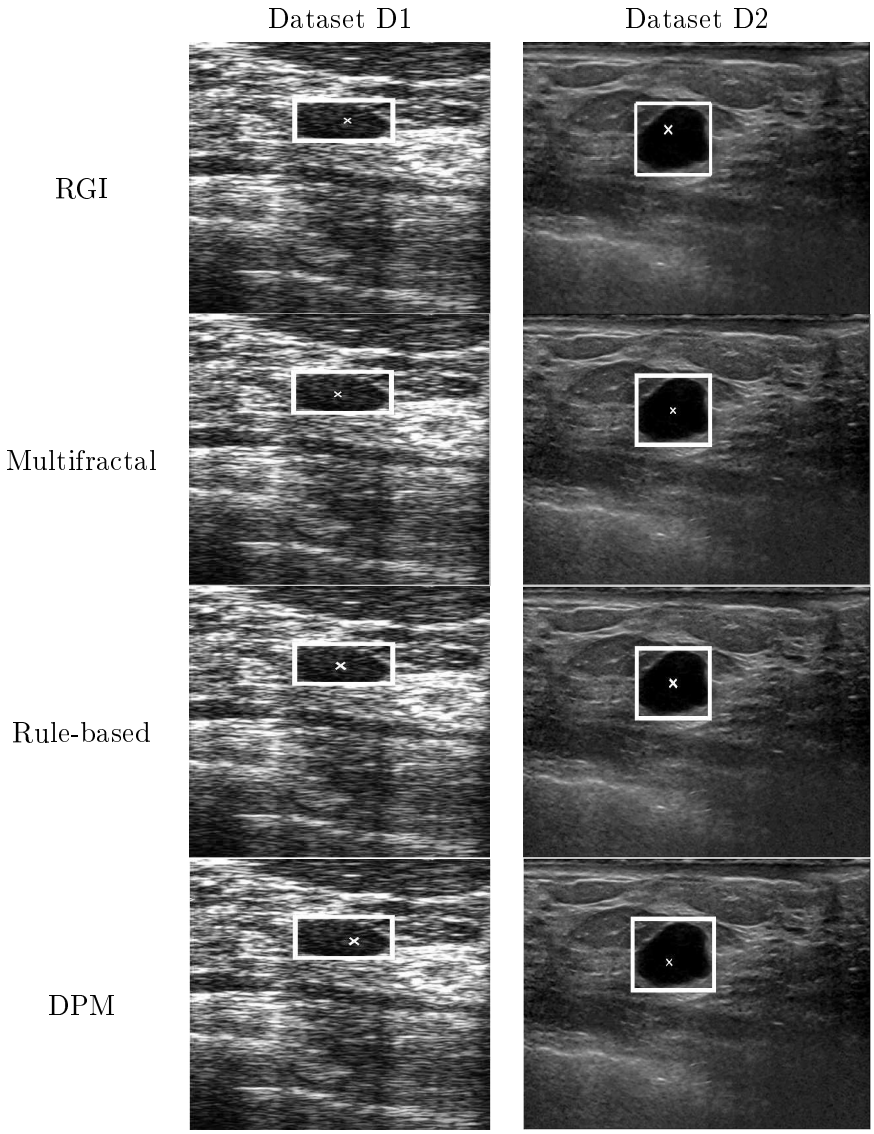


Figure 3.16: Examples from datasets D1 and D2 in which all the methodologies correctly detect the lesion. The ground truth is marked with a square, while the detection is marked with an X. In the first column are the results with an image from Dataset D1, and in the second column from Dataset D2. The first row shows the results of the RGI filtering technique [35], the second row shows the results of the Multifractal filtering technique [130], the third row shows the result of the Rule-based region ranking approach [109] and the last row shows the results of the proposed method.

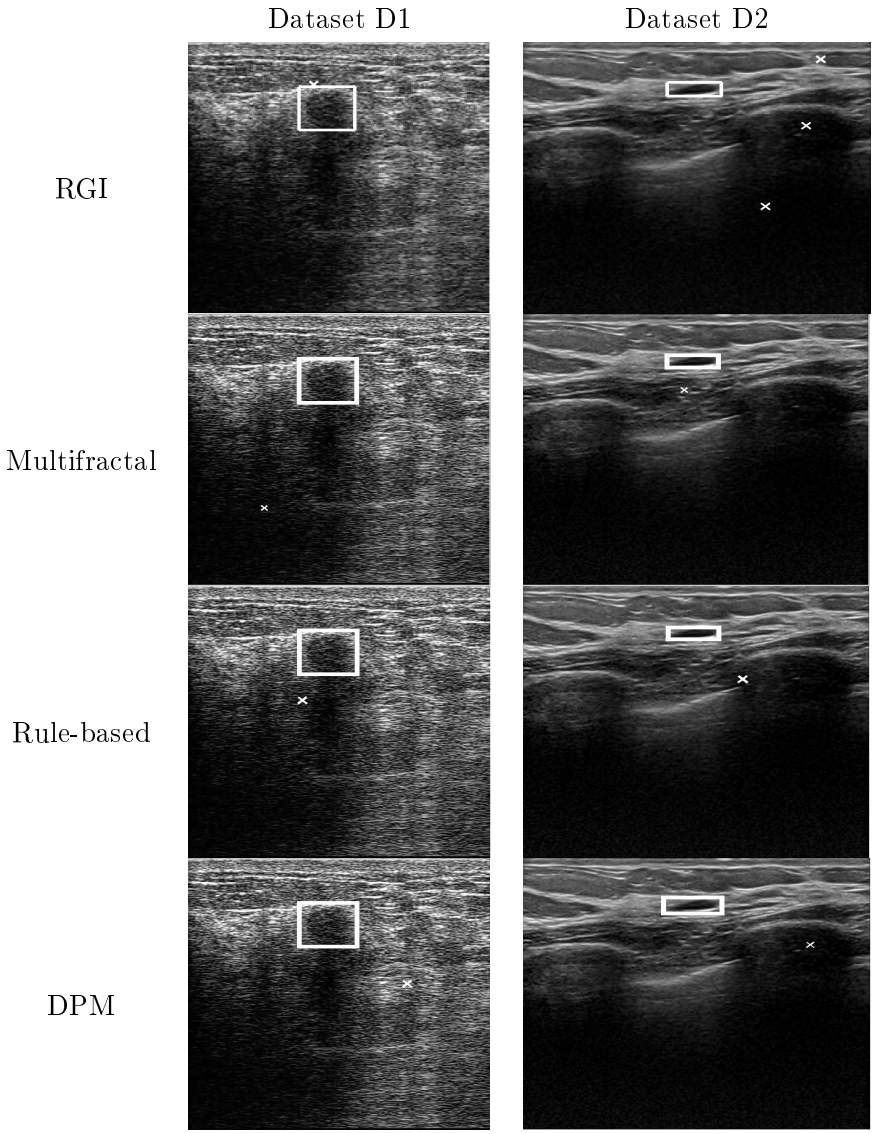


Figure 3.17: Examples from datasets D1 and D2 where all the methodologies miss the detection of the lesion. The ground truth is marked with a square, while the detection is marked with an X. In the first column are the results with an image from Dataset D1, and in the second column from Dataset D2. The first row shows the results of the RGI filtering technique [35], the second row shows the results of the Multifractal filtering technique [130], the third row shows the result of the Rule-based region ranking approach [109] and the last row shows the results of the proposed method.

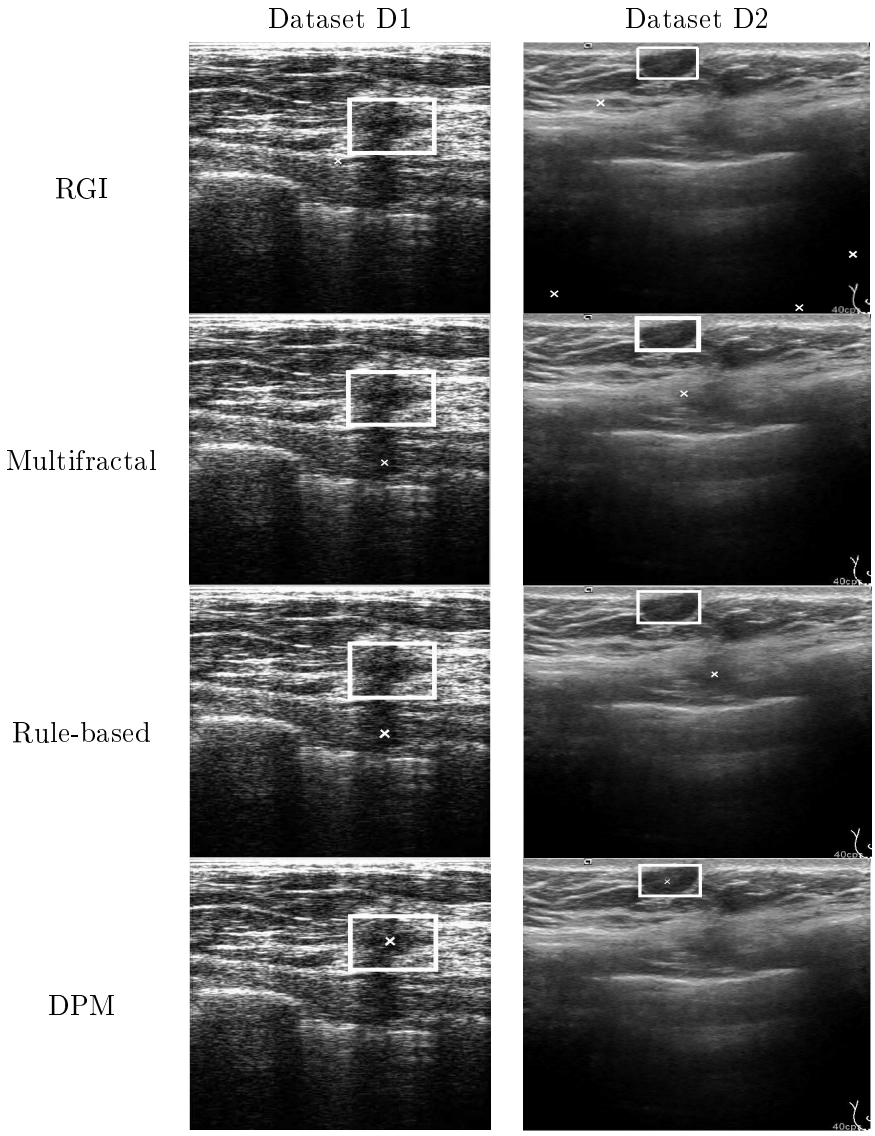


Figure 3.18: Examples from datasets D1 and D2 in which only the DPM approach correctly detects the lesion. The ground truth is marked with a square, while the detection is marked with an X. In the first column are the results with an image from Dataset D1, and in the second column from Dataset D2. The first row shows the results of the RGI filtering technique [35], the second row shows the results of the Multifractal filtering technique [130], the third row shows the result of the Rule-based region ranking approach [109] and the last row shows the results of the proposed method.

Table 3.6: Comparison of performance for different proposals.

Method	Dataset	TPR	FPs per image
Drukker et al. (2002) [35]	D1	0.76	1.57
	D2	0.72	2.47
Yap et al. (2008) [130]	D1	0.76	0.31
	D2	0.59	0.51
Shan et al. (2008) [109]	D1	0.75	0.50
	D2	0.60	0.54
Proposed method	D1	0.80	0.28
	D2	0.79	0.21

0.79 TPR and 0.21 FP/image in Dataset D2. The multifractal filtering technique [130] and rule-based region ranking [109] obtained good results for the images in Dataset D1, 0.76 and 0.75 TPR with 0.31 and 0.50 FP/image respectively, but not for the images in Dataset D2, 0.59 and 0.60 TPR with 0.51 and 0.54 FP/image respectively. Finally, the RGI filtering technique [35] showed a good performance in terms of TPR in both datasets (0.76 and 0.72 TPR) but with a large number of false-positive detections per image (1.57 and 2.47 respectively).

Comparing the performance of the methodologies according to which dataset is used, only the DPM provided similar results for both datasets. The rest of the methods were inconsistent and provided poor results when tested on Dataset D2 compared to Dataset D1. For that reason we made a further evaluation to see if there are significant differences between both datasets. Specifically, we compared the lesion size, the ratio between the area of the lesion and the area of the image, and the distance from the image center and the lesion centroid. Figure 3.19 shows the box plot charts for these comparisons where differences between both datasets are appreciable: the average size of the lesions in Dataset D1 is smaller than in Dataset D2 but the ratio between lesion pixels and total image pixels is higher. Regarding the spatial distribution of the lesions in the image, lesions in Dataset D1 are more centered than in Dataset D2. However, none of these differences are significant, hence, other characteristics such as the quality of the image may affect the behavior of the methodologies. The images in Dataset D2 present more defined structures such as ribs, muscles or parenchymal tissue that may lead some methodologies to err when detecting lesions.

Finally, the average computational time of each methods is compared, as shown in Table 3.7. Note that in most of the cases, average detection time for dataset D2 is higher

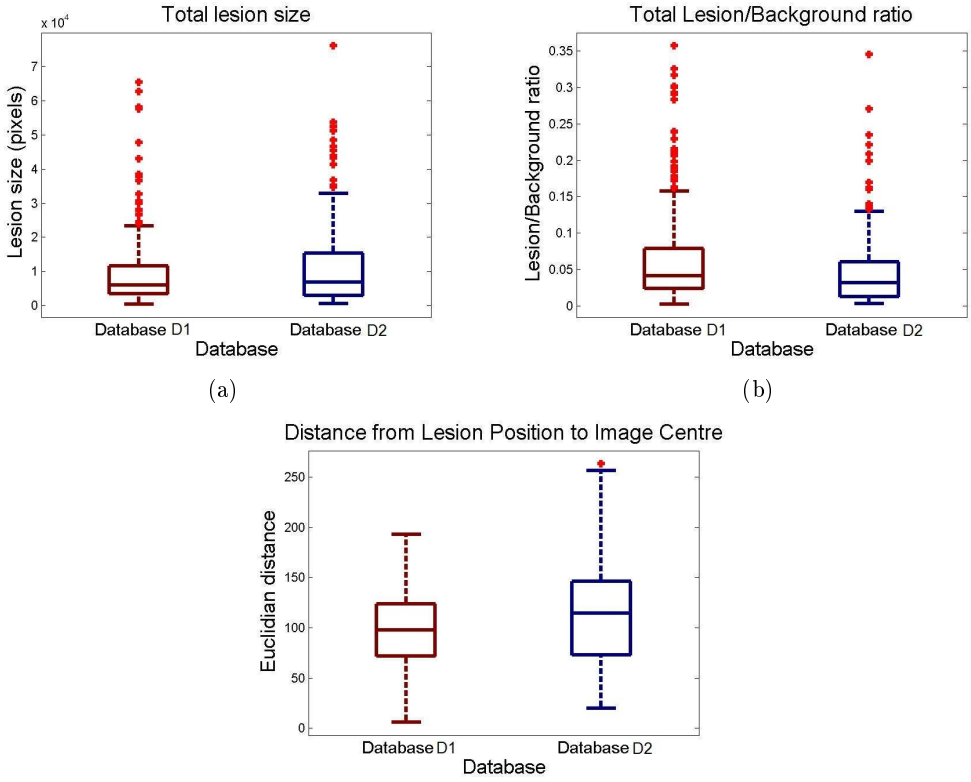


Figure 3.19: Dataset features comparison. Box plot chart comparing (a) the lesion size, (b) the ratio between the area of the lesion and the area of the image, and (c) the distance from the image center and the lesion centroid.

than D1 due to the resolution of these images. The Drukker et al. [35] approach is the slowest approach in detecting lesions, taking an approximate time of 45 seconds per image. This is caused by the fact that Drukker computes the RGI value for each point in the image, which is a time-consuming task. Shan et al. [109] is notably faster than the rest of the methodologies. The training time is also included for the DPM technique for both datasets. Each fold in the training process in the DPM takes 211.67 minutes for Dataset D1 and 196.71 minutes for Dataset D2.

3.3.7 Comparison of the reported results in other literature methods

In this subsection we compare the obtained results with the results reported in other important works in the literature but have not been implemented. Tables 3.8 and 3.9 summarize the results for lesion and cancer detection respectively.

Table 3.7: Comparison of computation time for different proposals.

Method	Dataset	Avg. detection time/image (s)	Avg. training time/fold (min)
Drukker et al. (2002) [35]	D1	45.56	-
	D2	44.42	-
Yap et al. (2008) [130]	D1	3.29	-
	D2	5.91	-
Shan et al. (2008) [109]	D1	0.42	-
	D2	0.68	-
Proposed method	D1	4.14	211.67
	D2	6.25	196.71

Table 3.8: Comparison of performance for other lesion detection methods

Method	A_z	LLF	NLF
Kutay et al. (2003) [80]	0.970	-	-
Drukker et al. (2005) [37]	0.950	0.80	0.60
Mogatadakala et al. (2006) [94]	0.910	-	-
Proposed method	0.975	0.86	0.28

Table 3.9: Comparison of performance for other cancer detection methods.

Method	A_z	LLF	NLF
Horsch et al. (2002) [65]	0.870	-	-
Drukker et al. (2003) [36]	-	0.660	0.25
Kutay et al. (2003) [80]	0.810	-	-
Drukker et al. (2005) [37]	0.860	0.80	0.83
Alvarenga et al. (2012) [6]	0.870	-	-
Proposed method	0.929	0.78	1.15

Analyzing the results of Table 3.8, in terms of lesion detection, our approach has obtained the best performance in the ROC analysis ($A_z = 0.975$), while in the FROC analysis it has obtained a sensitivity of 0.86 with 0.28 FP detections per image. These results slightly outperformed the results reported by Drukker et al. [35], who obtained a sensitivity of 0.87 with 0.76 FP detections per image. Hence, our proposal obtained a similar sensitivity with less FP detections per image.

In the task of cancer detection (see Table 3.9), our approach has also obtained the best performance in terms of the ROC analysis ($A_z = 0.929$), while in the FROC analysis it has obtained a sensitivity (0.78) close to Drukker et al. [37] (0.8) but with a higher number of

FP detections per image (1.15 compared to 0.83).

As mentioned before, the lack of standard datasets of US images makes the comparison between the ROC and the FROC results reported by the methods in the literature very difficult due to the high variability in image quality between US systems, the heterogeneity of lesion type, the size of the lesions and number of images. However, our proposal obtained the best results for lesion detection, and seems to be in line with the best approach for detecting cancerous lesions.

3.3.8 Discussion

After the extensive analysis of the different parameters involved in the DPM methodology, we have found that most of the parameters do not significantly influence the results of the DPM detecting lesions in breast US images. Although not significant, results indicate that the configuration that obtained the highest sensitivity detecting lesions with fewer FP results per image is a 3-component mixture model with 8 parts of 6×6 pixels, with a training threshold $t = 0.7$, a cell size of 8×8 pixels and a parameter $\lambda = 5$.

In terms of the training threshold value, we have found that thresholds lower than 0.5 and higher than 0.8 decrease the performance of the detector significantly. The selection of a threshold within this range is not critical for the method's performance according to the experimental results. We have also noticed a large difference of computation time between different numbers of components and parts. In this case, the selection of the number of components clearly depends on the specific problem and the available computational resources. In cases where computation time is not an issue, a three component model with 8 parts, which obtained slightly better results, is the recommended configuration.

An experiment for reducing the FP results was analyzed. The reduction in terms of sensitivity was not significant, which makes the results very attractive in case we want to adapt this methodology to 3D images where the number of FP detections needs to be strictly low without decreasing the sensitivity. Later, a bounding box estimation test was performed to evaluate the estimation of the detected bounding box as a function of different overlap thresholds. Comparing this with other methods in the literature, only [35] performed a similar experiment obtaining an accuracy close to 70% for an overlap value of 0.5. These results are similar to those presented in Figure 3.14.

Regarding the comparison made with the implementation of different methods with a common dataset, the methods based on image processing (RGI filtering [35], multifrac-

tal filtering [130] and rule-based region ranking [109]) obtained poor results when dealing with images acquired from two different US systems. One explanation is that most of the approaches take into consideration the characteristics of their datasets such as the lesion location, the influence of the speckle noise or the appearance of the lesions. These characteristics may not be the same in other datasets, reducing the accuracy of the algorithms. We can also detect differences in performance regarding the dataset used. Dataset D2 was acquired using a modern US system, which introduces new challenges for the existing techniques in lesion detection. These US systems acquire high-resolution images which may include other structures such as ribs, pectoral muscle or air in the lungs making the lesion detection more difficult. Dataset D1 was obtained from an older US system. The nature of the images is normally of a lower resolution and noisier. Hence, for a better visualization, the radiologist tends to locate the suspected lesion in the center of the image. Nowadays, with high quality US systems this is no longer necessary due to the fact that one image can capture larger regions of the breast. Hence, methodologies that made any kind of assumption about the lesion's location fail in most cases using modern US systems.

The proposed technique obtained the best results for both datasets. This is due to the fact that this approach adopts a training process that helps the method to build a particular characteristics model of each dataset. Thus, it is not as dataset dependent as the other methodologies.

For an exhaustive assessment, we have also compared the proposed methodology results with those published in some of the most important works for lesion detection. Our proposal slightly outperformed the other works in terms of both the ROC and FROC analysis, although different datasets were used.

Related to cancer detection, the performance of our approach in terms of the ROC and FROC analysis obtained worse results compared to lesion detection. This could be explained by the fact that cancerous lesions have a large variability in terms of shape, texture and intensity, making the process more complex. However, when comparing our results with others reported in the literature, our proposal obtained the best performance in terms of the ROC analysis. In terms of FROC analysis we obtained results similar to the best work [37], but considerably reducing the typical framework (detection, segmentation, feature extraction and classification).

3.4 Conclusions

In this chapter we have presented a computerized lesion detection system for breast US images using Deformable Part Models that has been evaluated in a large dataset. Different configurations of parameters have been tested to improve the results of the DPM in breast images. We have also performed a comparison with several of the most important methods in the literature, using the same datasets. Amongst the different methodologies compared, the proposed method obtained the best results for both datasets. This method is adaptable to the specific characteristics of any dataset, since it is machine-learning-based where a particular model is constructed for each dataset. In addition, we have compared the results obtained with the results published by other relevant methodologies in the literature, obtaining the best results in the ROC analysis for lesion detection. In the FROC analysis, our method outperformed the best state-of-the-art approach in detecting lesions but we reported a slightly higher number of false-positive detections per image when detecting malignant lesions. We also proved that by building a model to directly detect malignant lesions, we can considerably reduce the traditional CAD pipeline (detection, segmentation, feature extraction, lesion classification) obtaining results similar to those reported in the state-of-the-art.

In conclusion, we have shown that DPM can be used for lesion detection in breast US images and has the potential of being implemented in a clinical environment. This lesion detection approach can also be used to solve the initialization step of some lesion segmentation algorithms, which will be dealt with in the next chapter. Furthermore, the methodology does not make any assumption on the dimensionality of the data, hence it is our assertion that it can be adapted to a 3D volume such as Automated Breast Ultrasound (ABUS), which is currently being adopted in clinical practice and becoming a relevant topic of interest in medical imaging research.

4

Breast lesion segmentation

4.1 Introduction

Image segmentation plays an important role in the analysis of US images and is a relevant step in a CAD system. The aim of the lesion segmentation algorithms in medical imaging is to accurately describe the boundary of the lesion regions. This allows the radiologist to compare the same abnormality over time and extract conclusions about its behavior (if the lesion size grows or decreases) or extract discriminative features for a further classification stage. In addition, an accurate segmentation can provide the radiologists with a realistic measure of the lesion size before they start a biopsy procedure.

Lesion segmentation in breast sonography is an important topic of interest, as shown in Chapter 2. Reviewing the state-of-the-art, one's attention is drawn to the relatively low number of approaches based on MRF, which in principle also seems an interesting framework incorporating both spatial and intensity information. Among these, the work of Xiao et al. [129] seems to have a good potential for segmentation, but their work was based on a limited number of cases and also needed an important amount of manual interaction. The aim of this chapter is to develop and evaluate a lesion segmentation algorithm based

on the earlier work of Xiao et al. More specifically, we present different approaches to improve the Xiao et al. approach by means of reducing the human interaction involved in the method. Two datasets of US B-mode breast images obtained from two different hospitals using different US equipment are used to evaluate these approaches and compare them to the original method. Furthermore, the results are evaluated depending on the lesion type. This study allows obtaining conclusive results regarding the accuracy of an MRF based method, and the influence of the lesion type in the dataset of images used for the segmentation process. This information could be useful in terms of further CAD research for focusing the efforts in the pathologies that present the worst results. Finally, a comparative study with different approaches in the literature is performed.

4.2 MRF-MAP segmentation overview

The MRF-MAP segmentation method was originally proposed by Zhang et al. [136] to segment brain tissues on MRI by estimating a bias (or distortion) field while labeling tissues at the same time, and later was adapted to BUS images by Xiao et al. [129].

They considered the bias field as an additive artifact in the logarithmic ideal image. This distortion field is estimated to restore the ideal image while regions of similar intensity inhomogeneity are identified using a MRF and MAP framework; thus, this method will be referred to hereafter as the MRF-MAP approach (see Figure 4.1).

4.2.1 Image model

Attenuation-related intensity inhomogeneities are assumed to be a multiplicative field with low-frequency. A logarithmic transformation yields such a multiplicative model to an addition $y = y^* + d$, where d denotes the log-transformed intensity distortion field. It is assumed that y_i^*, ideal values at pixel i, follow a mixture of Gaussian distributions. Hence, assuming that the pixel intensities are statistically independent, the probability density for the entire image given the bias field is defined as

$$p(y|d) = \prod_{i \in \mathcal{S}} \sum_{j \in \mathcal{L}} g(y_i - d_i; \theta(x_i)) p(x_i = j) \tag{4.1}$$

where $\mathcal{L}$ denotes the label set; x_i is the corresponding class label of pixel i; and $\mathcal{S}$ denotes the image pixel set. The prior probability is $p(x_i = j)$ and the Gaussian distribution is

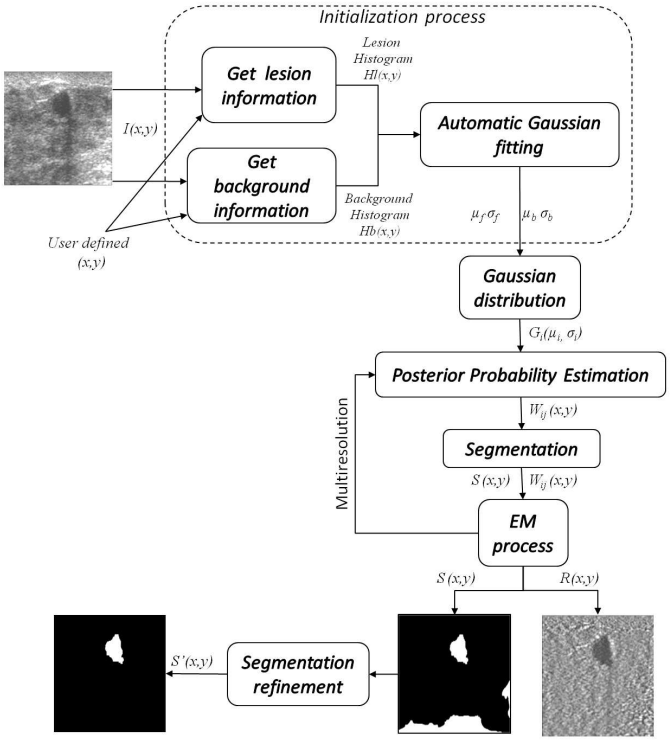


Figure 4.1: Block diagram of MRF-MAP method.

defined as $g(y_i^*; \theta(x_i))$ with parameters $\theta(x_i) = (\mu_{x_i}, \sigma_{x_i})$, μ_{x_i} and σ_{x_i} being the mean and standard deviation of class x_i, respectively.

Given the observed intensity values, the Baye's rule can be used to obtain the posterior probability of the bias field:

$$p(d|y) = \frac{p(y|d)p(d)}{p(y)} \tag{4.2}$$

where $p(y)$ is a normalization constant, and the prior probability of the bias field $p(d)$ is modeled as a Gaussian with zero mean to capture its smoothness property. The MAP principle can be used to obtain the optimal estimation of the bias field $\hat{d}$, given the observed intensity values:

$$\hat{d} = \arg\max_d p(d|y) \tag{4.3}$$

A zero-gradient condition is then used to assess this maximum, which leads to (see [128] for details).

$$W_{ij} = \frac{p(y_i|x_i, d_i)p(x_i = j)}{p(y_i|d_i)} \tag{4.4}$$

$$d_i = \frac{[FR]_i}{[F\psi^{-1}E]_i}\text{, with } E = (1, 1, ...)^T \tag{4.5}$$

where W_{ij} is the posterior probability that pixel i belongs to class j given the estimated bias field d; F is a low-pass filter, and R is the mean residual for a pixel i, and is defined as

$$R_i = \sum_{j \in \mathcal{L}} \frac{W_{ij}(y_i - \mu_j)}{\sigma_j^2} \tag{4.6}$$

and ψ is the mean inverse covariance, in which

$$\psi_{ik}^{-1} = \begin{cases} \sum_{j \in \mathcal{L}} W_{ij}\sigma_j^{-2} & \text{if } i = k \\ 0 & \text{otherwise} \end{cases} \tag{4.7}$$

If the prior probability $p(x_i = j)$ in equations 4.1 and 4.4 is set to be equal for all $j \in \mathcal{L}$ at every pixel i, W_{ij} is the normalized conditional probability and its estimation is given by a maximum-likelihood approach. Hence, an EM algorithm can be applied as follows:

Expectation step: calculates the posterior tissue class probability W_{ij} when the bias field d is known using equation 4.4.

Maximization step: the bias field d_i is estimated when W_{ij} is known using equation 4.5.

Initially, the bias field is initialized to zero everywhere in the image. The algorithm runs iteratively until estimating the bias field. Once the bias field is obtained, the ideal intensity I^* can be restored by dividing the observed image I by $exp(d)$. A class labeling x of the pixels is obtained by assigning to each pixel $i \in \mathcal{S}$ the label $j \in \mathcal{L}$ that has the largest value W_{ij}.

4.2.2 Label estimation using MRF-MAP framework

The maximum-likelihood approach is known to be sensitive to noise and thus one can argue that it would not be suitable for US images. Hence, Zhang et al. [136] proved that a full MAP estimation can be achieved incorporating a MRF prior model for the image tissue classes. MRF formulation provides a convenient way to keep spatial information coherence.

Let x indicate the true but unknown labeling of the given image, and $\hat{x}$ represents an estimate of x. Both variables can be interpreted as realization of a random field X, and the log-transformed image y can be interpreted as the realization of a random field Y. Then, the problem becomes one of estimating x, given y, which can be obtained using MAP estimation as follows:

$$\hat{x} = \arg\max_{x \in \mathcal{X}} p(x|y) \tag{4.8}$$

According to Baye's rule

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} \tag{4.9}$$

where $p(y|x)$ has the form of equation 4.1, $p(y)$ is a normalization constant, and $p(x)$ is discussed next.

In MRF, it is assumed that neighboring sites have direct relationships with each other, and tend to have the same class labels. The probability density of an MRF model can be described as a Gibbs distribution as follows

$$p(x) = Z^{-1}\, exp[-U(x)] \tag{4.10}$$

where Z is a normalization term and $U(x)$ an energy function defined as

$$U(x) = \beta \sum_{c \in \mathcal{C}} V_c(x) \tag{4.11}$$

where, $V_c(x)$ is a clique potential over all possible cliques $\mathcal{C}$, which is defined as a subset of pixels in $\mathcal{S}$ where every pair of pixels are neighbors and β is a positive constant that controls the size of that clique. The clique potential of x_i with respect to its clique neighbor $x_{i'}$ is of the form

$$V_{x_i, x_i'}(x_i) = 1 - \delta_{(x_i = x_{i'})} \tag{4.12}$$

$$\delta_{(x_i = x_{i'})} = \begin{cases} 1 & \text{if } x_i = x_i' \\ 0 & \text{if } otherwise \end{cases} \tag{4.13}$$

Once the energy function $U(x)$ is built, the MAP of class labels is obtained as

$$\hat{x} = \arg\min_{x \in \mathcal{X}} U(y|x) + U(x) \tag{4.14}$$

where $U(y|x)$ likelihood energy is the logarithm of the posterior probability $p(x|y)$, and it is defined as follows

$$U(y|x) = \sum_{i \in \mathcal{S}} \left[\frac{(y_i - \mu_{x_i})^2}{2\sigma_{x_i}^2} + log(\sigma_{x_i}) \right] \tag{4.15}$$

and the prior energy $U(x)$ can be defined as

$$U(x) = \sum_{i \in \mathcal{S}} \sum_{i' \in \mathcal{N}_i} \left[1 - \delta_{(x_i = x_{i'})} \right] \tag{4.16}$$

where $\mathcal{N}_i$ denotes the set of neighbor pixels of i.

Finding the global minimum in equation 4.14 is non-trivial, due to the large number of possible configurations for pixel labels. Hence, Iterated Conditional Modes (ICM) [18] algorithm is used to obtain a suboptimal solution, with a fast convergence.

Now, the MRF model is introduced in the segmentation framework by substituting the prior probability $p(x_i = j)$ in equation 4.4 for the MRF prior $p(x_i = j|x_{\mathcal{N}_i})$, which takes the labels of the neighbor pixels into account.

$$W_{ij} = \frac{p(y_i|x_i, d_i)p(x_i = j|x_{\mathcal{N}_i})}{p(y_i|d_i)} \tag{4.17}$$

where $p(x_i = j|x_{\mathcal{N}_i})$ has the form of equation 4.10.

The algorithm is implemented in a multiresolution Gaussian pyramid. As further discussed in Xiao et al. [129], this implementation justifies the assumption of a Gaussian Probability Density Function (PDF) for US images in general. Ideally, the PDF should be derived from the ultrasound physics of image formation. However, in a multiresolution implementation, lower-resolution ("blurred") solutions are shown to locally satisfy the Gaussian assumption. Thus, the classification and the distortion field estimation are performed at each scale, and both results are used to initialize the solution search at the next finer scale. Therefore, in each scale, the algorithm is only adjusting rather than solving from scratch for the parameters. Note that the Gaussian PDF assumption would not hold if the method were applied at a single resolution.

4.3 Automatic initialization process

Zhang et al. [136] proposed a MRF-MAP methodology for removing the bias field removal and segmenting brain tissue on MRI. Later, this work was adapted to US images by Xiao et al. [129]. One of the characteristics of brain MRI is its histogram, which is commonly bimodal (see Figure 4.2). Thus, the MRI histogram has two distinctive peaks, hence the mixture of Gaussian representation and the EM application seem a feasible approach. However, as far as we have observed in US images, the histogram of the background almost always overlaps with the histogram of the lesion information. Figure 4.2 shows a comparison between the histogram of a brain MRI and a BUS image where the MRI histogram is clearly bimodal, and on the US image, the two peaks overlap.

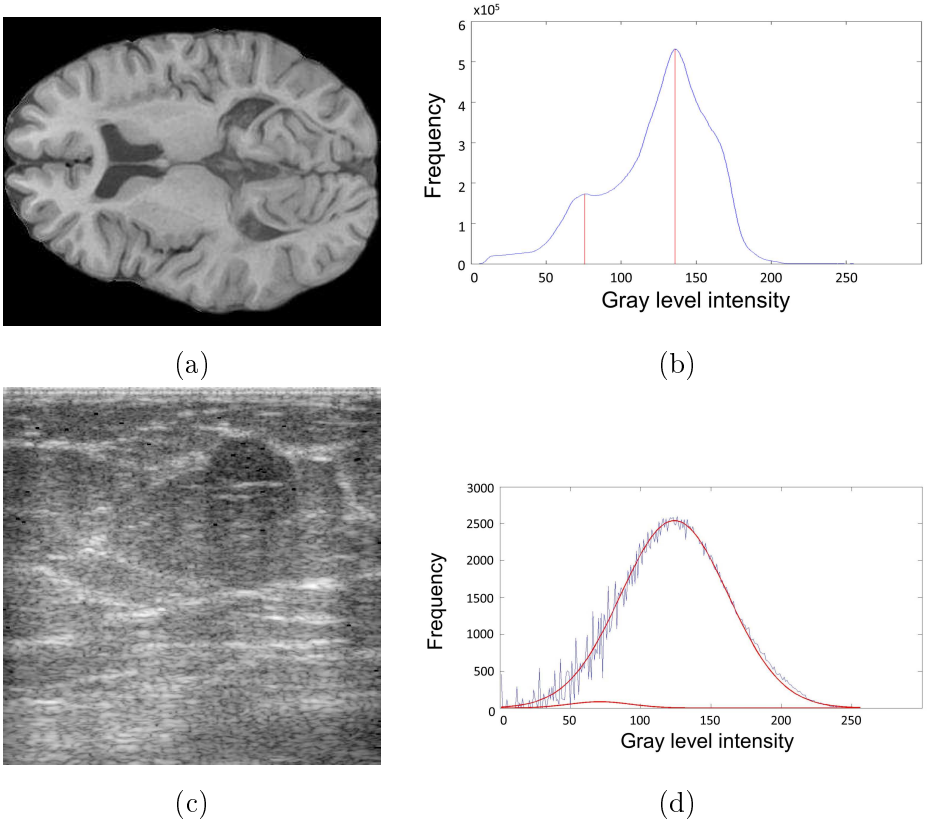
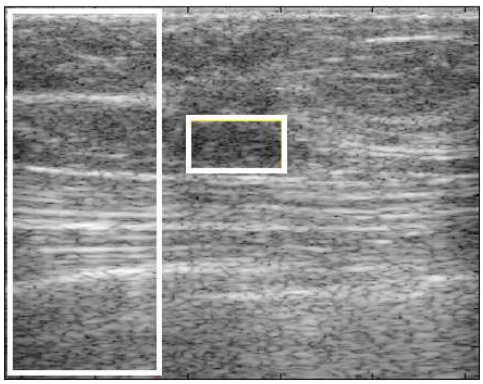


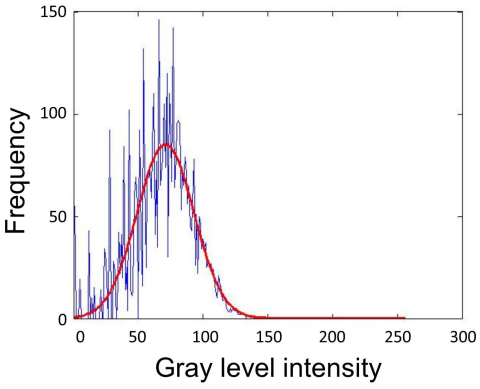
Figure 4.2: Intensity distributions of a brain MRI and a BUS. (a) Brain MRI. (b) Corresponding histogram. (c) Original BUS image. (d) Corresponding Gaussian distributions (in red). The larger Gaussian corresponds to the background and the smaller to the lesion. Both plotted over the image histogram (in blue).

The MRF-MAP proposal solved this problem by selecting the Gaussian distribution

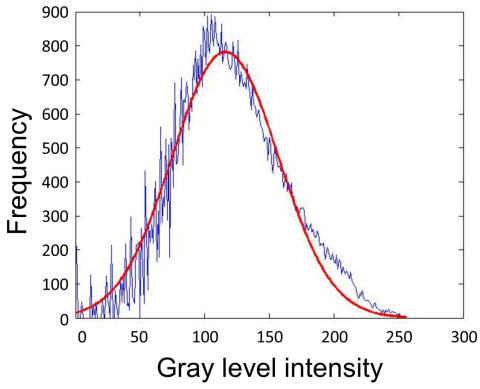
parameters (mean and standard deviation) empirically, analyzing the intensity histograms of the lesion region and the background by manually defining two different windows and manually fitting the best Gaussian distribution to them, as illustrated in Figure 4.3.



(a)



(b)



(c)

Figure 4.3: Histograms of two regions in an US image. (a) B-mode BUS image with the two regions delimited by rectangles. (b) Histogram of the lesion (obtained from the central window). (c) Histogram of the background (obtained from the left window). Both histograms depicted in blue and Gaussian distributions in red.

The main disadvantage of this initialization proposal is the cost in terms of user interaction. As far as we determined, the initialization step takes longer than 2 minutes and requires a previous image processing knowledge, which is an important drawback for radiologists. We propose to improve the original initialization step including spatial information. Due to the difficulty of implementing a fully automatic initialization using the intensity information alone, the initialization reduces the user interaction from the empirical choice of parameters to only one-click. Thus, the user selects the approximate location of the lesion, and a small window is defined to characterize the lesion pixels. Our experiments

show that most of the images in the dataset present lesions within 1 and 2 cm. Although this choice implies including unwanted background information in the lesion window in some cases, it is not critical for the results since the number of background pixels is much lower than the lesion information. In addition, it is important to ensure that the window always includes as much lesion information as possible to statistically model the information as a Gaussian distribution. Taking these considerations into account, the size of the lesion window is fixed to 2 cm. Subsequently, a larger window is opened to obtain a representation of the background information. To obtain the most accurate background information as possible, the pixels in the lesion window are ignored. However, some lesion pixels can still be part of the background in cases where the lesion is bigger than the predefined lesion window size, but according to our experiments, the number of these can be neglected compared to the size of the background. This methodology is named Local Lesion Information (LLI) strategy and can derive in two different initialization approaches depending on how the original MRF-MAP method is applied: Local Lesion Information at the Global image (LLIG), which uses the whole image, or Local Lesion Information at the Partial image (LLIP), which uses a partial image obtained by cropping the image by means of the background window. Figure 4.4 shows an example of LLI initialization using both windows and the point placed by the user.

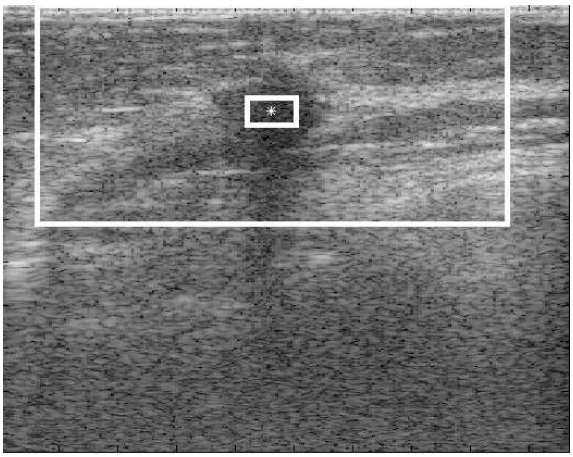


Figure 4.4: Local Lesion Information (LLI) acquisition. The inner rectangle contains pixel values for the lesion description and the outer rectangle contains background information.

A third alternative to improve the initialization step is also proposed, in which the histogram information is retrieved from 4 small windows surrounding the lesion to ensure that no pixel belonging to the lesion is used to compute the background histogram. This proposal is named Local Lesion Surrounding Information (LLSI). Figure 4.5 shows an

example.

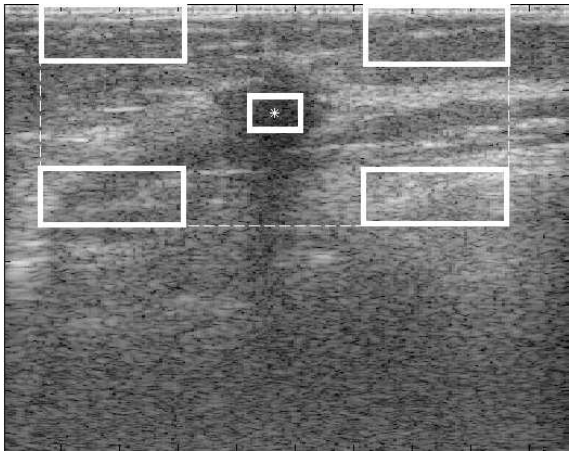


Figure 4.5: Local Lesion Surrounding Information (LLSI) acquisition. The central rectangle contains pixel values for the lesion description, and the outer rectangles contain background information.

4.4 Results

4.4.1 Image acquisition

Two breast B-mode US image datasets containing 212 lesion cases in total (98 malignant and 114 benign) were used to evaluate the performance of this proposal. The breast B-mode US images were collected from two different hospitals and ultrasound machines. Dataset S1 was acquired from the UDIAT Diagnostic Center of Parc Taulí Corporation (Sabadell, Spain) with a Siemens Acuson Sequoia C512 system 17L5 HD linear array transducer (8.5 MHz), and included 140 real BUS images: 44 malignant and 96 benign. Dataset S2 was obtained from the Churchill Hospital (Oxford, England) with a Zonare z.one system and a L10-5 linear array transducer (8.5 MHz), and included 72 real breast US images: 54 malignant and 18 benign. In this study, only 1 image per case was used in the experiments and an experienced radiologist performed the case selection on the basis of the most representative cases for all pathologic types. Table 4.1 provides details of the two datasets in terms of lesion type. All of the images involved in this work were previously made anonymous to accomplish the rules issued by the Ethical Committee of both hospitals concerning the data confidentiality.

One of the aims of this work is the evaluation of the segmentation results taking into

Table 4.1: B-mode ultrasound dataset

		Dataset S1	Dataset S2	Total	Percentage
Benign	Cyst	56	2	58	27.35%
	FA	34	13	47	22.17%
	Other benign	6	3	9	4.24%
Malign	IC (common type)	27	42	69	32.55%
	IC (uncommon type)	14	8	20	9.44%
	DCIS	3	6	9	4.25%
	Total	140	72	212	100%

account the type of lesion. Within malignant and benign cases, this study focuses on different specific pathologies: cyst, 58 images; FA, 47 images; other benign masses, 9 images; Invasive Carcinoma (IC;common type), 69 images; IC (uncommon type), 20 images; and DCIS, 9 images. IC (common type) lesions include IDC and DCIS plus IDC. IC (uncommon type) includes mucinous carcinoma and ILC. Other benign masses includes papilloma, fibrosis and lymphoma. An example of each lesion type is shown in Figure 4.6. The classification of these types is based on the radiologists' experience, in line with the existing literature [50].

4.4.2 Ground truth generation

For the evaluation of the segmentation results, Ground Truth (GT) provided by expert radiologists is needed. However, US images present a high inter-observer variability, and a consensus GT is commonly used. To generate this consensus GT, the one of the most extended techniques was used: Simultaneous Truth And Performance Level Estimation (STAPLE) [126]. This algorithm formulates the scenario as a missing-data problem, which provides a hidden GT estimation inferred from the experts' segmentation using an EM algorithm.

4.4.3 Inter-observer study

To generate the consensus ground truth, we first analyzed the inter-observer agreement to determine whether the differences between the experts are significant. We randomly selected 50 images regardless of the dataset they belonged to, and performed manual segmentations by an expert radiologist and 5 biomedical engineers to generate a consensus Ground Truth (GT) using the STAPLE algorithm. Table 4.2 shows the results obtained

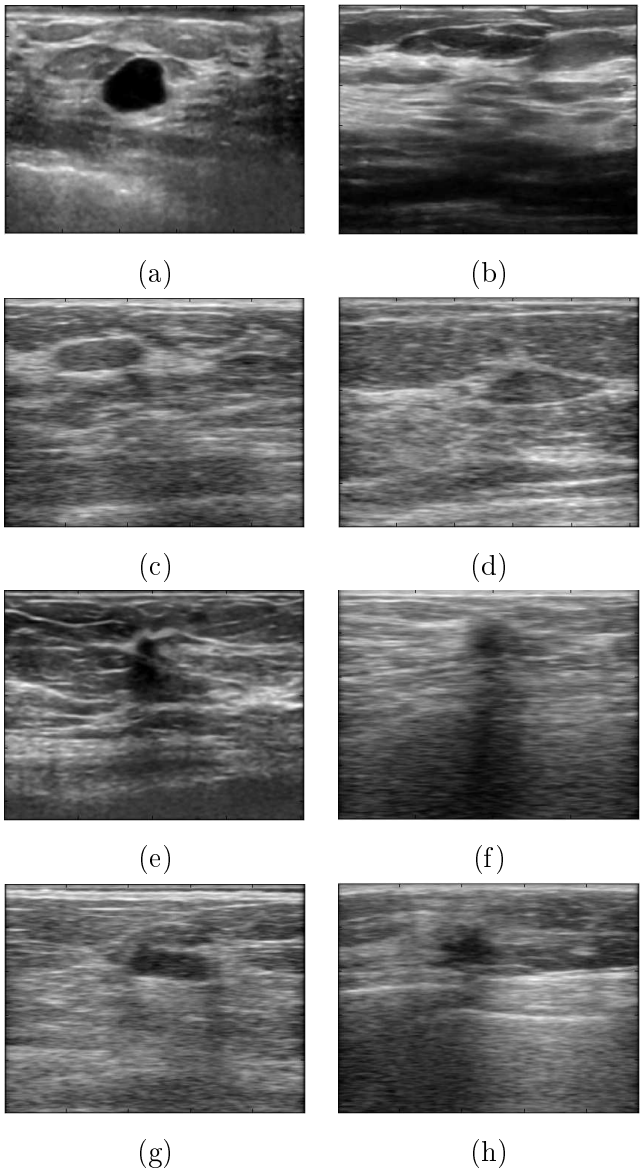


Figure 4.6: Example of each type of lesion: (a) cyst, (b) FA, (c) fibrosis, (d) mucinous carcinoma, (e) IDC, (f) DCIS, (g) DCIS+IDC and (h) ILC.

to compare each expert segmentations with the consensus GT. Using the DSC value as a reference, note that there are no significant differences between the radiologist's and the engineers' segmentations as shown in Figure 4.7. This finding indicates that all raters follow a similar segmentation criterion and that their accuracy is equivalent. Thus, to evaluate the entire dataset, we have generated a consensus ground truth using the segmentations from a radiologist and 5 biomedical engineers for the entire dataset.

Table 4.2: Comparison between of biomedical engineers and a radiologist segmentations

User	Sensitivity	Specificity	PPV	AO	DSC
Radiologist	0.871	0.997	0.968	0.845	0.914
Engineer 1	0.806	0.999	0.991	0.800	0.885
Engineer 2	0.859	0.998	0.997	0.840	0.910
Engineer 3	0.873	0.997	0.970	0.846	0.912
Engineer 4	0.901	0.994	0.919	0.827	0.900
Engineer 5	0.878	0.998	0.969	0.852	0.915

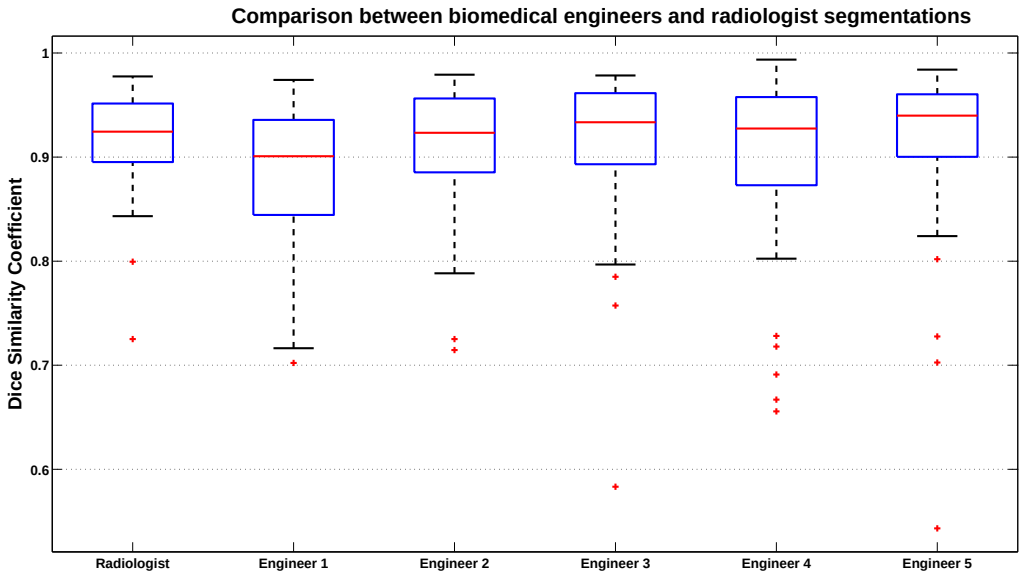


Figure 4.7: Box plot charts comparing the biomedical engineers' and the radiologist's segmentations.

4.4.4 Quantitative and qualitative results

An initial experiment was performed to compare the original MRF-MAP work [129] with the different proposals based on the LLI initialization (LLIG, LLIP, LLSI). Figure 4.8 shows an example of qualitative segmentation results for different images from both datasets using the LLSI approach, where (a and e) show the original BUS images. Their estimated distortion fields, and the restored images are shown in (b and f) and (c and g) respectively. Finally, (d and h) show the segmentation results.

Table 4.3 details the mean values for all the measures chosen. In this experiment, the cases where the algorithm results do not overlap with the actual lesion were included with a DSC value of 0. This decreases the mean values reported in Table 4.3. Analysis of

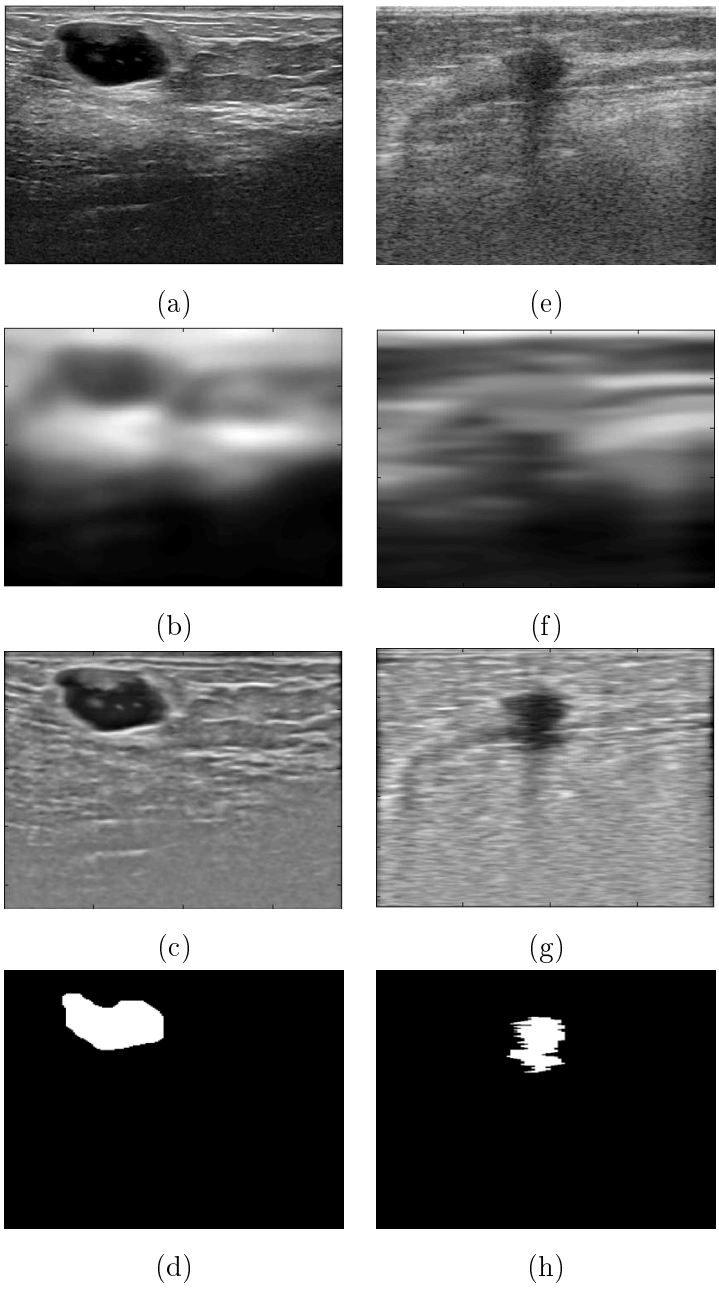


Figure 4.8: Lesion segmentation results. (a-d) Results for an image from Dataset S1. (e-h) Results for an image from Dataset S2. (a, e) Original image. (b, f) Estimated distortion field. (c, g) Corrected image. (d, h) Segmentation result.

this table shows that 2 of the 3 initialization proposals (LLIG and LLSI) improve the segmentation results and considerably reduce the user interaction in comparison with the original MRF-MAP proposal.

Table 4.3: Segmentation results using different evaluation criteria for each segmentation proposal

Method	Sensitivity	Specificity	PPV	AO	DSC
Original	0.565	0.991	0.729	0.508	0.610
LLIG	0.557	0.996	0.806	0.518	0.635
LLIP	0.451	0.997	0.693	0.425	0.519
LLSI	**0.604**	**0.995**	**0.804**	**0.550**	**0.663**

Figure 4.9 shows the segmentation results using DSC and AO values for all the images. Analyzing these diagrams for both measures, we can see that all median values of the results for each method are within the confidence interval of the other methods, which means that there are no significant differences between the methods. However, as shown in Table 4.3, the LLSI approach obtained the best results, while the LLIP approach obtained the worst. The way in which the LLSI approach acquires the background information (4 windows surrounding the lesion) avoids the inclusion of shadows and other intensity related artifacts, improving the results.

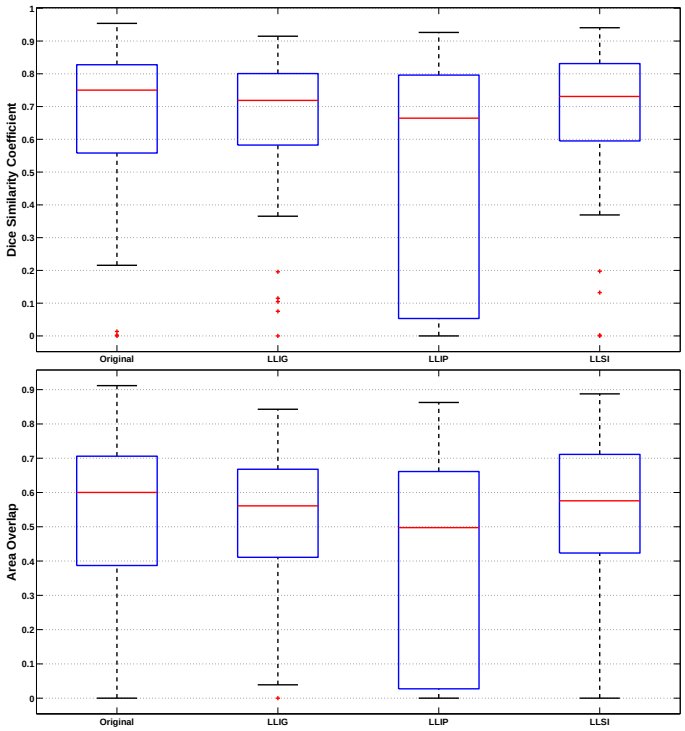


Figure 4.9: Box diagrams of DSC and AO measures for all the methods.

4.4.5 Lesion type comparison

In this subsection, the segmentation proposals are compared depending on the lesion typology. This allows to obtain conclusive results regarding the best segmentation approach for both datasets and in relation to the lesion type, as well as which pathologies are more complex for the segmentation process. This information is very valuable in terms of further work in order to focus efforts on specific pathologies. Table 4.4 shows the comparison results for the different cancerous lesion types and Table 4.5 summarizes the results for benign lesions.

Table 4.4: Comparison of measure means obtained for cancerous lesions (98 images)

	Method	Sensitivity	Specificity	PPV	AO	DSC
IC (common type) (69 images)	Original	0.551	0.989	0.688	0.472	0.578
	LLIG	**0.605**	**0.993**	**0.795**	**0.536**	**0.665**
	LLIP	0.435	0.997	0.703	0.409	0.517
	LLSI	0.611	0.991	0.752	0.529	0.646
IC (uncommon type) (20 images)	Original	0.429	0.997	0.753	0.411	0.525
	LLIG	0.426	0.994	0.78	0.390	0.520
	LLIP	0.439	0.996	0.831	0.410	0.545
	LLSI	**0.476**	**0.994**	**0.818**	**0.428**	**0.567**
DCIS (9 images)	Original	0.616	0.995	0.742	0.549	0.655
	LLIG	**0.606**	**0.997**	**0.796**	**0.585**	**0.686**
	LLIP	0.553	0.996	0.758	0.508	0.613
	LLSI	0.611	0.997	0.709	0.572	0.655

Analysis of these tables makes it possible to conclude that LLSI approach obtains better results for IC (uncommon type), cyst, FA, and other benign lesions (134 images of 212). The LLIG approach is better for DCIS (9 images) and IC (uncommon type; 69 images), obtaining slightly better results than LLSI. The LLIP approach and the original MRF-MAP [129] proposal do not provide better performances for any type of lesions. This finding is also graphically depicted in Figure 4.10, where (a) shows the DSC values for each type of lesion and approach, and (b) shows the percentage of mis-segmented images for each lesion typology and approach. A mis-segmented image is defined by a DSC of less than 0.6, which is below the commonly acceptable range. In summary, LLSI obtains the best results comparing all of the images (see Table 4.3) and the best results for 4 different types of lesions, which represent the 63% of the dataset.

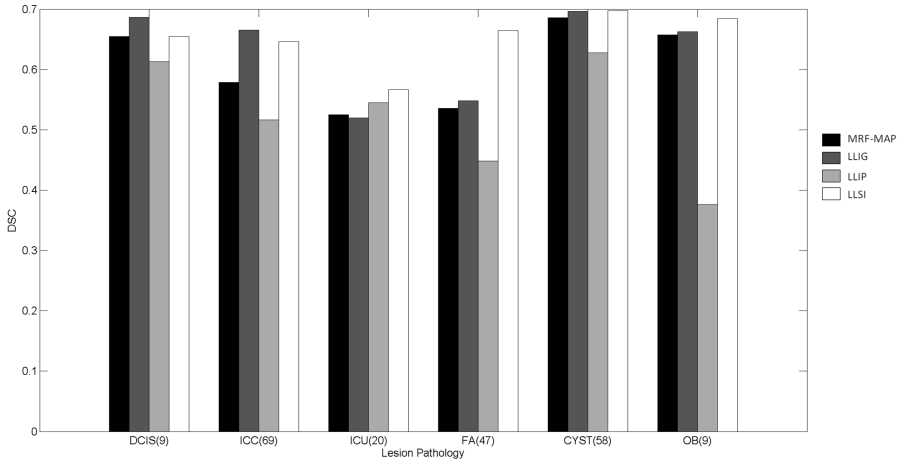
Finally, Table 4.6 plots a summary of the results of the LLSI method, which has obtained

Table 4.5: Comparison of measure means obtained for benign lesions (114 images)

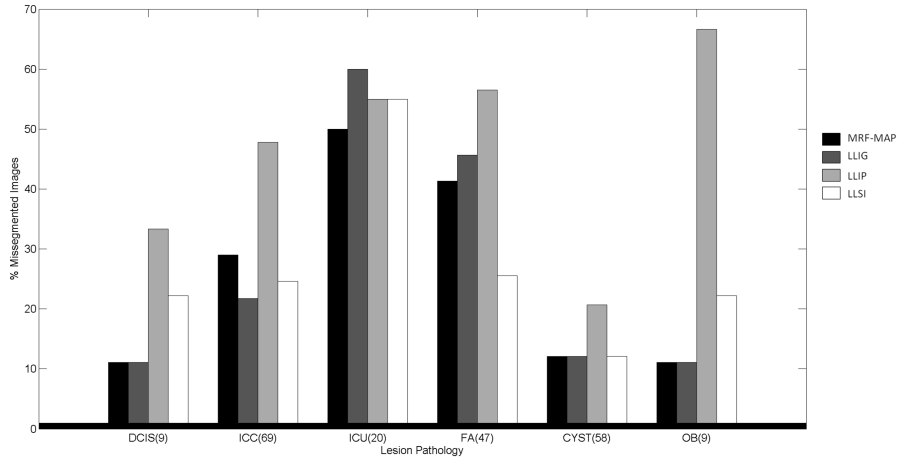
	Method	Sensitivity	Specificity	PPV	AO	DSC
Cyst (58 images)	Original	0.584	0.999	0.870	0.576	0.686
	LLIG	0.605	0.999	0.873	0.587	0.696
	LLIP	0.546	0.998	0.786	0.528	0.628
	LLSI	**0.607**	**0.998**	**0.854**	**0.588**	**0.698**
Fibroadenoma (47 images)	Original	0.543	0.982	0.590	0.451	0.535
	LLIG	0.488	0.996	0.757	0.445	0.548
	LLIP	0.387	0.998	0.685	0.363	0.448
	LLSI	**0.618**	**0.994**	**0.782**	**0.557**	**0.664**
Other benign masses (9 images)	Original	0.543	0.999	0.852	0.530	0.657
	LLIG	0.544	0.998	0.873	0.537	0.663
	LLIP	0.303	0.994	0.547	0.299	0.376
	LLSI	**0.546**	**0.999**	**0.986**	**0.541**	**0.684**

the best results for both datasets, indicating the DSC value, the DSC value excluding the mis-segmented images, the percentage of mis-segmented images and the number of cases for each type of lesion. In summary, the table indicates the performance of each segmentation approach on every type of lesion.

Focusing on the total dataset results (datasets S1 and S2), the best performance is obtained on cyst segmentation with a DSC of 0.69, and only 12.06% of cystic cases are mis-segmented, as they usually show a well-defined hypoechoic lesion. FA lesions obtained the worst results among the benign lesions (25.53% of mis-segmented images). These kinds of abnormalities usually show intensity inhomogeneities within the lesion (see Figure 4.11(a)) leading to under-segmented lesions. Cancerous lesions (common IC, uncommon IC and DCIS) obtained a worse performance than benign ones, all with a higher percentage of mis-segmented images than cysts (24.63%, 55%, and 22.22%, respectively). These kinds of lesions usually show a prominent shadow (see Figure 4.11(b)). Although the LLSI approach tends to avoid these artifacts, however, the segmentation fails when the image contains a large lesion. This problem could be solved in the future by using adaptable windows depending on the lesion size to improve the background information acquisition. The number of mis-segmented images of uncommon IC is higher than the other carcinomas due to the fact that the images are mostly composed of ILC, which has a spiculated shape with blurred contours, increasing the complexity of the segmentation process. Although the quantity of DCIS images was not enough to extract definitive conclusions, the results indicate that this type of lesion follows the tendency of common IC in terms of mis-



(a)



(b)

Figure 4.10: Segmentation results for each lesion type and segmentation approach. (a) Dice similarity coefficient. (b) Percentage of mis-segmented images. ICC indicates IC (common type); ICU, IC (uncommon type); and OB, other benign lesions.

segmented images. However, when removing the mis-segmented images, the lesions were segmented better than the other carcinomas. Finally, the number of other benign images was also not conclusive, but the high number of mis-segmented images can be explained by the lesion heterogeneity within this group.

Table 4.6 also gives information about the behavior of the segmentation method using the two different datasets, S1 and S2. Although the number of cases in each dataset was

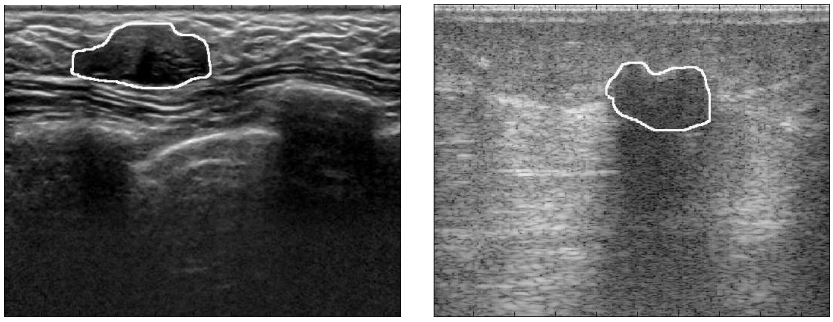
Table 4.6: Measures for each lesion type using LLSI (best proposal)

	Typology	DSC	DSC (no mis-segmented)	Error	Cases
Dataset S1	IC (common type)	0.654	0.775	25.92%	27
	IC (uncommon type)	0.528	0.737	57.14%	14
	DCIS	0.603	0.904	33.33%	3
	Cyst	0.694	0.793	12.5%	56
	FA	0.686	0.815	27.27%	33
	Other benign masses	0.768	0.768	0%	6
	All lesions	0.680	0.794	22.14%	140
Dataset S2	IC (common type)	0.641	0.762	23.8%	42
	IC (uncommon type)	0.554	0.735	50%	6
	DCIS	0.681	0.817	16.67%	6
	Cyst	0.820	0.820	0%	2
	FA	0.614	0.846	30.7%	13
	Other benign masses	0.518	0.670	66.66%	3
	All lesions	0.632	0.787	27.77%	72
Total Dataset	IC (common type)	0.646	0.764	24.63%	69
	IC (uncommon type)	0.567	0.748	55%	20
	DCIS	0.684	0.924	22.22%	9
	Cyst	0.698	0.792	12.06%	58
	FA	0.664	0.798	25.53%	47
	Other benign masses	0.684	0.735	22.22%	9
	All lesions	0.662	0.792	24.05%	212

different, it can be observed that the results were quite similar for both datasets. The image quality differences between both datasets can be observed in Figure 4.11, where an image from each dataset is depicted. Dataset S1 images present more details of the breast tissues, while images from Dataset S2 present a higher influence of speckle.

4.4.6 Comparison with other methods in the literature

In this subsection, a comparison of the obtained results with the results reported in the most relevant works in the literature is shown. Specifically, the segmentation methods in this comparison are recent algorithms (from 2009), which need practically no user interaction (1 click) or none interaction at all. The results reported by these works are summarized in Table 4.7.



(a) (b)

Figure 4.11: Example of FA and IC lesions. (a) Fibroadenoma lesion in which intensity inhomogeneities within the lesion are present. (b) Invasive carcinoma in which a prominent shadow is formed. The images were acquired from different ultrasound systems. Lesion boundaries have been provided by expert manual delineation.

Table 4.7: Comparison of performance with other recent lesion segmentation methods

Method	DSC	Number of images
Cui et al. (2009) [30]	0.85	250
Liu et al. (2010) [86]	0.88	46
Massich et al. (2010) [92]	0.78	25
Zhang et al. (2010) [134]	0.91	347
Hao et al. (2012) [64]	0.85	480
Proposed method	0.66	212

Analysis of this table shows that our proposal obtained the worst results. However, there is a set of conditional factors that makes this straight comparison unreliable. The main factor is the dataset used in the assessment of the methods. Each dataset used is completely different; acquired from different ultrasound systems, differs in the number of images, and contains different lesion types. The results of the segmentation methods vary depending on the image quality, the presence of artifacts, and the lesion types. As we have shown before, it is easier to segment cystic lesions than, for instance, fibroadenomas. In addition, we can not know if the methods are tuned to obtain good results in their respective datasets, or if they would work properly with different ones. It is clear then, that for a fair comparison, all the algorithms should be tested using the same datasets.

Due to the fact that we could not implement these other methods, the correctness of our segmentation proposal must be assessed in other terms. We can take advantage of the inter-observer study to see how well an expert radiologist segments lesions in the

dataset used. In this case, the radiologist obtained a DSC value of 0.91. Thus, it is clear that a computerized method can hardly obtain better results than an experienced trained radiologist.

The proposed method obtained a DSC value close to 0.7, which is commonly considered a good result for a segmentation method. However, this average DSC measure contains 24.05% mis-segmented images. Focusing on Dataset S1, the images that induce our method to segmentation errors (22.14%) are mostly images where the lesions are zoomed-in on (see Figure 4.12(a) and (b)), or where the appearance of the lesion is similar to healthy tissue (see Figure 4.12(c) and (d)). Excluding these errors, a DSC close to 0.8 was obtained, which indicates that the method segments the lesions well when the above mentioned artifacts are not present (almost 80% of the images in the dataset).

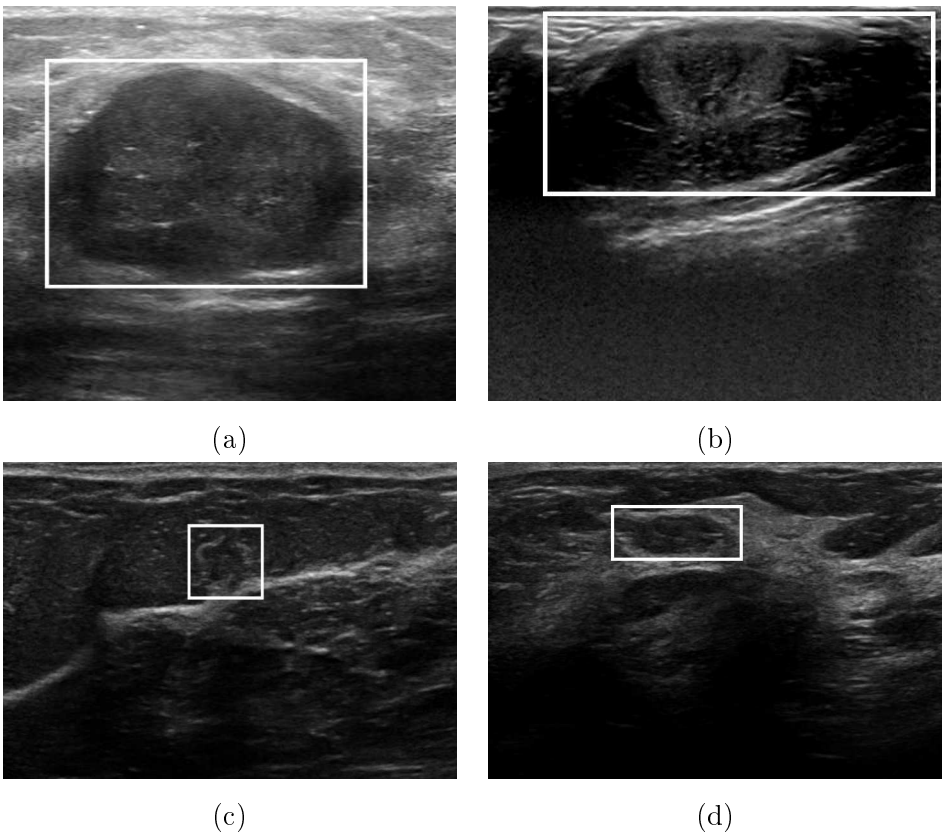
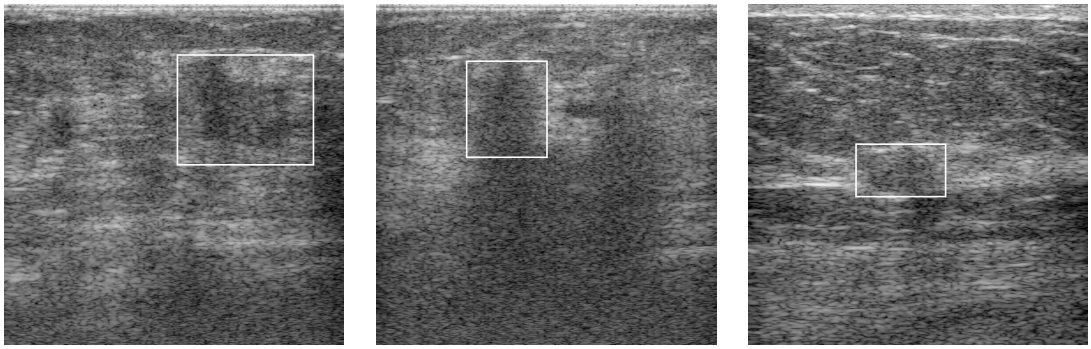


Figure 4.12: Example of Dataset S1 images which induce a segmentation error. (a) and (b) are zoomed-in lesions. (c) and (d) lesions with similar appearance to the normal tissue. The location of the lesion is labeled with a rectangle.

Regarding Database S2, most of the segmentation errors are caused by the low quality of the images. Figure 4.13 shows different images that induce the algorithm to a

mis-segmentation results (27.77% of the images in Dataset S2) due to their low quality. Excluding these errors, a DSC close to 0.8 was also obtained.



(a) (b) (c)

Figure 4.13: Example of Dataset S2 images that induce a segmentation error due to their low quality. The location of the lesion is labeled with a rectangle.

4.5 Summary and conclusions

In this chapter, an exhaustive analysis of the proposal by Xiao et al. [129] (MRF-MAP) and a set of initialization improvements was performed using breast US images supplied by two different hospitals. A detailed analysis of the results obtained shows that the LLSI initialization proposal obtained the best results for both datasets. Moreover, the original proposal by Xiao et al. [129] is considerably improved in terms of segmentation results and the initialization procedure. This enhancement is due to the information acquisition procedure in LLSI in which the background information is obtained by using 4 windows surrounding the lesion. Thus, the shadowing effect is minimized, and only background information is collected to correctly initialize the Gaussian distribution, which models the non-lesion tissue.

Another goal of this work was to evaluate the robustness of the segmentation approaches. This evaluation was described in the previous section, with similar results obtained from both databases, which indicates that the method is not highly influenced by the acquisition system used. However, the method is induced to segmentation errors in some images. One reason is the low quality of some images, and secondly, some artifacts present in the images, such as zooming. The results obtained excluding the mis-segmented images indicate that we must focus on solving the influence of these artifacts to enhance the performance of our method as future work. In addition, future work will also focus on multiple lesion

segmentation as this is a current limitation of the method.

Finally, we studied the segmentation results by lesion type. As expected, we conclude that cystic lesions are the easiest to segment. Fibroadenoma and cancerous lesions have more mis-segmented images and consequently lower mean DSC values due to the inherent artifacts related to the lesion type. These results indicate that more efforts in computer-aided diagnosis system research must be done to improve the segmentation of malignant lesions due to their impact on patient health.

5

Breast lesion segmentation using elastography information

5.1 Introduction

Mammography and ultrasonography are the diagnostic methods that have shown the highest sensitivity in the diagnosis of breast cancer. However, both techniques present some limitations. Mammography performed in dense breasts may yield false-positive results [105]. US is sensitive in the detection of lesions, but the specificity in cancer detection is low since most of the detected solid lesions are benign. This leads to an increase of unnecessary biopsies causing discomfort to the patients and increasing costs [58].

To overcome these limitations and obtain a more accurate characterization of breast lesions, US elastography was introduced. Breast US elastography is a recent diagnostic technique based on imaging the difference in stiffness between cancerous and benign tissues. Elastography is based on the principle that, in general, normal tissue is easily deformed while cancerous tissue is stiffer. Hence, cancerous tissues are more easily identifiable in an elastogram than in other screening techniques. Clinical studies support the

use of this technique to enrich current screening methods [56, 73, 137]. Following the prominent relevance of elastography in clinical environments, it is reasonable to assume that lesion segmentation methods could also benefit from this complementary information. In this chapter, we propose the inclusion of elastography information in the MRF-MAP segmentation framework presented in the previous chapter. This involves an adaption using a bivariant formulation as will be described in the following sections. This proposal is assessed using data acquired using two different elastography techniques, which are also presented in the following section.

5.2 Breast Elastography

Elasticity is the property of a tissue to be deformed by an external force and, when the force is removed, to return to its original shape. The elasticity is measured by a physical quantity called Young's modulus and expressed in pressure units (kPa). Young's modulus E can be defined as the relationship between an external compression (or stress S) applied to a solid tissue and its deformation (or strain e) inside the tissue (see Figure 5.1).

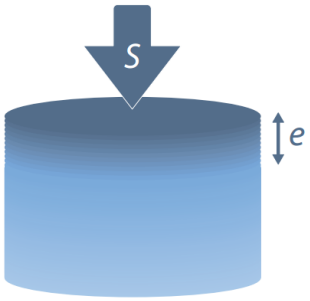


Figure 5.1: Deformation (e) of a soft solid under an external stress (S).

Often a linear relationship is assumed between stress and strain, and is expressed in equation 5.1.

$$E = \frac{S}{e} \tag{5.1}$$

Different tissues are expected to respond differently to the external forces. This is the basis of manual palpation, a common cancer diagnosis technique in medical practise. In general, fatty tissue is easily deformed while cancerous tissues are stiffer [58]. Elastography is an imaging technique that evaluates the elasticity of the tissues, but is more sensitive and

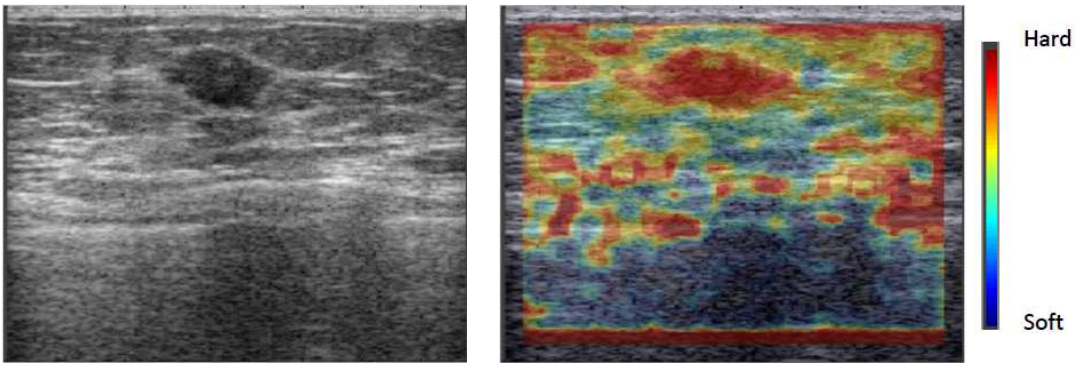
less subjective than palpation. Breast elastography can be performed using two different techniques: Compression Ultrasound Elastography (USE) and Shear-Wave Elastography (SWE).

5.3 Compression Ultrasound Elastography (USE)

Compression or conventional elastography is based on the application of a compressive force using a conventional transducer on the breast and on the measurement of the deformation of the tissues.

This technique allows only qualitative assessment due to the fact that the external compression force is unknown, allowing only the calculation of the deformability ratio (strain) by measuring variations in the RF of the US beam before and after compression [100], and not the absolute elasticity value [58].

In qualitative assessment, the tissue elasticity is encoded in a color map shown as an overlay in a B-mode, as shown in Figure 5.2. The different colors represent different elasticity levels (in Figure 5.2 stiff tissues are represented in red while soft tissues in blue). Note that some works encoded the color map regarding the strain values (high strain represented in red while low strain in blue), which is the inverse mapping of Figure 5.2 (high strain indicates soft tissue while low strain indicates stiffness).



(a) (b)

Figure 5.2: Figure (a) plots the B-mode image of a carcinoma and (b) its elastography information represented by a color overlay. Note that red represents high values of stiffness (hard tissue) and blue low values (soft tissue).

In order to improve the objectivity of the elastography, several works proposed scoring systems to describe the elastic behavior of a lesion [49, 60, 73, 81]. They proposed different

scales to classify an elastogram based on colors. For instance, Itoh et al. [73] proposed a scale, assuming blue as the expression of stiffness and green as softness, that includes: score 1 for soft strain nodules with the entire lesion colored in green; score 2 for lesions with a mosaic pattern of green, red and blue; score 3 for lesions with blue in the middle and green in the periphery; score 4 for stiff strain nodules with the entire lesion colored in blue without the surrounding area and finally, score 5 for stiff strain lesions with a blue surrounding area (see Figure 5.3). Later, alternative, but similar, scales were proposed [49, 60, 81].

In order to obtain a more objective assessment of the elastic behavior between lesions and normal tissue, a numerical ratio (called strain ratio) between the deformation of the lesion and the normal tissue, and the difference in lesion size before and during the compression (referred to as length ratio) were proposed . The strain ratio is measured from a ROI adjusted to the lesion boundaries and a comparable ROI placed in the adjacent fat tissue [5, 120, 138]. The length ratio (or width ratio) is calculated by measuring the maximal horizontal length of the lesion in the USE image divided by the corresponding length measured in the B-mode image [5, 12, 104].

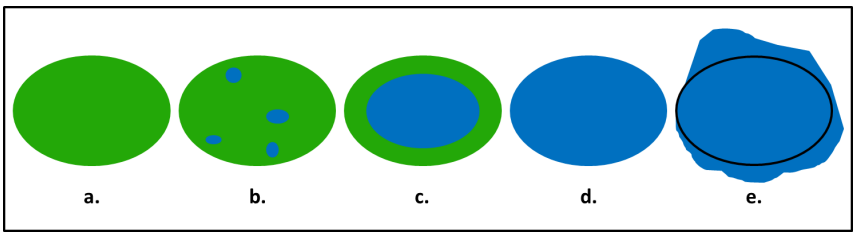


Figure 5.3: Five point scoring criteria according to Itoh et al. [73] for breast lesions. (a) Score 1: strain appears in the entire lesion. (b) Score 2: Strain is not seen in parts of the lesion. (c) Score 3: Strain appears only at the periphery of the lesion. (d) Score 4: No strain appears in the entire lesion. (e) No strain appears either in the lesion or the surrounding area.

5.4 Shear-Wave Elastography (SWE)

Shear-wave elastography is a technique of obtaining elastography images based on the combination of stress induced in a tissue by an ultrasonic beam, and an ultrafast imaging sequence capable of acquiring the propagation of the resulting shear waves in real time [14]. The system induces mechanical vibrations created by a focused US beam. An ultrafast (5000 frames/s) US acquisition sequence is used to capture the propagation of

the shear-waves. The tissue's elasticity is directly deduced by measuring the speed of wave propagation c, as indicated in equation 5.2

$$E = 3\rho c^2 \tag{5.2}$$

where ρ is the density of the tissue, which is assumed to be a constant in the body, i.e. it is very close to the density of water (1000 kg/m^3).

Shear-waves travel faster through hard tissue and therefore, hard tissue has larger Young's modulus value compared to soft tissue. After the local shear-wave velocity is recovered, the production of a two-dimensional map of shear elasticity is generated in real-time, where stiffer tissues are coded in red and softer tissues in blue, as shown in Figure 5.4

The production of the stress by the transducer rather than the operator (as applied in conventional ultrasound elastography) means that shear-wave elastography is more reproducible than conventional elastography. Since the absolute elasticity values of the tissues are obtained, values for maximum stiffness, mean stiffness and standard deviation can be calculated from a given ROI in the ultrasound acquisition machine.

Some studies established a relationship between the elasticity measures obtained with SWE and the Breast Imaging and Reporting Data System (BIRADS) classification, and established thresholds to differentiate between benign and malignant lesions [11, 42, 44, 119]. Other studies proposed a set of features extracted from the SWE, and analyzed the discriminative properties of these features in the task of distinguishing benign from malignant lesions [15, 63, 78].

5.5 The role of elastography on CAD

Elastography is a recent technique and, although it has been shown to improve the specificity of diagnosing lesions [15, 58], it is still not widely used in routine screening programs. For this reason, the role of the elastography on CAD systems is not relevant yet, and only several works investigated the use of computerized techniques with elastography imaging. These works can be grouped as to the stage of the CAD system in which they use elastography: lesion detection, segmentation or classification.

Regarding automatic lesion detection, only Zhang et al. [135] proposed a fully-automatic method for locating cystic lesions in BUS images. They used the same detection algorithm

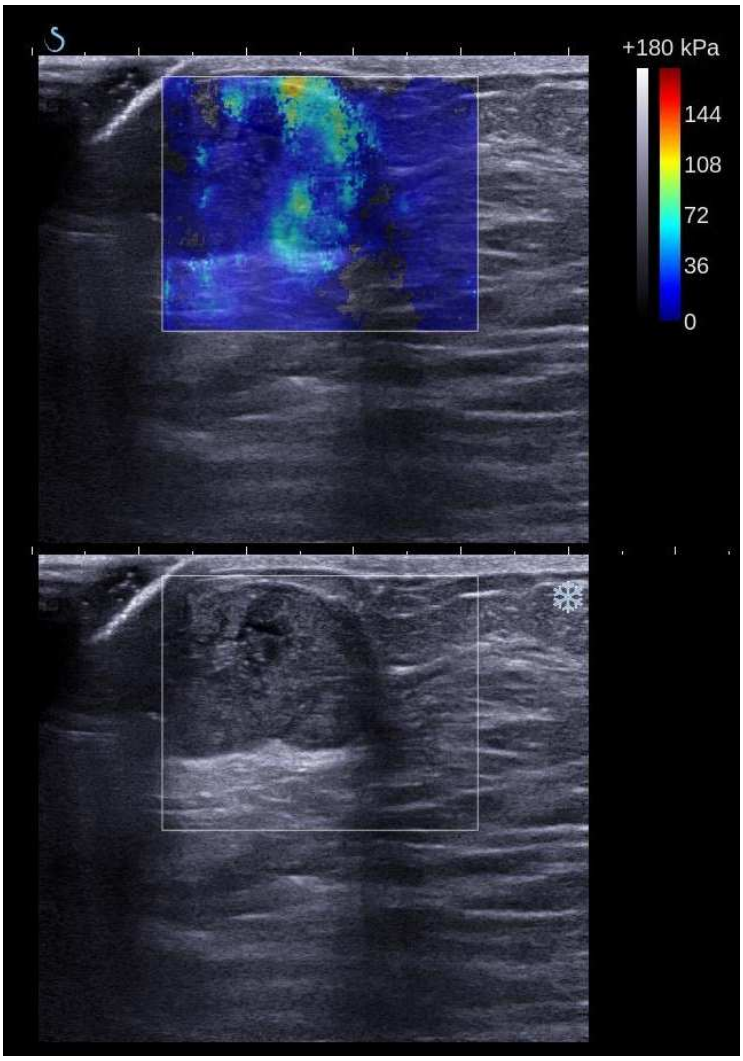


Figure 5.4: SWE of a fibroadenoma. The bottom image depicts the B-mode and the top image the B-mode with the elastogram overlayed.

presented in [134], which used a machine learning framework with a PBT classifier and Haar-like features [123] extracted from the image, but only extracting the features from pairs of B-mode and elastography images instead of a single B-mode.

More efforts in tissue segmentation can be found in the literature. Regarding lesion segmentation on BUS images, we can find the work of Von Lavante and Noble [124], who added strain features extracted from the RF signal to a graph-cut segmentation framework. Later, Nedevschi et al. [97] proposed the segmentation of elastography images using the EM algorithm. The method automatically initializes the EM, analyzing the peaks in the elastography histogram. Finally, Zhou et al. [139] proposed the inclusion of elastography

in a level-set segmentation framework, but they only evaluated their approach in phantom images.

Elastography data can also be useful in the segmentation of other organs, such as the liver or the prostate. This is the case of Techavipoo et al. [118], who proposed a semi-automatic algorithm to segment elastography images of the liver. In this work, the user interaction is needed to place two ROIs to initialize a histogram thresholding segmentation process. The segmentation is then refined with a morphological operation to remove artifacts. Liu et al. [87] also proposed the segmentation of elastography images of the liver. They presented an ACM method, where the contour is initialized with a coarse-to-fine transformation (Gaussian pyramid) [21]. In the field of prostate segmentation, we find the work of Mahdavi et al. [89], where they presented a method that combines B-mode and elastography information in an Active Shape Model.

As we mentioned in the previous section, elastography features are used to distinguish benign from malignant lesions in BUS images in clinical studies. Hence, some works focused their efforts on using elastography features for lesion classification. This is the case of Chang et al. [23], who extracted features of the lesion before and after the compression (volume, shape and contour differences) to train a SVM classifier. Moon et al. [96] evaluates a set of features (5 from the elastography and 6 from the B-mode) extracted from manual segmentations of the tumors, using a NN. Subsequently in [95], the same authors proposed an algorithm to pick the best image in an elastography video and used the best combination of features to classify a previously segmented lesion. Also Selvan et al. [108] evaluated a set of combined features from B-mode and elastography extracted from a segmentation performed with a level-set method.

5.6 Bivariate MRF-MAP segmentation approach

Elastography information can be seen as a grayscale image. Indeed, in clinical environments elastography information is commonly shown as an overlay in B-mode screening process assigning intensity value ranges to a different color to visually help radiologists, as seen in Figures 5.2 and 5.4. Hence, elastography information can be combined with B-mode information to improve segmentation results, mostly when the B-mode image information is not conclusive in providing the lesion location. We propose to extend the MRF-MAP approach to take both B-mode and elastography into account in a unique framework, by modeling both data in a bi-variate Gaussian PDF.

5.6.1 Image model

We assume that both images (B-mode and elastography) present intensity inhomogeneities and these are described by a multiplicative field. A logarithmic transformation yields this multiplicative model to an addition. Similar to the notation described in Chapter 2, let the observed and ideal log-transformed intensities be denoted respectively by $\mathbf{y}$ and $\mathbf{y}^*$, then $\mathbf{y} = \mathbf{y}^* + \mathbf{d}$ where $\mathbf{d}$ denotes the log-transformed intensity bias field. Given the class label $x_i \in \mathcal{L}$, where $\mathcal{L}$ denotes the label set. Note that now, $\mathbf{y}_i^*$ at pixel i refers to both B-mode and elastography values at that pixel. It is assumed that those values, $\mathbf{y}_i^*$, follow a bivariate Gaussian distribution with parameter $\theta(x_i) = (\boldsymbol{\mu}_{x_i}, \Sigma_{x_i})$, where $\boldsymbol{\mu}_{x_i}$ now is a vector of means for B-mode and elastography, and Σ_{x_i} the covariance matrix of class x_i, and defined as follows

$$p(\boldsymbol{y}_i^*|x_i) = g(\boldsymbol{y}_i^*; \theta(x_i)) = \frac{1}{2\pi det(\Sigma_{x_i})^{\frac{1}{2}}} exp\left[\frac{1}{2}(\boldsymbol{y}_i^* - \boldsymbol{\mu}_{x_i})^T \Sigma_{x_i}^{-1}(\boldsymbol{y}_i^* - \boldsymbol{\mu}_{x_i})\right] \tag{5.3}$$

Variables in bold type represent vectors to distinguish between notation of the MRF-MAP formulation in Chapter 4 and bivariate formulation. Experimental results corroborate the validity of the assumption of a bivariate Gaussian distribution for both lesion and background information in B-mode and elastography, as illustrated in Figure 5.5, where the intensity distribution for lesion and background are shown for elastography and B-mode, and a clear bivariant Gaussian distribution can be seen.

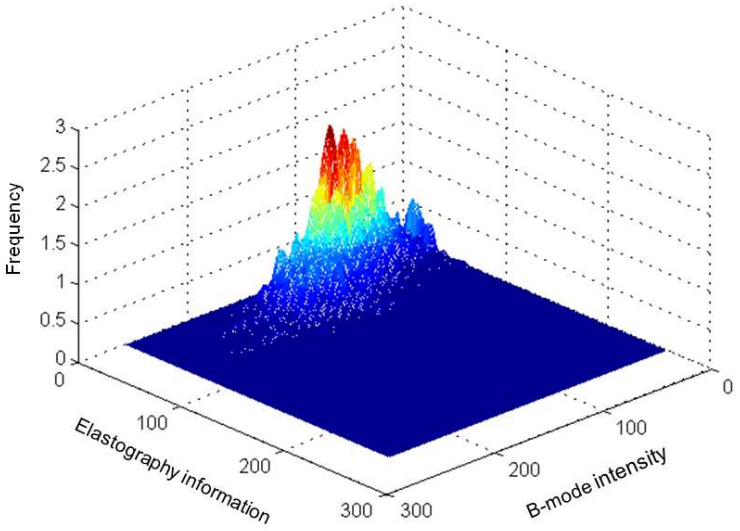
Considering the bias field, the above distribution can be rewritten in terms of the observed intensities $\boldsymbol{y}_i$ as

$$p(\boldsymbol{y}_i|x_i, \boldsymbol{d}_i) = g(\boldsymbol{y}_i - \boldsymbol{d}_i; \theta(x_i)) \tag{5.4}$$

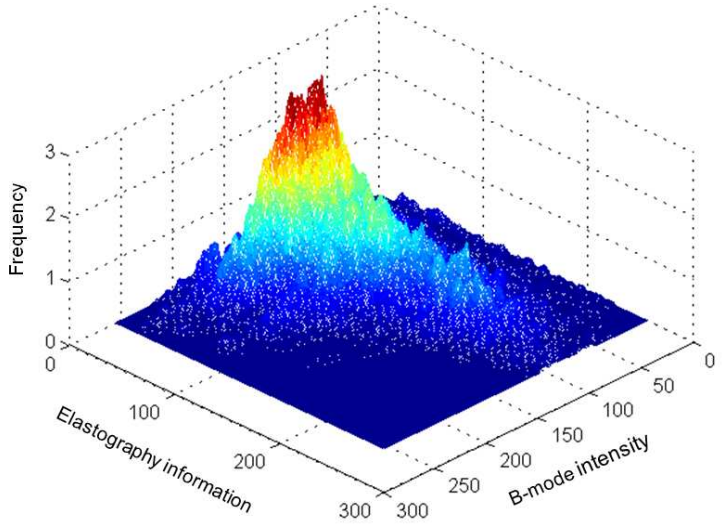
and the class-independent intensity distribution is denoted by

$$p(\boldsymbol{y}_i|\boldsymbol{d}_i) = \sum_{j \in \mathcal{L}} g(\boldsymbol{y}_i - \boldsymbol{d}_i; \theta(x_i)) p(x_i = j) \tag{5.5}$$

Thus, the intensity distribution is modeled as a bivariate Gaussian mixture and, assuming that the pixel intensities are statistically independent, the probability density for the entire image given the bias field is



(a)



(b)

Figure 5.5: Examples of bivariate distributions: 2D histograms showing B-mode and Elastography information of (a) lesion and (b) background.

$$p(\boldsymbol{y}|\boldsymbol{d}) = \prod_{i \in \mathcal{S}} \sum_{j \in \mathcal{L}} g(\boldsymbol{y}_i - \boldsymbol{d}_i; \theta(x_i)) p(x_i = j) \tag{5.6}$$

where $\mathcal{S}$ denotes the image pixel set.

5.6.2 Label estimation using MRF-MAP framework

Similarly to the B-mode formulation, the MRF prior model is introduced to keep the spatial information. It is assumed that neighboring sites have a direct relationship with each other and tend to have the same class labels. The probability density of a MRF model can be described as a Gibbs distribution [17] (see equation 4.10 in Chapter 4).

Following the same notation as in Chapter 4, the MAP of class labels is defined as

$$\hat{x} = \arg\min_{x \in \mathcal{X}} U(\boldsymbol{y}^*|x) + U(x) \tag{5.7}$$

where $U(\boldsymbol{y}^*|x)$ likelihood energy is the logarithm of the posterior probability, and it is defined as follows

$$U(\boldsymbol{y}^*|x) = \sum_{i \in \mathcal{S}} \left[\frac{1}{2} (\boldsymbol{y}_i^* - \boldsymbol{\mu}_{x_i})^T \Sigma^{-1} (\boldsymbol{y}_i^* - \boldsymbol{\mu}_{x_i}) + log(det(\Sigma)^{\frac{1}{2}}) \right] \tag{5.8}$$

and the prior energy $U(x)$ can be defined as in equation 4.16 of Chapter 4.

5.6.3 Expectation Maximization (EM) algorithm

To obtain the estimation of the distortion field, the EM algorithm is also used to update the label image and the intensity inhomogeneity field iteratively, initializing the bias field $\boldsymbol{d}$ to be zero. Since a fast convergence is needed in a few iterations, the ICM algorithm [18] is used (see Chapter 4).

Expectation step: calculates the posterior tissue class probability W_{ij} using the MRF prior model when the bias field $\boldsymbol{d}$ is known and using $p(x_i = j|x_{\mathcal{N}_i})$

$$W_{ij} = \frac{p(\boldsymbol{y}_i|x_i, \boldsymbol{d}_i) p(x_i = j|x_{\mathcal{N}_i})}{p(\boldsymbol{y}_i|\boldsymbol{d}_i)} \tag{5.9}$$

Maximization step: the bias field $\boldsymbol{d}$ is estimated when W_{ij} is known.

$$\boldsymbol{d}_i = \frac{[F\boldsymbol{R}]_i}{[F\boldsymbol{\psi}^{-1}E]_i}, \text{ with } E = (1, 1, ...)^T \tag{5.10}$$

where F is a low-pass filter, and $\boldsymbol{R}$ is the mean residual values for both B-mode and elastography images, in which for pixel i are defined as

$$\boldsymbol{R}_i = \sum_{j \in \mathcal{L}} W_{ij}(\boldsymbol{y}_i - \boldsymbol{\mu}_j)\Sigma^{-1} \tag{5.11}$$

The difference from Chapter 4 is that the mean residual value $\boldsymbol{R}_i$ is a vector of size 1×2, and that the mean inverse covariance matrix $\boldsymbol{\psi}_i^{-1}$ at pixel i is a vector of size 2×2 defined as follows

$$\boldsymbol{\psi}_i^{-1} = \sum_{j \in \mathcal{L}} W_{ij}\Sigma^{-1} \tag{5.12}$$

5.7 Compression ultrasound elastography results

5.7.1 Image dataset

In this study, two datasets provided by different research institutions containing B-mode and elastography information were used to assess the proposed segmentation method.

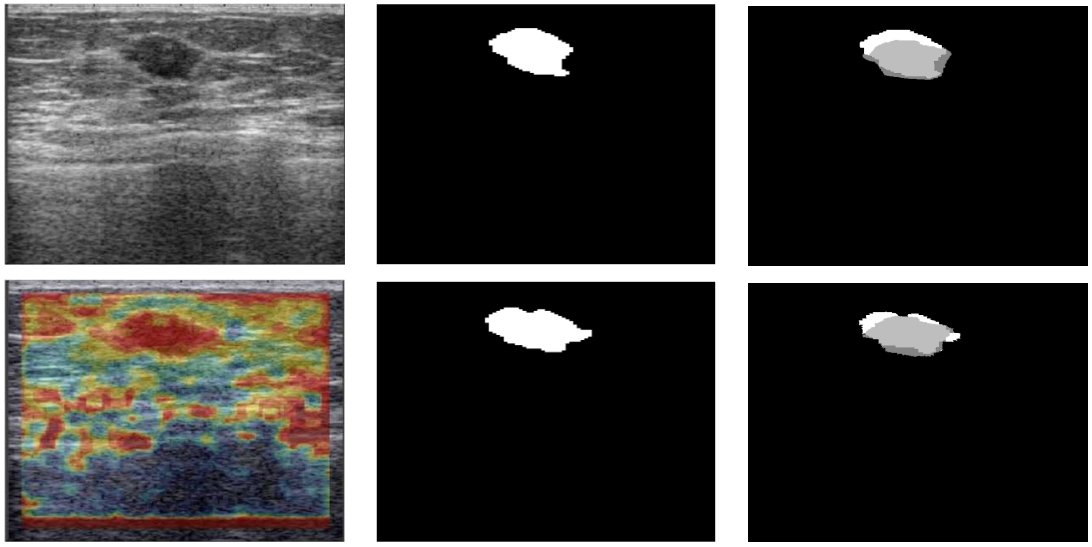
Dataset E1 is composed of 12 images obtained from the Churchill Hospital (Oxford, England) with a Zonare z.one system and an L10-5 linear array transducer (8.5 MHz). The strain information was provided by the BiomedIA research group of the University of Oxford, based on estimating the strain information by tracking the displacement of the RF signal [82]. Dataset E2, composed of 21 images, was provided by the Medical Imaging Group of the Cambridge University Engineering Department. The scans were obtained with a Dynamic Imaging Diasus ultrasound machine with a 5-10 MHz linear array transducer, and the strain information was generated using a tissue displacement tracking algorithm proposed by the same research group [83]. Manual delineations of the tumors were performed by an expert radiologist, and all of the images involved in this work were previously made anonymous to preserve the confidentiality of the patients.

5.7.2 Qualitative results

A qualitative evaluation is presented here to show the behavior of the algorithm dealing with illustrative cases, one in which the lesion is easily segmented using B-mode information alone (example 1) and one where this modality is clearly insufficient (example 2). Firstly, the images are segmented using the previously proposed LLSI algorithm with B-mode information alone. Subsequently, the image is segmented including both B-mode and elastography information. We have compared the different results in order to determine

the benefits of combining both ultrasound B-mode and elastography information in the lesion segmentation problem.

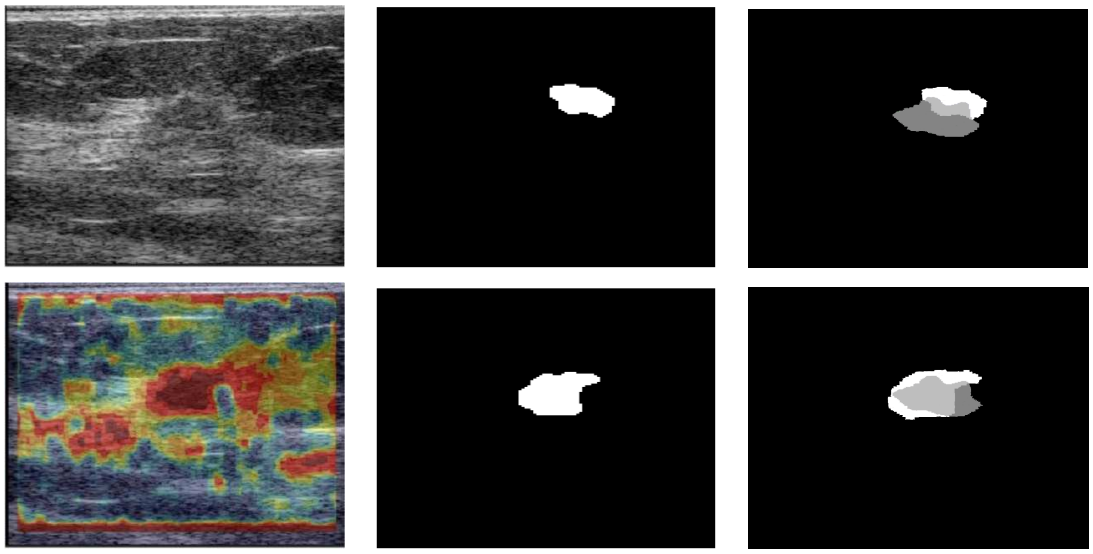
Figure 5.6 shows the segmentation results of the first example, where the tumor is well defined in the B-mode image and the elastography does not provide essential additional information. Note that the results are similar for both performances. Although it is not clearly appreciated in this example, the inclusion of elastography information might yield an over-segmented result due to the fact that tumors invade surrounding tissues making them appear stiffer.



(a) (b) (c)

Figure 5.6: Example 1: Segmentation results using B-mode information alone (first row), and combining B-mode and elastography (second row). Column (a) plots the original image, (b) the segmentation result and (c) the overlap between the result and the GT, where the light gray colour denotes TP pixels, dark gray represents FN and white denotes FP.

On the other hand, Figure 5.7 shows the segmentation results of the second example. In this case, the B-mode does not provide enough information to clearly segment the tumor and the method fails. In this kind of image, elastography provides better information on the location of the tumor, as shown in the second row of Figure 5.7. Using this information, the segmentation result is clearly improved.



(a) (b) (c)

Figure 5.7: Example 2: Segmentation results using B-mode information alone (first row), and combining B-mode and elastography (second row). Column (a) plots the original image, (b) the segmentation result and (c) the overlap between the result and the GT, where the light gray colour denotes TP pixels, dark gray represents FN and white denotes FP.

5.7.3 Quantitative results

In this subsection, quantitative results are presented. Specifically, we compared the results obtained by using B-mode information alone and combining it with elastography information in the segmentation framework. The results are presented in relationship with the dataset used in Table 5.1. Analysis of this table shows that the inclusion of the elastography in the segmentation framework considerably improves the segmentation results in the two different datasets independently and in the total dataset (Dataset E1 and E2). For a better comparison of these results, Figure 5.8 shows a set of box plot charts comparing the DSC values obtained. Analyzing this figure, we can observe a significant improvement in the results using Dataset E1, while no-significant improvements are appreciable in Dataset E2. This difference in results between datasets can be explained by looking at the images of each dataset: while most of the lesions appear well defined in Dataset E2, they are more difficult to locate in Dataset E1. Hence, the impact of including elastography is higher in datasets where B-mode information alone is not sufficient. However, the inclusion of elastography information improves (being significant or not) the performance of the segmentation algorithm in all cases.

Table 5.1: Quantitative results using B-mode alone or including elastography information for USE.

	Information	Sensitivity	Specificity	PPV	AO	DSC
Dataset E1	B-mode	0.427	0.907	0.722	0.370	0.501
	B-mode + elasto	0.661	0.990	0.839	0.583	0.728
Dataset E2	B-mode	0.732	0.942	0.865	0.6752	0.780
	B-mode + elasto	0.750	0.989	0.861	0.673	0.789
Total Dataset	B-mode	0.621	0.292	0.813	0.564	0.679
	B-mode + elasto	0.717	0.990	0.853	0.641	0.767

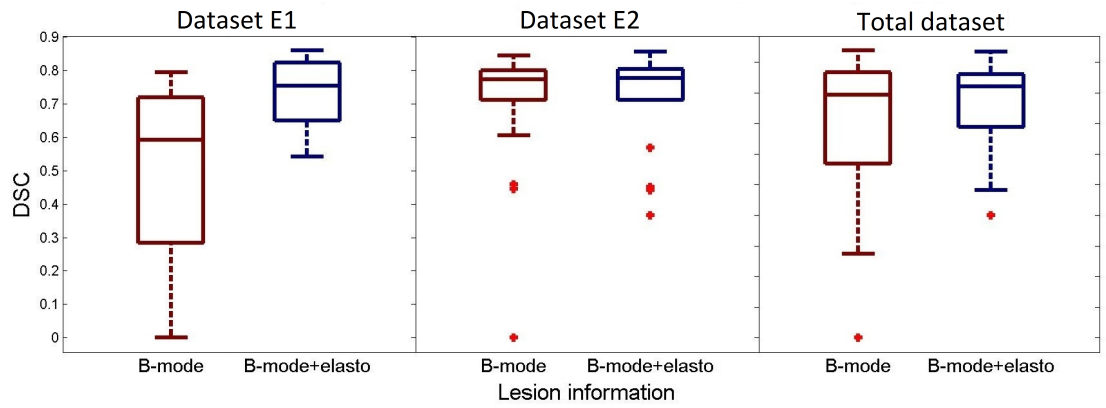


Figure 5.8: Box plot charts comparing the DSC values for Dataset E1, E2 and the total dataset.

5.8 Shear-wave elastography results

5.8.1 Image dataset

The dataset of SWE images was collected recently (2013) from the UDIAT Diagnostic Centre of the Parc Taulí Corporation, Sabadell (Spain) with an Aixplorer V4 US system (SuperSonic Imagine, Aix-en-Provence, France), which was equipped with a 4-15 MHz linear array transducer. The dataset consists of 24 images from different patients. Figure 5.4 shows an example of a SWE image from this dataset. The B-mode and the elasticity map are depicted on the same screen for a better visualization. The values of elasticity are embedded in the DICOM file when acquiring the image, and can be extracted using a software tool provided by SuperSonic Imagine.

5.8.2 Qualitative results

A qualitative evaluation is presented here to show the behavior of the algorithm dealing with two representative cases. The first case, Figure 5.9(a-e), shows a well defined lesion in the B-mode. The second case, Figure 5.9(f-j), shows a lesion which appears in the B-mode with similar intensity values to normal tissues. In this case, the inclusion of elastography gives essential information for the lesion identification.

Shear-wave elastography provides reliable information on the location of the lesion. However, when performing this study, we found the presence of an artifact in most of the SWE cases, caused by the peritumoral stiffness [43, 44, 132]. As depicted in Figure 5.10, in some cases with cancerous lesions, the stiffest tissue is found in the peritumoral region rather than in the cancer itself. In addition, in some cases the elasticity information within the lesion is not even measured. The presence of the peritumoral stiffness is helpful for the radiologist when detecting malignant lesions. However, it induces our segmentation method to fail. Thus, no quantitative results using SWE can be reported as most of the cases show peritumoral stiffness. This will need further investigation to assess the nature of this artifact.

5.9 Conclusions

In this chapter, a novel unified framework for simultaneously estimating the bias field and segmenting lesions in breast ultrasound using both B-mode and elastography information was proposed.

First, the inclusion of compression elastography data was evaluated. Qualitative results were performed using two illustrative cases, one where the B-mode shows a well defined lesion and another where elastography provides more meaningful information. The segmentation results were compared by using the B-mode information alone or combining the B-mode and elastography. These qualitative results show that combining both the B-mode and elastography information in a unique framework improves the segmentation results, especially when B-mode images are not conclusive, which is the often the case in non-cystic lesions. These results are corroborated later in the quantitative results section. The proposal has been assessed in two datasets (E1: 12 images, E2: 21 images) obtaining better results when including elastography information in both datasets.

Finally, we evaluated the inclusion of SWE in the segmentation framework. The presence

of the peritumoral stiffness in the majority of the cases in the dataset induced our approach to mis-segment the images, and made its evaluation impossible with this kind of image. Since the presence of peritumoral stiffness is helpful in the detection of cancerous lesions, we consider that it would be interesting to include SWE information in a novel lesion detection method.

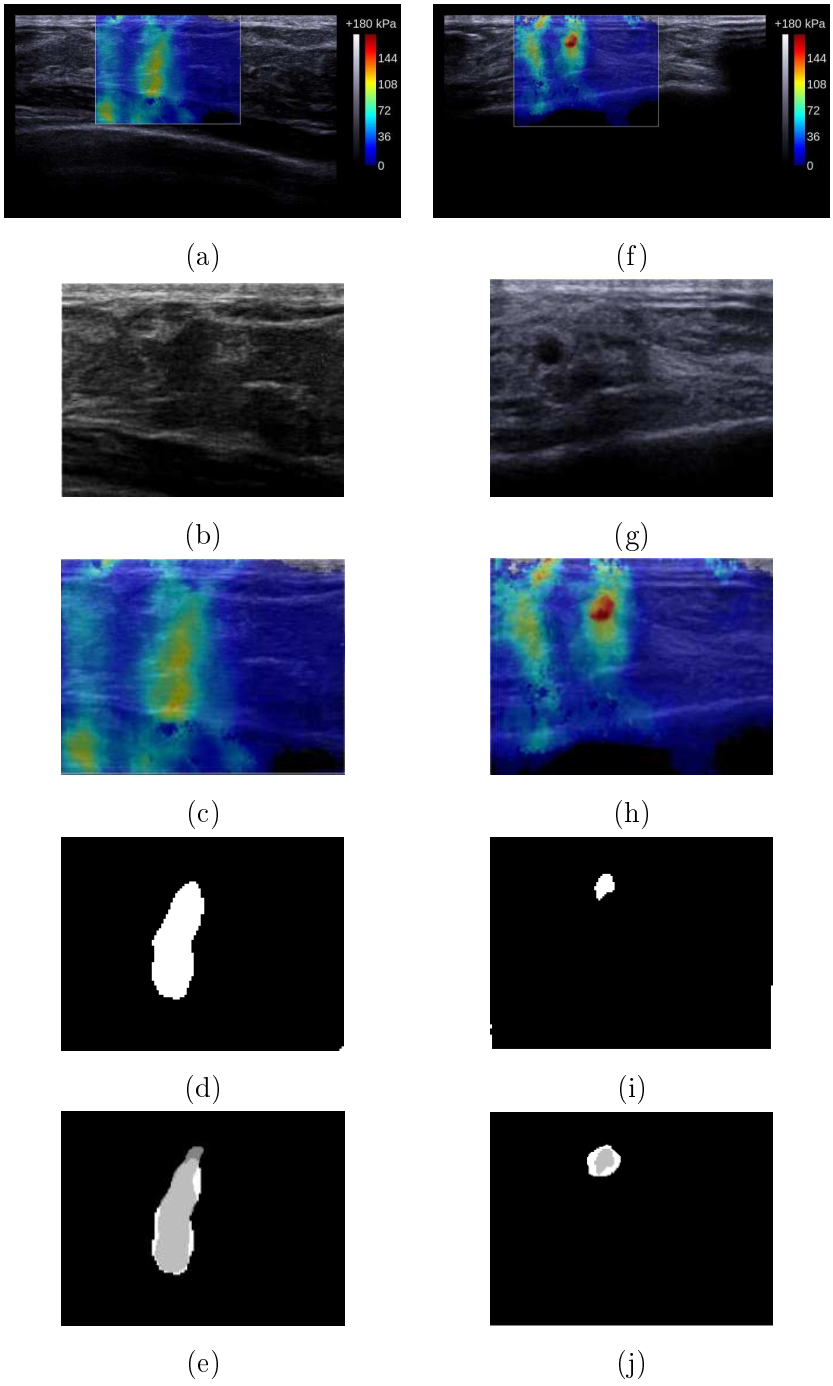
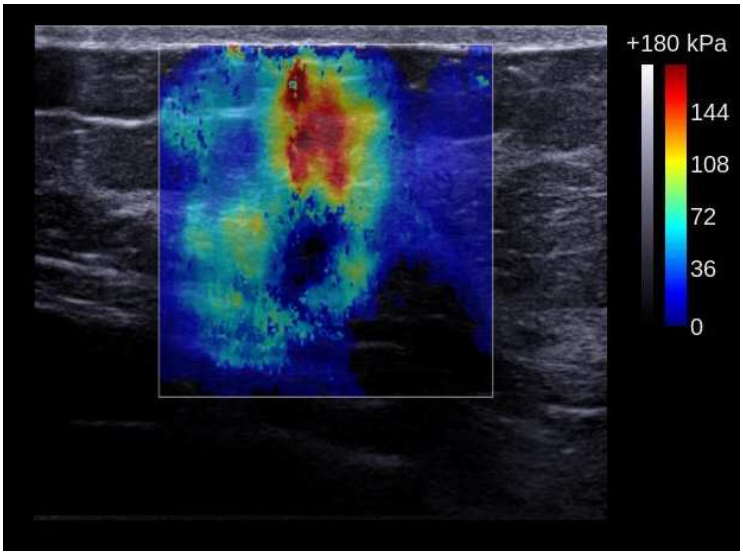


Figure 5.9: Qualitative results for two SWE breast images. (a-e) show the results for the first case and (f-j) for the second case. (a and f) show the original image with the elasticity map, measured within a ROI placed by the radiologist, superimposed. (b and g) show the B-mode image corresponding to the ROI. (c,h) show the elastography map of the ROI. (d and i) show the segmentation results. (e and j) show the overlap between the segmentation and the GT, where the light gray color denotes TP pixels, dark gray represents FN and white denotes FP.



(a)

(b)

Figure 5.10: Examples of the SWE peritumoral artifact. (a) Malignant case where the peritumoral tissues appear as the stiffer regions. (b) Elasticity values of the lesion are not measured.

6

Conclusions and future work

The current and final chapter summarizes the methodology and results of the studies on the different CAD stages presented in this thesis. It further describes the original contributions made by this thesis to the research area of computer-aided breast cancer detection. It also highlights possible future improvements and directions of research related to the work presented.

6.1 Conclusions of the thesis

The aim of this thesis has been the proposal of a new pipeline capable of detecting, segmenting, and classifying breast lesions in ultrasound imaging. Starting with an initial study of the state-of-the-art of breast lesion detection, it was concluded that machine learning-based methods adapt better to the specific characteristics of the target dataset, and do not need the use of an extra false positive reduction step, necessary in most of the detection methods based on image processing.

A further analysis of the breast lesion segmentation literature was also provided in order to highlight the strengths and weaknesses of each segmentation technique analyzed.

Analysis of these works shows that machine learning-based methods obtained good results and do not require any kind of user interaction. Through this analysis, it was assessed that few approaches were based on MRF, which, in principle, can be regarded as an interesting framework incorporating both spatial and intensity information.

Subsequently, a study of the state-of-the-art of breast lesion classification was also carried out in order to highlight the most frequent chosen important factors, such as features, classifier, train/test procedure and dataset composition, for the classification results. Regarding the features used to classify the lesions, most are extracted from manual delineations or from a ROI placement, and the majority of the methods studied used a combination of texture and morphological features. It has also been demonstrated that most of the methods used the k-fold cross validation in the training/testing process as well as unbalanced datasets, which contained fewer malignant than benign images.

Therefore, a computerized lesion detection system for breast US images using Deformable Part Models (DPM) has been proposed. Different configurations of parameters were extensively evaluated to improve the results of the DPM in breast US images. A comparison of several of the most important methods in the literature was performed, evaluating all the methods with the same datasets. Amongst the different methodologies compared, the proposed method obtained the best results. Furthermore, the detection system was trained to detect malignant lesions directly, obtaining similar results to the best proposals in the literature, but considerably simplifying the traditional cancer detection pipeline.

After the detection proposal, different initialization processes to include spatial information in the MRF-MAP segmentation framework to reduce the interaction of the user with the method were introduced. It was concluded that the LLSI initialization proposal obtained the best results. Moreover, the segmentation results were studied in relationship with the lesion type, concluding that cystic lesions obtained the best results due to their well defined boundaries and the homogeneous appearance, while fibroadenoma and cancerous lesions performed poorly due to inherent artifacts related to the lesion type, such as shadowing or heterogeneities within the lesion.

In order to improve the segmentation results, a novel algorithm including elastography information has been proposed. The MRF-MAP segmentation framework was redesigned to include both B-mode and elastography ultrasound data by modeling both types of information as bivariate Gaussian PDFs. The proposed method was evaluated with compression elastography, outperforming the results obtained when using B-mode information alone. It was also evaluated with shear-wave elastography images, but the presence of peritumoral

stiffness in most of the images induced the method to fail, and only qualitative results were reported.

6.1.1 Contributions

The goal of this thesis is to aid radiologists in the day-to-day practise by assisting them in the challenging task of cancer detection in breast sonography by providing them with a second opinion and help to increase the diagnosis accuracy, reducing the biopsy rate and saving time and effort. This second opinion should ideally be provided automatically so as to be effective in a real clinical environment. From this point of view, the main contributions of this thesis to both the scientific and clinical community are:

- A novel lesion detection algorithm in breast ultrasound images using Deformable Part Models (DPM) that outperformed the current methods in the literature. This development has been achieved after elaborating an extensive survey of the most relevant breast ultrasound lesion detection algorithms, an exhaustive analysis of the parameters involved in the method to determine the best configuration for sonography, an exhaustive test using two different datasets, and an experimental comparison between our lesion detection proposal and relevant methods in the state-of-the-art.

- A cancer detection technique in breast ultrasound images using DPM. The model is trained to detect malignant lesions directly, simplifying the traditional cancer detection pipeline, which first detect candidate lesions, segments them for extracting features, and finally, classifies them into benign or malignant.

- A proposal of different initialization improvements of the MRF-MAP method that reduces the interaction process involved in the original proposal to one-click. This development has been achieved after elaborating an extensive survey of the most relevant breast ultrasound lesion detection algorithms, an exhaustive test using two different datasets, and an experimental comparison between our lesion segmentation proposal and relevant methods in the state-of-the-art.

- A study of the influence of lesion type in the MRF-MAP segmentation results. To our knowledge, this type of study has not been performed before, and indicates which types of lesions obtain better and worst segmentation results.

- A novel lesion segmentation algorithm in breast ultrasound images including elastography information in a unified framework, assessed including conventional com-

pressing elastography and the recent shear-wave elastography. This development has been achieved after elaborating an extensive survey of the most relevant works which investigated the use of computerized techniques with elastography imaging.

6.2 Future work

The analysis of breast US images is a complex topic involving several aspects and multiple research lines. This notion is exemplified in the research of CAD systems, where different processes play an important role. The research efforts can focus on a particular stage of the CAD pipeline, or on a combination of them. Furthermore, other interesting topics arise from the needs of current clinical practise, with the introduction of new screening techniques, such as elastography or ABUS.

Hence, future directions are presented divided into two categories: those related to increasing the reliability of our proposal, and future research lines departing from this thesis.

6.2.1 Short-term proposal improvements

After the analysis of the proposed methodologies, some potential improvements could be made in the near future. In regard to malignant lesion detection, a more exhaustive parameter analysis could be performed to improve the detection results. The sensitivity of this proposal is in line with the methods of the state-of-the-art, but with a higher number of FP per image, as commented in Chapter 3. One proposal could be a two-stage detection method, where first, lesions are detected, and then, cancers are detected amongst the previously detected lesions. This proposal will increase the malignant lesion pipeline, but it will continue having fewer stages than the traditional pipeline.

Regarding the segmentation process, more efforts need to be made when dealing with images with specific artifacts that induce the method to fail. As was shown in Chapter 4, the method obtained results in line with the other state-of-the-art methods when excluding these conflictive images. To solve the influence of such artifacts, we propose using adaptable size windows in the lesion and background information acquisition and the inclusion of a pre-processing step to enhance the lesions. Adaptable windows could manage zoomed-in lesions, and the pre-processing step could enrich the visualization of lesions with similar intensity to normal tissue. Another limitation of the method that would be interesting to solve is the fact that only one lesion per image can be segmented.

A new approach for lesion segmentation including elastography has been presented in this thesis. Promising results were obtained, but an additional further evaluation in a larger dataset should be performed. We have also shown the potential of SWE imaging. However, further investigation of the acquisition of these images should be performed to solve the presence of artifacts.

6.2.2 Future research lines

In the long term, there are several new research lines departing from this thesis that could be studied. One that was initially explored in this thesis was the inclusion of elastography information in the segmentation framework. The benefits of the introduction of elastography in current screening programs have been proved in different studies. Hence, we consider that elastography can play an important role in computerized lesion detection and classification.

Regarding the MRF-MAP segmentation proposal, we consider that it would be necessary to include spatial information in the methodology. Thus, the method could take advantage of the lesion location provided by the user (one click) or by the lesion detection method.

Analyzing the results obtained in lesion detection by the DPM, we also consider that this methodology has the potential of being adapted to detect lesions in 3D volumes, such as ABUS, which is currently being adopted in clinical practice and becoming a relevant topic of interest in medical imaging research. Because of the large number of slices or planes in a 3D breast volume, one of the main problems in 3D lesion detection is the number of FP. The DPM approach obtained a low ratio of FPs per image, which could possibly be improved by including a new FP reduction post-processing technique to analyze the propagation of the detections in the volume.

A

Summary of datasets

In this thesis, different datasets have been used for evaluating the proposed methodologies. Here, a summary of the datasets used is given.

Name: Dataset D1.
Provided by: Manchester Metropolitan Univeristy, Manchester (UK).
Images: Breast sonography.
Machine: B&K Medical Panther 2002 and B&K Medical Hawk 2102.
Transducer: 8-12 MHz linear array transducer.
Year: 2001.
Cases: 406.
Benign cases: 246.
Malignant cases: 60.
Other cases: 100 healthy images.
Avg. image size: 377x396 pixels.
Ground truth: Posterior biopsy/pathological examination, and boundary delineation of a radiologist.
Used for: Lesion detection.

Name: Dataset D2.
Provided by: UDIAT Diagnostic Centre of the Parc Taulí Corporation, Sabadell (Spain).
Images: Breast sonography.
Machine: Siemens ACUSON Sequoia C512.
Transducer: 8.5 MHz 17L5 HD linear array.
Year: 2012/13.
Cases: 326.
Benign cases: 110.
Malignant cases: 53.
Other cases: 163 healthy images.
Avg. image size: 760x570 pixels.
Ground truth: Posterior biopsy/pathological examination, and boundary delineation of a radiologist.
Used for: Lesion detection.

Name: Dataset S1.
Provided by: UDIAT Diagnostic Centre of the Parc Taulí Corporation, Sabadell (Spain).
Images: Breast sonography.
Machine: Siemens ACUSON Sequoia C512.
Transducer: 8.5 MHz 17L5 HD linear array.
Year: 2012.
Cases: 140.
Benign cases: 96.
Malignant cases: 44.
Other cases: N/A.
Avg. image size: 760x570 pixels.
Ground truth: Posterior biopsy/pathological examination, and boundary delineation of a radiologist.
Used for: Lesion segmentation.

Name: Dataset S2.
Provided by: Churchill hospital, Oxford (UK).
Images: Breast sonography.

Machine: Zonare z.one.
Transducer: 8.5 MHz L10-5 linear array.
Year: 2009.
Cases: 72.
Benign cases: 18.
Malignant cases: 54.
Other cases: N/A.
Avg. image size: 226x1017 pixels.
Ground truth: Posterior biopsy/pathological examination, and boundary delineation of a radiologist.
Used for: Lesion segmentation.

Name: Dataset E1.
Provided by: Churchill hospital, Oxford (UK).
Images: Breast sonography and elastography.
Machine: Zonare z.one.
Transducer: 8.5 MHz L10-5 linear array.
Year: 2010.
Cases: 12.
Benign cases: N/A.
Malignant cases: N/A.
Other cases: N/A.
Avg. image size: 226x1017 pixels.
Ground truth: Boundary delineation of a radiologist.
Used for: Lesion segmentation with elastography.

Name: Dataset E2.
Provided by: Medical Imaging Group of the Cambridge University, Cambridge (UK).
Images: Breast sonography and elastography.
Machine: Dynamic Imaging Diasus.
Transducer: 5-10 MHz linear array.
Year: 2008.
Cases: 21.

Benign cases: 17.

Malignant cases: 4.

Other cases: N/A.

Avg. image size: 384x294 pixels.

Ground truth: Posterior biopsy/pathological examination, and boundary delineation of a radiologist.

Used for: Lesion segmentation with elastography.

Name: Dataset E3.

Provided by: UDIAT Diagnostic Centre of the Parc Taulí Corporation, Sabadell (Spain).

Images: Breast sonography and shear-wave elastography.

Machine: SuperSonic Aixplorer V4.

Transducer: 4-15 MHz linear array.

Year: 2013.

Cases: 24.

Benign cases: N/A.

Malignant cases: N/A.

Other cases: N/A.

Avg. image size: 540x317 pixels.

Ground truth: Boundary delineation of a radiologist.

Used for: Lesion segmentation with elastography.

Bibliography

[1] Càncer. Generalitat de Catalunya, September 2013. URL: `www20.gencat.cat/portal/site/cancer/`.

[2] K.Z. Abd-Elmoniem, A.B.M. Youssef, and Y.M. Kadah. Real-time speckle reduction and coherence enhancement in ultrasound imaging via nonlinear anisotropic diffusion. *IEEE Transactions on Biomedical Engineering*, 49(9):997–1014, 2002.

[3] A. Achim, A. Bezerianos, and P. Tsakalides. Novel bayesian multiscale method for speckle removal in medical ultrasound images. *IEEE Transactions on Medical Imaging*, 20(8):772–783, 2001.

[4] M. Alemán-Flores, L. Álvarez, and V. Caselles. Texture-oriented anisotropic filtering and geodesic active contours in breast tumor ultrasound segmentation. *J. Math. Imaging Vis.*, 28(1):81–97, 2007.

[5] S.M. Alhabshi, K. Rahmat, N. Abdul Halim, S. Aziz, S. Radhika, G.C. Gan, A. Vijayananthan, C.J. Westerhout, M.N. Mohd-Shah, S. Jaszle, N. Harlina Mohd Latar, and R. Muhammad. Semi-quantitative and qualitative assessment of breast ultrasound elastography in differentiating between malignant and benign lesions. *Ultrasound in Medicine and Biology*, 39(4):568–578, 2013.

[6] A.V. Alvarenga, A.F.C. Infantosi, W.C.A. Pereira, and C.M. Azevedo. Assessing the combined performance of texture and morphological parameters in distinguishing breast tumors in ultrasound images. *Medical Physics*, 39(12):7350–7358, 2012.

[7] American Cancer Research. Breast cancer facts and figures 2011-2012. *Atlanta, GA: American Cancer Society*, 2011.

[8] B.O. Anderson, R. Shyyan, A. Eniu, R.A. Smith, C.H. Yip, N.S. Bese, L.W.C. Chow, S. Masood, S.D. Ramsey, and R.W. Carlson. Breast cancer in limited-resource

countries: an overview of the breast health global initiative 2005 guidelines. *Breast Journal*, 12(1):S3–S15, 2006.

[9] M.P. André, M. Galperin, L.K. Olson, K. Richman, S. Payrovi, and P. Phan. Improving the accuracy of diagnostic breast ultrasound. *Acoustical Imaging*, 26:453–460, 2002.

[10] J. Austoker. Screening and self examination for breast cancer. *BMJ*, 309(6948):168–174, 1994.

[11] F.M. Awad. Role of supersonic shear wave imaging quantitative elastography (SSI) in differentiating benign and malignant solid breast masses. *The Egyptian Journal of Radiology and Nuclear Medicine*, 44(3):681–685, 2013.

[12] R.G. Barr. Real-time ultrasound elasticity of the breast: initial clinical results. *Ultrasound Quarterly*, 26(2):61–6, 2010.

[13] J. Bartko. Measurement and reliability: statistical thinking considerations. *Schizophr. Bull.*, 17(3):483–489, 1991.

[14] J. Bercoff, M. Tanter, and M. Fink. Supersonic shear imaging: a new technique for soft tissue elasticity mapping. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 51(4):396–409, 2004.

[15] W.A. Berg, D.O. Cosgrove, C.J. Dor, F. Schfer, W. Svensson, R. Hooley, R. Ohlinger, E.B. Mendelson, C. Balu-Maestro, M. Locatelli, C. Tourasse, B. Cavanaugh, V. Juhan, A.T. Stavros, A. Tardivon, J. Gay, J.P. Henry, and C. Cohen-Bacrie. Shearwave elastography improves the specificity of breast ultrasound: the BE1 multinational study of 939 masses. *Radiology*, 262:435–449, 2012.

[16] W.A. Berg, L. Gutierrez, M.S. NessAiver, W. B. Carter, M. Bhargavan, R.S. Lewis, and O.B. Ioffe. Diagnostic accuracy of mammography, clinical examination, US, and MR imaging in preoperative assessment of breast cancer. *Radiology*, 233(3):830–849, 2004.

[17] J. Besag. Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society*, 36(2):192–326, 1974.

[18] J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society*, B-48:259–302, 1986.

[19] 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.

[20] D. Boukerroui, A. Baskurt, J.A. Noble, and O. Basset. Segmentation of ultrasound images: multiresolution 2D and 3D algorithm based on global and local statistics. *Pattern Recognition*, 24(4-5):779–790, 2003.

[21] P.J. Burt and E.H. Adelson. A multiresolution spline with application to image mosaics. *ACM Trans Graphics*, 2:217–236, 1983.

[22] D. Chakraborty, H.J. Yoon, and C.M. Thoms. Spatial localization accuracy of radiologists in free-response studies: inferring perceptual FROC curves from mark-rating data. *Academic Radiology*, 14(1):4–18, 2007.

[23] R.F. Chang, D.R. Chen, W.K. Moon, and W.R. Lai. 3-D ultrasound strain images for breast cancer diagnosis. *International Congress Series*, 1281(0):1069–1074, 2005.

[24] C.M. Chen, Y. H. Chou, K. C. Han, G. S. Hung, C. M. Tiu, H. J. Chiou, and S. Y. Chiou. Breast lesions on sonograms: computer-aided diagnosis with nearly setting-independent features and artificial neural networks. *Radiology*, 226:504–514, 2003.

[25] D.R. Chen, R.F. Chang, W.J. Kuo, M.C. Chen, and Y.L. Huang. Diagnosis of breast tumors with sonographic texture analysis using wavelet transform and neural networks. *Ultrasound in Medicine and Biology*, 28(10):1301–1310, 2002.

[26] H.D. Cheng, J. Shan, W. Ju, Y. Guo, and L. Zhang. Automated breast cancer detection and classification using ultrasound images: a survey. *Pattern Recognition*, 43(1):299–317, 2010.

[27] H.D. Cheng, X.J. Shi, R. Min, L.M. Hu, X.P. Cai, and H.N. Du. Approaches for automated detection and classification of masses in mammograms. *Pattern Recognition*, 39(4):646–668, 2006.

[28] J.Z. Cheng, Y.H. Chou, C.S. Huang, Y.C. Chang, C.M. Tiu, F.C. Yeh, K.W. Chen, C.H. Tsou, and C.M. Chen. ACCOMP: augmented cell competition algorithm for breast lesion demarcation in sonography. *Medical Physics*, 37(12):6240–6252, 2010.

[29] H.H. Chiang, J.Z. Cheng, P.K. Hung, C.Y. Liu, C.H. Chung, and C.M. Chen. Cell-based graph cut for segmentation of 2D/3D sonographic breast images. *IEEE Inter-*

national Symposium on Biomedical Imaging: From Nano to Macro, pages 177–180, 2010.

[30] J. Cui, B. Sahiner, H.P. Chan, A. Nees, C. Paramagul, L.M. Hadjiiski, C. Zhou, and J. Shi. A new automated method for the segmentation and characterization of breast masses on ultrasound images. *Medical Physics*, 36(5):1553–1565, 2009.

[31] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1:886–893, 2005.

[32] P. L. Davis and K. S. McCarty Jr. Sensitivity of enhanced MRI for the detection of breast cancer: new, multicentric, residual, and recurrent. *European Radiology*, 7(5):S289–S298, 1997.

[33] L. R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

[34] K. Doi. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized Medical Imaging and Graphics*, 31(4):198–211, 2007.

[35] K. Drukker, M. L. Giger, K. Horsch, M. A. Kupinski, C. J. Vyborny, and E. B. Mendelson. Computerized lesion detection on breast ultrasound. *Medical Physics*, 29(7):1438–1446, 2002.

[36] K. Drukker, M. L. Giger, and E. B. Mendelson. Computerized analysis of shadowing on breast ultrasound for improved lesion detection. *Medical Physics*, 30(7):1833–1842, 2003.

[37] K. Drukker, M. L. Giger, and C. E. Metz. Robustness of computerized lesion detection and classification scheme across different breast US platforms. *Radiology*, 237(3):834–840, 2005.

[38] K. Drukker, M. L. Giger, C. J. Vyborny, and E. B. Mendelson. Computerized detection and classification of cancer on breast ultrasound. *Academic Radiology*, 11(5):526–535, 2004.

[39] H. El-Wakeel and H.C. Umpleby. Systematic review of fibroadenoma as a risk factor for breast cancer. *Breast*, 12(5):302–7, 2003.

[40] D. Ensminger and F.B. Stulen. *Ultrasonics: data, equations, and their practical uses.* Taylor & Francis Group, 2009.

[41] Eurostat. Health statistics at regional level - statistics explained (2013/6/4). http://epp.eurostat.ec.europa.eu/statistics_explained/index.php.

[42] A. Evans, P. Whelehan, K. Thomson, K. Brauer, L. Jordan, C. Purdie, D. McLean, L. Baker, and A. Thompson. Differentiating benign from malignant solid breast masses: value of shear wave elastography according to lesion stiffness combined with greyscale ultrasound according to BI-RADS classification. *British Journal of Cancer*, 107:224–229, 2012.

[43] A. Evans, P. Whelehan, K. Thomson, D. McLean, K. Brauer, C. Purdie, L. Baker, L. Jordan, P. Rauchhaus, and A. Thompson. Invasive breast cancer: relationship between shear-wave elastographic findings and histologic prognostic factors. *Radiology*, 263(3):673–677, 2012.

[44] A. Evans, P. Whelehan, K. Thomson, D. McLean, K. Brauer, C. Purdie, L. Jordan, L. Baker, and A. Thompson. Quantitative shear wave ultrasound elastography: initial experience in solid breast masses. *Breast Cancer Research*, 12(6):R104, 2010.

[45] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2010 (VOC2010) Results. http://www.pascal-network.org/challenges/VOC/voc2010/workshop/index.html, 2010.

[46] K. Evers. Diagnostic breast imaging mammography, sonography, magnetic resonance imaging, and interventional procedures. *American Journal of Roentgenology*, 177(5):1094–1094, 2001.

[47] T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.

[48] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1627–1645, 2010.

[49] E.F. Fleury, J.C. Fleury, S. Piato, and Roveda D. New elastographic classification of breast lesions during and after compression. *Diagnostic and Interventional Radiology*, 15(2):96–103, 2009.

[50] The International Agency for Research on Cancer. *World Health Organization: tumours of the breast and female genital organs.* IARCPress-WHO, 1 edition, 2003.

[51] 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.

[52] Y. Freund and R.E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

[53] D. Gan and Y.T. Soon. A multifractal approach for auto-segmentation of SAR images. *IEEE International Geoscience and Remote Sensing Symposium.*, 5:2301–2303, 2001.

[54] L. Gao, X. Liu, and W. Chen. Phase- and GVF-based level set segmentation of ultrasonic breast tumors. *Journal of Applied Mathematics*, 2012:22, 2012.

[55] L. Gao, W. Yang, Z. Liao, X. Liu, Q. Feng, and W. Chen. Segmentation of ultrasonic breast tumors based on homogeneous patch. *Medical Physics*, 39(6):3299–3318, 2012.

[56] B.S. Garra, E.I. Cespedes, J. Ophir, S.R. Spratt, R.A. Zuurbier, C.M. Magnant, and M.F. Pennanen. Elastography of breast lesions: initial clinical results. *Radiology*, 202(1):79–86, 1997.

[57] R. B. Girshick, P. F. Felzenszwalb, and D. McAllester. Discriminatively trained deformable part models, release 5. http://people.cs.uchicago.edu/ rbg/latent-release5/.

[58] A. Goddi, M. Bonardi, and S. Alessi. Breast elastography: a literature review. *Journal of Ultrasound*, 15(3):192–198, 2012.

[59] W. Gómez, L. Leija, A. V. Alvarenga, A. F. C. Infantosi, and W. C. A. Pereira. Computerized lesion segmentation of breast ultrasound based on marker-controlled watershed transformation. *Medical Physics*, 37(1):82–95, 2010.

[60] X. Gong, Q. Xu, Z. Xu, P. Xiong, W. Yan, and Y. Chen. Real-time elastography for the differentiation of benign and malignant breast lesions: a meta-analysis. *Breast Cancer Research and Treatment*, 130(1):11–18, 2011.

[61] L. Grady. Random walks for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(11):1768–1783, 2006.

[62] M. Guray and A.A. Sahin. Benign breast diseases: classification, diagnosis, and management. *The Oncologist*, 11(5):435–449, 2006.

[63] H.M. Gweon, J.H. Youk, E.J. Son, and J.A. Kim. Clinical application of qualitative assessment for breast masses in shear-wave elastography. *European Journal of Radiology*, 82(11):e680–e685, 2013.

[64] Z. Hao, Q. Wang, Y. Seong, J.H. Lee, H. Ren, and J.Y. Kim. Combining CRF and multi-hypothesis detection for accurate lesion segmentation in breast sonograms. *Medical Image Computing and Computer-Assisted Intervention MICCAI*, 7510:504–511, 2012.

[65] K. Horsch, M. L. Giger, L. A. Venta, and C. J. Vyborny. Computerized diagnosis of breast lesions on ultrasound. *Medical Physics*, 29(2):157–164, 2002.

[66] Karla Horsch, Maryellen L. Giger, Luz A. Venta, and Carl J. Vyborny. Automatic segmentation of breast lesions on ultrasound. *Medical Physics*, 28(8):1652–1659, 2001.

[67] Q.H. Huang, S.Y. Lee, L.Z. Liu, M.H. Lu, L.W. Jin, and A.H. Li. A robust graph-based segmentation method for breast tumors in ultrasound images. *Ultrasonics*, 52(2):266–275, 2012.

[68] Y.L. Huang and D.R. Chen. Watershed segmentation for breast tumor in 2-D sonography. *Ultrasound in Medicine and Biology*, 30(5):625–632, 2004.

[69] Y.L. Huang and D.R. Chen. Automatic contouring for breast tumors in 2-D sonography. *27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3225–3228, 2005.

[70] Y.L. Huang, D.R. Chen, and Y.K. Liu. Breast cancer diagnosis using image retrieval for different ultrasonic systems. *International Conference on Image Processing (ICIP)*, 5:2957–2960, 2004.

[71] Y.L. Huang, Y.R. Jiang, D.R. Chen, and W.K. Moon. Level set contouring for breast tumor in sonography. *Journal of Digital Imaging*, 20(3):238–247, 2007.

[72] Y.L. Huang, K.L. Wang, and D.R. Chen. Diagnosis of breast tumors with ultrasonic texture analysis using support vector machines. *Neural Comput. Appl.*, 15(2):164–169, 2006.

[73] A. Itoh, E. Ueno, E. Tohno, H. Kamma, H. Takahashi, T. Shiina, M. Yamakawa, and T. Matsumura. Breast disease: clinical application of US elastography for diagnosis. *Radiology*, 239(2):341–350, 2006.

[74] P. Jaccard. The distribution of flora in the alpine zone. *New Phytol.*, 11(2):37–50, 1912.

[75] P. Jiang, J. Peng, G. Zhang, E. Cheng, V. Megalooikonomou, and H. Ling. Learning-based automatic breast tumor detection and segmentation in ultrasound images. *IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1587–1590, 2012.

[76] P.C. Johns and M.J. Yaffe. X-ray characterisation of normal and neoplastic breast tissues. *Physics in Medicine and Biology*, 32(6):675, 1987.

[77] S. Joo, Y.S. Yang, W.K. Moon, and H.C. Kim. Computer-aided diagnosis of solid breast nodules: use of an artificial neural network based on multiple sonographic features. *IEEE Transactions on Medical Imaging*, 23(10):1292–1300, 2004.

[78] H. Kim, J.H. Youk, H.M. Gweon, J.A. Kim, and E.J. Son. Diagnostic performance of qualitative shear-wave elastography according to different color map opacities for breast masses. *European Journal of Radiology*, 82(8):e326–e331, 2013.

[79] M.A. Kupinski and M.L. Giger. Automated seeded lesion segmentation on digital mammograms. *IEEE Transactions on Medical Imaging*, 17(4):510–517, 1998.

[80] M. A. Kutay, A. P. Petropulu, and C. W. Piccoli. Breast tissue characterization based on modeling of ultrasonic echoes using the power-law shot noise model. *Pattern Recognition Letters*, 24(4-5):741–756, 2003.

[81] L.C.H. Leong, L.S.J. Sim, Y.S. Lee, F.C. Ng, C.M. Wan, S.M.C. Fook-Chong, A.R. Jara-Lazaro, and P.H. Tan. A prospective study to compare the diagnostic performance of breast elastography versus conventional breast ultrasound. *Clinical Radiology*, 65(11):887–894, 2010.

[82] J. Li, Y. Cui, M. Kadour, and J.A. Noble. Elasticity reconstruction from displacement and confidence measures of a multi-compressed ultrasound RF sequence. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 55(2):319–326, 2008.

[83] J.E. Lindop, G.M. Treece, A.H. Gee, and R.W. Prager. Estimation of displacement location for enhanced strain imaging. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 54(9):1751–1771, 2007.

[84] B. Liu, H.D. Cheng, J. Huang, J. Tian, J. Liu, and X. Tang. Automated segmentation of ultrasonic breast lesions using statistical texture classification and active contour based on probability distance. *Ultrasound in Medicine and Biology*, 35(8):1309–1324, 2009.

[85] B. Liu, H.D. Cheng, J. Huang, J. Tian, X. Tang, and J. Liu. Fully automatic and segmentation-robust classification of breast tumors based on local texture analysis of ultrasound images. *Pattern Recognition*, 43(1):280–298, 2010.

[86] B. Liu, H.D. Cheng, J. Huang, J. Tian, X. Tang, and J. Liu. Probability density difference-based active contour for ultrasound image segmentation. *Pattern Recognition*, 43(6):2028–2042, 2010.

[87] W. Liu, J.A. Zagzebski, . Varghese, C.R. Dyer, U. Techavipoo, and T.J. Hall. Segmentation of elastographic images using a coarse-to-fine active contour model. *Ultrasound in Medicine and Biology*, 32(3):397–408, 2006.

[88] A. Madabhushi and D.N. Metaxas. Combining low-, high-level and empirical domain knowledge for automated segmentation of ultrasonic breast lesions. *IEEE Transactions on Medical Imaging*, 22(2):155–169, 2003.

[89] S. S. Mahdavi, M. Moradi, W. J. Morris, S. L. Goldenberg, and S. E. Salcudean. Fusion of ultrasound B-mode and vibro-elastography images for automatic 3-D segmentation of the prostate. *IEEE Transactions on Medical Imaging*, 31(11):2073–2082, 2012.

[90] M. Martinez-Alonso, E. Vilaprinyo, R. Marcos-Gragera, and M. Rue. Breast cancer incidence and overdiagnosis in Catalonia (Spain). *Breast Cancer Research*, 12(4):R58, 2010.

[91] F.J. Massey. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253):68–78, 1951.

[92] J. Massich, F. Meriaudeau, E. Pérez, R. Martí, A. Oliver, and Joan Martí. Lesion segmentation in breast sonography. In *Digital Mammography / IWDM*, pages 39–45, 2010.

[93] J. Massich, F. Meriaudeau, M. Sentís, S. Ganau, E. Pérez, R. Martí, A. Oliver, and J. Martí. Automatic seed placement for breast lesion segmentation on US images. *Breast Imaging*, 7361:308–315, 2012.

[94] K. V. Mogatadakala, K. D. Donohue, C. W. Piccoli, and F. Forsberg. Detection of breast lesion regions in ultrasound images using wavelets and order statistics. *Medical physics*, 33(4):840–849, 2006.

[95] W.K. Moon, S.C. Chang, J.M. Chang, N. Cho, C.S. Huang, J.W. Kuo, and R.F. Chang. Classification of breast tumors using elastographic and B-mode features: comparison of automatic selection of representative slice and physician-selected slice of images. *Ultrasound in Medicine and Biology*, 39(7):1147–1157, 2013.

[96] W.K. Moon, C.S. Huang, W.C. Shen, E. Takada, R.F. Chang, J. Joe, M. Nakajima, and M. Kobayashi. Analysis of elastographic and B-mode features at sonoelastography for breast tumor classification. *Ultrasound in Medicine and Biology*, 35(11):1794–1802, 2009.

[97] S. Nedevschi, C. Pantilie, T. Marita, and S. Dudea. Statistical methods for automatic segmentation of elastographic images. *4th International Conference on Intelligent Computer Communication and Processing (ICCP)*, pages 287–290, 2008.

[98] A. Oliver. *Automatic Mass Segmentation in Mammographic Images.* PhD thesis, University of Girona, July 2007.

[99] M. O'Mahony. *Sensory evaluation of food : statistical methods and procedures*, page 487. New York : Dekker, 1986.

[100] J. Ophir, I. Céspedes, H. Ponnekanti, Y. Yazdi, and X. Li. Elastography: a quantitative method for imaging the elasticity of biological tissues. *Ultrasonic Imaging*, 13(2):111–134, 1991.

[101] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1):62–66, 1979.

[102] G. Pons, J. Martí, R. Martí, S. Ganau, J.C. Vilanova, and J.A. Noble. Evaluating lesion segmentation on breast sonography as related to lesion type. *Journal of Ultrasound in Medicine*, 32(9):1659–1670, 2013.

[103] S. Prapavesis, B. Fornage, A. Palko, C.F. Weismann, and P. Zoumpoulis. *Breast Ultrasound and US-Guided Interventional Techniques: a Multimedia Teaching File.* Thessaloniki, Greece, 2003.

[104] D.M. Regner, G.K. Hesley, N.J. Hangiandreou, M.J. Morton, M.R. Nordland, D.D. Meixner, T.J. Hall, M.A. Farrell, J.N. Mandrekar, W.S. Harmsen, and J.W. Charboneau. Breast lesions: evaluation with US strain imaging-clinical experience of multiple observers. *Radiology*, 238(2):425–437, 2006.

[105] I. Saarenmaa, T. Salminen, U. Geiger, P. Heikkinen, S. Hyvärinen, J. Isola, V. Kataja, M.L. Kokko, R. Kokko, E. Kumpulainen, A. Kärkkäinen, J. Pakkanen, P. Peltonen, A. Piironen, A. Salo, M.L. Talviala, and M. Hakama. The effect of age and density of the breast on the sensitivity of breast cancer diagnostic by mammography and ultasonography. *Breast Cancer Research and Treatment*, 67(2):117–123, 2001.

[106] R. Schulz-Wendtland, M. Fuchsjäger, T. Wacker, and K.-P. Hermann. Digital mammography: an update. *European Journal of Radiology*, 72(2):258–265, 2009.

[107] C.M. Sehgal, T.W. Cary, S.A. Kangas, S.P. Weinstein, S.M. Schultz, P.H. Arger, and E.F. Conant. Computer-based margin analysis of breast sonography for differentiating malignant and benign masses. *Journal of Ultrasound in Medicine*, 23(9):1201–1209, 2004.

[108] S. Selvan, M. Kavitha, S. Shenbagadevi, and S. Suresh. Feature extraction for characterization of breast lesions in ultrasound echography and elastography. *Journal of Computer Science*, 6(1):67–74, 2010.

[109] J. Shan, H.D. Cheng, and Y. Wang. A novel automatic seed point selection algorithm for breast ultrasound images. In *19th International Conference on Pattern Recognition (ICPR)*, pages 1–4, 2008.

[110] J. Shan, H.D. Cheng, and Y. Wang. Completely automated segmentation approach for breast ultrasound images using multiple-domain features. *Ultrasound in Medicine and Biology*, 38(2):262 – 275, 2012.

[111] W.C. Shen, R.F. Chang, W.K. Moon, Y.H. Chou, and C.S. Huang. Breast ultrasound computer-aided diagnosis using BI-RADS features. *Academic Radiology*, 14(8):928–939, 2007.

[112] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:888–905, 1997.

[113] A.P. Smith, P.A. Hall, and D.M. Marcello. Emerging technologies in breast cancer detection. *Radiology management*, 26(4):16–24, 2004.

[114] R.A. Smith, D. Saslow, K.A. Sawyer, W. Burke, M.E. Costanza, W. Evans, R.S. Foster, E. Hendrick, H.J. Eyre, and S. Sener. American Cancer Society guidelines for breast cancer screening: update 2003. *CA: a cancer journal for clinicians*, 53(3):141–169, 2003.

[115] P. Soille. *Morphological Image Analysis*. Springer-Verlag, Heidelberg, 1999.

[116] A.T. Stavros, D. Thickman, C.L. Rapp, M.A. Dennis, S.H. Parker, and G.A. Sisney. Solid breast nodules: use of sonography to distinguish between benign and malignant lesions. *Radiology*, 196(1):123–134, 1995.

[117] K.J. Taylor, C. Merritt, C. Piccoli, R. Schmidt, G. Rouse, B. Fornage, E. Rubin, D. Georgian-Smith, F. Winsberg, B. Goldberg, and E. Mendelson. Ultrasound as a complement to mammography and breast examination to characterize breast masses. *Ultrasound in Medicine and Biology*, 28(1):19–26, 2002.

[118] U. Techavipoo, T. Varghese, J.A. Zagzebski, Q. Chen, and W. Liu. Semiautomated thermal lesion segmentation for three-dimensional elastographic imaging. *Ultrasound in Medicine and Biology*, 30(5):655–664, 2004.

[119] A. Thitaikumar, L.M. Mobbs, C.M. Kraemer-Chant, B.S. Garra, and J.Ophir. Breast tumor classification using axial shear strain elastography: a feasibility study. *Physics in Medicine and Biology*, 53(17):4809, 2008.

[120] A. Thomas, F. Degenhardt, A. Farrokh, S. Wojcinski, T. Slowinski, and T. Fischer. Significant differentiation of focal breast lesions: calculation of strain ratio in breast sonoelastography. *Academic Radiology*, 17(5):558–563, 2010.

[121] Z. Tu. Probabilistic boosting-tree: learning discriminative models for classification, recognition, and clustering. *10th IEEE International Conference on Computer Vision (ICCV)*, 2:1589–1596, 2005.

[122] L.A. Venta, C.M. Dudiak, C.G. Salomon, and M.E. Flisak. Sonographic evaluation of the breast. *Radiographics : a review publication of the Radiological Society of North America, Inc*, 14(1):29–50, 1994.

[123] P. Viola and M. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57:137–154, 2004.

[124] E. von Lavante and J.A. Noble. Segmentation of breast cancer masses in ultrasound using radio-frequency signal derived parameters and strain estimates. *5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 536–539, 2008.

[125] C. Wang, N. Komodakis, and N. Paragios. Markov Random Field modeling, inference and learning in computer vision and image understanding: a survey. *Computer Vision and Image Understanding*, 117(11):1610–1627, 2013.

[126] S.K. Warfield, K.H. Zou, W.M. Wells, and M. William. Simultaneous Truth and Performance Level Estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging*, 23:903–921, 2004.

[127] J. Weickert. Nonlinear diffusion filtering. *Handbook on computer vision and applications*, 2:423–450, 1999.

[128] W.M. Wells, W.E.L. Grimson, R. Kikinis, and F.A. Jolesz. Adaptive segmentation of MRI data. *IEEE Transactions on Medical Imaging*, 15(4):429–442, 1996.

[129] G. Xiao, M. Brady, J.A. Noble, and Y. Zhang. Segmentation of ultrasound B-mode images with intensity inhomogeneity correction. *IEEE Transactions on Medical Imaging*, 21(1):48–57, 2002.

[130] M. H. Yap, E. A. Edirisinghe, and H. E. Bez. A novel algorithm for initial lesion detection in ultrasound breast images. *Journal of Applied Clinical Medical Physics*, 9(4):181–199, 2008.

[131] C.K. Yeh, Y.S. Chen, W.C. Fan, and Y.Y. Liao. A disk expansion segmentation method for ultrasonic breast lesions. *Pattern Recognition*, 42(5):596–606, 2009.

[132] J. Youk, H. Gweon, E. Son, J. Kim, and J. Jeong. Shear-wave elastography of invasive breast cancer: correlation between quantitative mean elasticity value and immunohistochemical profile. *Breast Cancer Research and Treatment*, 138(1):119–126, 2013.

[133] Y. Yu and S.T. Acton. Speckle reducing anisotropic diffusion. *IEEE Transactions on Image Processing*, 11(11):1260–1270, 2002.

[134] J. Zhang, S.K. Zhou, S. Brunke, C. Lowery, and D. Comaniciu. Database-guided breast tumor detection and segmentation in 2D ultrasound images. *Proceedings of SPIE Medical Imaging 2010: Computer-aided diagnosis.*, 7624, 2010.

[135] J. Zhang, S.K. Zhou, S. Brunke, C. Lowery, and D. Comaniciu. Detection and retrieval of cysts in joint ultrasound B-mode and elasticity breast images. *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 173–176, 2010.

[136] Y. Zhang, M. Brady, and S. Smith. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1):45–57, 2001.

[137] H. Zhi, B. Ou, B.M. Luo, X. Feng, Y.L. Wen, and H.Y. Yang. Comparison of ultrasound elastography, mammography, and sonography in the diagnosis of solid breast lesions. *Journal of Ultrasound in Medicine*, 26(6):807–815, 2007.

[138] H. Zhi, X.Y. Xiao, H.Y. Yang, B. Ou, Y.L. Wen, and B.M. Luo. Ultrasonic elastography in breast cancer diagnosis: strain ratio vs 5-point scale. *Academic Radiology*, 17(10):1227–1233, 2010.

[139] X. Zhou, T. Liang, and W. Yu. Accurate segmentation of ultrasound images using the motion cue. *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 332–335, 2010.